Elementary Linear Algebra

Chris Johnson

Fall 2024

Contents

Introduction to the Course

Mathematics is the art of reducing any problem to linear algebra.

WILLIAM STEIN

What is linear algebra?

Most mathematicians see mathematics as falling into two broad categories: applied math and pure math. Applied mathematics is the math that is concerned with solving "real world" problems that occur in engineering, economics, and the sciences, and is probably what most educated non-mathematicians think of when they think of math. Pure mathematics, however, is more akin to philosophy or art: it is the mathematics that is studied because it is considered interesting or beautiful.

Linear algebra is one of the few mathematical disciplines that falls squarely in both categories. It is a collection of ideas and techniques that are intrinsically interesting, but also profoundly useful. It is extremely difficult, and perhaps impossible, to name an area of modern mathematics, science, technology, or engineering that does not use linear algebra in some way. Not only are the tools of linear algebra useful in practical applications, they are fundamental in other advanced areas. For example, much of algebraic topology involves associating certain vector spaces (which are the principal objects of study in linear algebra) to topological spaces, reducing topological problems into linear algebraic problems.

What is the structure of this course?

Some of the topics that we will study in this course are things you may have seen before in a high school algebra class; systems of linear equations and matrices are often (but not always) studied in high school. Some ideas, such as sets, will be familiar to students that have already taken courses in discrete mathematics or logic and proof, but will be new to other students. And some topics will be completely alien to everyone in the class.

This course will start at the beginning by reviewing systems of linear equations. Because this is a topic that should be somewhat familiar to most students in the course, we will go through this material relatively quickly in lecture, but detailed lecture notes will be provided if you want or need to see a deeper explanation than what is mentioned in class.

After we have covered the basics (solving systems of linear equations and matrix algebra) we will begin to move into more advanced territory, and we will consider topics which at first will seem very abstract and esoteric. However, we will see that even these strange and exotic ideas have real world applications. One of the reasons we will go through the early material quickly is so that we will have more time to discuss the more difficult material in the second half of the course.

There is no denying that this course will require a level of abstraction and sophistication that you are likely not used to. This may seem very jarring, especially if you are the type of student that has typically understood material in previous mathematics courses without a great deal of trouble. I mention this not to scare you away from the course, but just to reassure you that feeling frustrated and confused is a normal part of the learning process, especially when you are faced with ideas that are unlike anything you have seen before. I have every confidence that each student is capable of understanding the material in this course, but you have to accept that this course will require you to work hard and sometimes you will be frustrated.

How can students succeed?

I firmly believe that every student is capable of succeeding in this course, but I also know that some students will struggle and so I want to mention a few concrete things that you can do to succeed in the course.

Recognize that this class is difficult

In the first couple of weeks of the semester this class may seem relatively straight-forward, but you should not let this lull you into thinking the entire semester will be easy. One of the difficult things about this course is that we will cover a lot of material, and each new topic will build off the previous ones. If you start to fall behind, you will have a very small window of time to get back up to speed before being behind will negatively affect your grade.

Study every day

To do well in this course you will need to invest a significant amount of time into studying outside of class. Sitting in lecture, even if you feel like you understand what is going on during the lecture, will not be enough. You should get into the habit of studying every day: not just the days the class meets, not just on weekdays, but every day. Something as simple as putting aside one hour for individual studying outside of class each day can have a huge impact on your grade and keep you from falling behind. Sometimes you will have other commitments that prevent you from getting an hour each day, but when at all possible, you should really try to study at least one hour each day. If you can't manage an hour one day, do what you can; even fifteen or twenty minutes of studying is *much* better than no studying at all.

When you study you should first review your notes from class; notice this implies that you need to be taking notes in class. If there is something from your notes that you don't understand, try to figure it out. It's best if you try to figure things out on your own first without having to look in a book or online: sometimes you just need to spend a few minutes thinking through the details of some calculation or the logic behind some argument before it starts to make sense.

Read your book and the posted lecture notes. The book and the lecture notes cover the same material, but sometimes presented in different ways. By reading both you see the same ideas from two different points of view. This can be helpful because one point of view may "click" when the other does not.

Be incredulous

To do well in advanced math courses you should try to think like a mathematician. This means trying to understand the ins and outs of every argument, why each step in a computation was performed, what earlier results were used, etc. In general, you should be incredulous: you should not simply take it on faith that what we have learned in class is true (even though it is!), but you should instead always ask why it's true and try to figure out the reason. This one piece of advice, if taken to heart, subsumes everything else.

Practice, then practice some more

You know that you understand a concept well and are prepared for an exam when you have practiced so much that solving problems becomes mechanical. For example, think about solving for x in an equation such as $x^3 = x^2 + 2x$. The first time you started learning algebra this may have seemed very odd and difficult, but as you did more examples you started to notice the patterns and the tricks, and now you (hopefully) are able to solve for x in equations like the one above without any trouble.

Similarly, in this class you will probably find some computations and some logical arguments very difficult and time-consuming at first. If you do enough examples, however, then the things that at first seem difficult and confusing will over time become second nature. The only way for this to happen is to invest time in practicing. When you review your notes and see an example that we did in class, try to reproduce the result without looking at the notes and then look back at the notes if you get stuck or make a mistake. Do the practice problems at the end of each set of lecture notes; pick and choose extra problems at the end of the sections in the book; make up your own problems; look for extra problems online. The more practice you do, the easier everything will be when you actually sit down to take an exam.

Prepare for exams

The biggest mistake you can make when an exam is coming up is to put off studying for it. The earlier you start preparing for an exam, the better. When an exam is coming up, start adding more time to your usual study sessions. Ideally you should add an extra hour each day for a week leading up to an exam. It's probably best to try to split this up into two one-hour study sessions each day instead of doing two hours at once.

When you're preparing for an exam, you should study as if the upcoming exam is the hardest one you have ever taken in your life. (This isn't to say that it necessarily will be the hardest exam you ever take, but it's better to over-prepare than to under-prepare.) If there is a topic you don't feel comfortable with or are worried about, don't ignore it! Study as if the problems you dislike and find hard will be on the exam: chances are at least a couple types of problems you dislike will make their way onto an exam at some point.

To prepare for an exam you should review all of the relevant notes, look over any returned homeworks or quizzes, and try to understand any mistakes that you made. It is very likely that problems from homework or quizzes will reappear on an exam.

Come to class

For some strange reason there always seem to be people who think it's okay to skip class. You should come to class each and every day which you are physically able. In class you should be actively paying attention to the lecture and trying to think through examples as they are done on the board. If you have questions in class, then that's good! Having questions means that you're thinking, which is what you should be doing in class.

You *should not* be daydreaming, working on assignments for other classes, playing games on your phone, or checking Facebook / X (formerly known as Twitter) / Tumblr / Instagram / Snapchat / Tinder / YikYak / etc.

Start assignments early and work on them often

You will usually have at least a week to do a homework assignment, and that is for a reason. Some of the questions on these assignments will be difficult and you will have to spend some time thinking in order to do the assignment. You should really try to start on assignments early, meaning the day they are assigned, and try to do a few problems each day. You should also anticipate that some questions are going to require a lot of time – maybe an hour or more for the hardest questions. If you wait until the last minute to do an assignment, you won't have time to get it done.

Get help when you need help

Sometimes this class will be hard, but if you're willing to ask questions and get help when you really need help, you will find the material much easier. I encourage you to work with other students and look up resources online like Khan Academy and MIT's Open Course Ware when you're having difficulty with a problem or concept. You can also email me (`cjohnson@wcu.edu`) or drop in during my office hours.

In general when you have a specific concern about an assignment or a topic from class, you should try to address your concern by taking advantage of resources in the following order:

1. Try to figure things out yourself. There will be plenty of times when you just need to spend a little bit of time thinking on your own and you can figure things out.
2. Check in the book or lecture notes. Many questions you have will be answered in the book or notes, you just have to take the time to look through the book/notes and find it.
3. Ask a classmate. Sometimes you may have a misconception or misheard something in lecture, and asking a friend might be all it takes for you to realize your mistake.
4. Ask other students in the Math Tutoring Center.
5. Email me or come to office hours. I put this at the end of the list not because I'm trying to avoid seeing you or talking to you, but just as a matter of practicality. If everyone in the class came to me the instant they had a question I would spend my entire day answering their questions. I am fine with answering your questions or talking to you when you have concerns, but I also have other classes and other responsibilities that I have to attend to.

If you have more serious concerns about your standing in the class – not simply a homework problem you can't figure out – then by all means contact me first.

Don't stress out (too much)

There will be times when this class frustrates you: maybe there is a topic you can't seem to wrap your head around, or a problem that you feel like you're staring at and have no idea how to get started. This is completely normal and you shouldn't get too stressed out about it. This class is going to be hard and you are going to get confused and feel stuck sometimes, but that is just a normal part of learning difficult material. The important thing to remember is that you should persevere. If you're getting frustrated, take a break: go get something to eat, play a game, read a book, take a nap; do something you enjoy for a little while and then get back to work when you're ready.

The lecture notes

The notes you are reading are in their third incarnation, having evolved from the handwritten examples I used when I first taught a version of this course at Wake Forest University, and then updated when I taught the course at Indiana University. There may be places where the notes are "rough around the edges" and may contain typos and mistakes (though hopefully those are all minor). If you see something in the notes you think is a mistake, it may very well be, and it would be greatly appreciated if you would email me (cjohnson@wcu.edu) to let me know about any mistakes. While these notes are my primary resource for the examples I use in the lectures, they should not be a substitute for the textbook. Besides the fact that your textbook has fewer mistakes than these notes (probably not mistake-free, but relatively few and minor mistakes) since it was professionally edited, the textbook also has lots of exercises and practice problems, which these notes do not. I hope these notes are helpful to you, but you should not use them as your only source of study material.

Chris Johnson
Fall 2024

Part I

Linear Systems and Matrices

1

Systems of Linear Equations

> *"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."*
>
> LEWIS CARROLL
> *Alice in Wonderland*

Linear algebra is one of the most fundamental tools in mathematics, engineering, and the sciences. Many objects in both mathematics and physics are defined in terms of linear algebra, and the tools of linear algebra are then used to study those objects. From differential geometry and Einstein's theory of general relativity to the practical, real-world optimization problems that occur in industry, linear algebra is everywhere. One reason for this is because linear algebra is extremely well-understood, but also because linear algebra can be done very efficiently on a computer. This means that describing a problem of interest in terms of linear algebra is often the first step to understanding and ultimately solving that problem.

In this class we will describe the fundamentals of linear algebra, including linear transformations, matrix algebra, determinants, eigenvectors & eigenvalues, and inner products & orthogonality (if time allows). The starting point for all of this, however, is solving systems of linear equations. Systems of linear equations appear throughout mathematics, and over the years people have developed a set of algorithms (mathematical recipes) for how to solve such systems, or determine that they can not be solved. Everything else in linear algebra is built upon the ideas involved in solving a system of linear equations, so understanding these systems is necessarily the proper place to begin a study of linear algebra.

1.1 Linear Equations

A ***real linear function*** is a map from $\mathbb{R}^n$ to $\mathbb{R}$ which has the form

$$(x_1, ..., x_n) \mapsto a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

where the $x_1, x_2, ..., x_n$ are independent variables (i.e., values that are allowed to change), and the $a_1, ..., a_n$ are constants that have been chosen and will not change. These $a_1, ..., a_n$ values are called the ***coefficients***. Here we are assuming the coefficients and the variables are all real numbers.

We could define a ***complex linear function*** similarly: this is a map from $\mathbb{C}^n$ to $\mathbb{C}$ of the form

$$(z_1, ..., z_n) \mapsto \alpha_1 z_1 + \alpha_2 z_2 + \cdots + \alpha_n z_n$$

where the α_i and z_i are complex numbers.

The majority of the material in this class will apply equally well to both real and complex numbers, and other more exotic number systems. Whenver we say a function is ***linear*** without specifying whether we meal real linear or complex linear, we mean that either a real linear function or a complex linear function can be used.

The number n tells us how many variables and coefficients there are. In this class n will be allowed to be any positive integer or infinity. Usually if there are only two variables (i.e., we have a function from $\mathbb{R}^2$ to $\mathbb{R}$), then we will call the variables x and y instead of x_1 and x_2; similarly, if there are only three variables (the linear function is from $\mathbb{R}^3$ to $\mathbb{R}$), then we will use x, y, and z as the variables.

Example 1.1.
Each of the following are linear functions:

(a) $T(x, y, z) = 3x - 2y + \pi z$

(b) $f(x, y) = -6x + \frac{17}{3} y$

(c) $6x_1 - 7x_2 + 4x_4 - e^2 x_5$

(d) $(z_1, z_2) \mapsto iz_1 + 2z_2$

(e) $(x_1, x_2, ..., x_n) \mapsto 0$

Example 1.2.
The following *are not* linear functions:

(a) $g(x, y, z) = x^2 + 2y - 3z$

(b) $h(x, y) = xy$

(c) $T(x, y, z) = 3x - 2y + \pi z - 4$

(d) $(z_1, z_2) \mapsto z_1 + e^{z_2}$

(e) $(x_1, x_2, ..., x_n) \mapsto 1$

Exercise 1.1.
Why is Example 1.1(e) linear, but Example 1.2(e) not?

A ***linear equation*** is an equation where each side of the equation is a linear function or a constant.

Example 1.3.
Each of the following is a linear equation:

(a) $6x + 2y = 3$

(b) $x_1 - x_2 + 3x_3 + \frac{22}{7}x_4 = 2$

(c) $x - y + z = 0$

Example 1.4.
The following *are not* linear equations:

(a) $xy = 1$

(b) $\frac{x_1+x_2}{x_3} = x_4 - x_5$

(c) $x^2 = x + y$

Let's go ahead and notice at this point that the set of all solutions to a linear equation in two variables gives us a line. For example, the collection of all (x, y) pairs that satisfy the linear equation

$$6x + 2y = 3$$

is a line. This might be easiest to see if we take our linear equation and rewrite in the more familiar slope-intercept form of a line by solving for y:

$$\begin{aligned} 6x + 2y &= 3 \\ \implies 2y &= -6x + 3 \\ \implies y &= -3x + \frac{3}{2} \end{aligned}$$

This is a line of slope -3 which passes through the point $(0, 3/2)$ as in Figure 1.1.

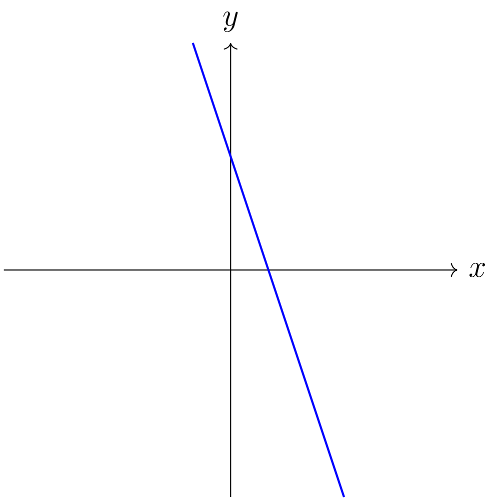


Figure 1.1: The set of points satisfying the linear equation $6x + 2y = 3$ is a line in the plane.

It is because of this graphical interpretation that equations of the form $ax + by = c$ in two dimensions give lines that we call the functions above *linear* functions and equations with linear functions are *linear equations*.

Notice that in three dimensions the set of solutions to a linear equation give a plane and not a line, as in Figure 1.2, but we still use the term "linear."

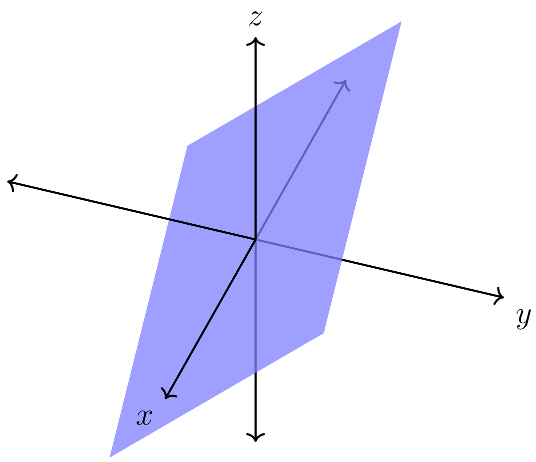


Figure 1.2: The set of points satisfying the linear equation $x - y + z = 0$ is a plane in 3-space.

A ***system of linear equations*** is a collection of linear equations, all in the same number of variables.

Example 1.5.
Each of the following are systems of linear equations:

(a)

$$\begin{aligned} 3x + 2y &= 4 \\ 6x - y &= 9 \end{aligned}$$

(b)

$$\begin{aligned} 4x + 3y &= 3 \\ 4x + 3y &= 2 \end{aligned}$$

(c)

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

A ***solution*** to a system of a linear equations is a collection of numbers, one for each variable, which makes *all* of the equations true simultaneously. In the case of Example 1.5(a) it's easy to check that $(x, y) = ({}^{22}/_{15}, -{}^{1}/_{5})$ is a solution to the system by simply plugging $x = {}^{22}/_{15}$ and $y = -{}^{1}/_{5}$ into each equation in the system and verifying that both equations are true:

$$\begin{aligned} 3 \cdot \frac{22}{15} + 2 \cdot \frac{-1}{5} =& \frac{22}{5} - \frac{2}{5} \\ =& \frac{20}{5} \\ =& 4 \end{aligned}$$

$$\begin{aligned} 6 \cdot \frac{22}{15} - \frac{-1}{5} =& \frac{44}{5} + \frac{1}{5} \\ =& \frac{45}{5} \\ =& 9 \end{aligned}$$

In Example 1.5(b) it's also easy to see that there are *no solutions* to the system: there is no choice of x and y that can make $4x + 3y = 3$ and $4x + 3y = 2$ at the same time, since $3 \neq 2$.

It's a little bit harder to see, but there are actually infinitely-many different solutions to Example 1.5(c). Let's try to explain why this is. If we solve the first equation for y we have

$$y = 3 - x - z.$$

So triple (x, y, z) solving the system has to also satisfy this equation (since this is just the first equation rewritten). Now if (x, y, z) is a solution to the system, then it must also solve the second equation as well as $y = 3 - x - z$.

This means we can rewrite the second equation as

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

If we now solve this equation for z we have

$$\begin{aligned} & 2x - (3 - x - z) + 3z = 0 \\ \implies & 3x + 4z - 3 = 0 \\ \implies & z = \frac{3 - 3x}{4}. \end{aligned}$$

If we now plug this back into $y = 3 - x - z$ we have

$$\begin{aligned} y =& 3 - x - \frac{3 - 3x}{4} \\ =& \frac{12 - 4x}{4} - \frac{3 - 3x}{4} \\ =& \frac{9 - x}{4}. \end{aligned}$$

So, what does this mean? It means if (x, y, z) is a solution to the system, then y and z are both determined by x:

$$y = \frac{9 - x}{4} \quad \text{and} \quad z = \frac{3 - 3x}{4}.$$

Here x can be whatever value you'd like (in a situation like this we sometimes call x a ***free variable***) and once you've chosen x, you know what y and z must be. Since there are infinitely-many different choices for x (x can be any real number you'd like), there are infinitely-many solutions.

Right now the above algebra probably seems tedious – easy, but a little bit of boring work to figure out. We will quickly see that there are some algorithms that make finding solutions to systems like this much easier. Before doing that, however, let's talk about the geometry of the set of solutions to a system of linear equations.

1.2 Solution Sets

Given a system of linear equations our goal will typically be to find all possible solutions to the system. The collection of all possible solutions is called the ***solution set*** of the system. In principle, the solution set of an arbitrary system of equations could be very complicated, but for systems of *linear* equations, the solution sets are actually very nice.

In fact, the solution set of a system of linear equation comes in one of three flavors: it could be empty (no solutions), it could contain exactly one point (a unique solution), or it could contain infinitely-many points. Let's think about why this is in two variables.

Let's suppose that you had a system of linear equations in two variables: say there are n equations, and the i-th equation has the form $a_i x + b_i y = c_i$, so the system looks something like the following:

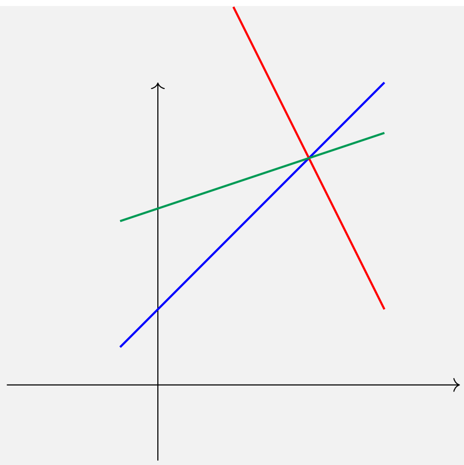
$$\begin{aligned} a_1 x + b_1 y &= c_1 \\ a_2 x + b_2 y &= c_2 \\ &\vdots \\ a_n x + b_n y &= c_n. \end{aligned}$$

Each one of the equations determines a line in the plane. The solution set of the system is the collection of points that are simultaneously on all of the lines. It could be that all of the lines intersect at a single point giving a unique solution; it could be that no point is on all of the lines at the same time (no solution); or it could be that all the lines are actually the same and there are infinitely-many solutions (every point on the line is a solution).

Let's consider one example of each situation just by considering the graphs of the lines.

Example 1.6.
The following system has one unique solution:

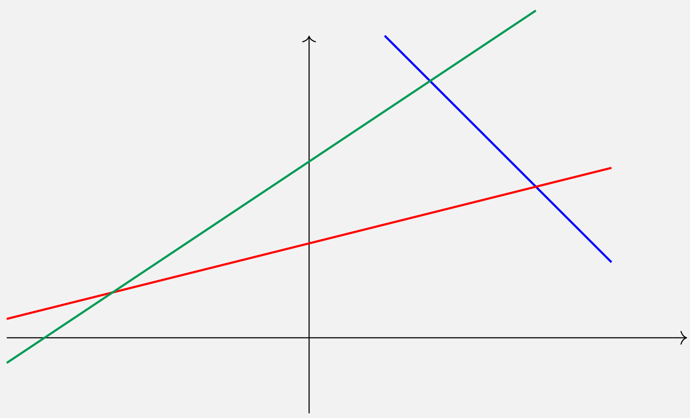
$$\begin{aligned} x - y &= -1 \\ 2x + y &= 7 \\ -3x + 9y &= 21 \end{aligned}$$



Here there is a unique solution because there is exactly one point that is on all three lines.

Example 1.7.
The following system has no solutions:

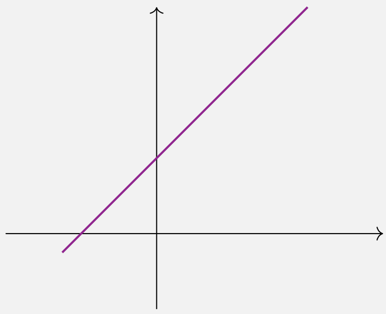
$$\begin{aligned} x + y &= 5 \\ -2x + 8y &= 10 \\ 2x - 3y &= 7 \end{aligned}$$



There are no solutions because there is not a point that is on all three lines simultaneously.

Example 1.8.
The following system has infinitely-many solutions:

$$\begin{aligned} -x + y &= 1 \\ 2x - 2y &= -2 \\ -3x + 3y &= 3 \end{aligned}$$



In this case all three lines are actually the same line on top of one another, so any solution to one equation is instantly a solution to both of the other equations.

The same situation can happen in any number of variables: regardless of whether your system of equation is in 2 variables, 3 variables, or 457 variables, a system of linear equations has either no solutions, one unique solution, or infinitely-many solutions.[1]

Exercise 1.2.

[1]This is true if we're talking about real or complex linear systems. We will see later in the semester that there are some applications where we'd like to use another type of number system, and in that case it could be possible to have only finitely-many distinct solutions to a system.

(a) Plot each of the lines in the system below, and determine if the system has no solutions, one solution, or infinitely-many solutions. If there is one unique solution, determine what that solution is.

$$\begin{aligned} x + y =& 1 \\ x - 2y =& 3 \end{aligned}$$

(b) Repeat part (a), but add $4x + 7y = 9$ as a third equation to the system.

Whenever a system of equations has a solution (regardless of whether it has one solution or infinitely-many) we say the system is ***consistent***. If the system does not have any solutions, however, we say the system is ***inconsistent***. So the systems in Example 1.6 and Example 1.8 are consistent, while the system in Example 1.7 is inconsistent.

1.3 Solving a System of Linear Equations

Now that we what a linear system is and the different "flavors" the solution can come in, how do we go about determining the solutions?

Let's consider "manually" solving a couple of different systems before we try to come up with an algorithm.

Example 1.9.
Solve the following system:

$$\begin{aligned} x - y =& 7 \\ y =& 3 \end{aligned}$$

This is a system everyone can solve without any knowledge of linear algebra: the second equation tells us explicitly what y has to be, so all of the (x, y) solutions to the system have to take y to be three. Plugging this back into the first equation we then have $x - 3 = 7$

and we can easily solve $x = 10$. Thus $(x, y) = (10, 3)$ is a solution, the unique solution, to this system.

Example 1.10.
Solve the following system:

$$\begin{aligned} x - y =& 7 \\ 2x - y =& 17 \end{aligned}$$

To solve this system we might first try to isolate a variable in one of the equations. For example, if we subtract twice the first equation from the second equaton we would have

$$\begin{aligned} & 2x - y - 2(x - y) = 17 - 2 \cdot 7 \\ \implies & 2x - y - 2x + 2y = 17 - 4 \\ \implies & y = 3 \end{aligned}$$

We are then back to exactly the same situation as the previous example and so the solution is again $(10, 3)$.

When two different systems of equations have the same set of solutions, such as the examples above, we say the systems are ***equivalent***.

Notice that certain types of systems of equations are very easy to solve, and others can seem more complicated. Here is another easy example:

Example 1.11.
Solve the following system

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

This system is very easy to solve because one of the equations instantly tells us what one of the variables has to be: we know that z must be -1. If we plug this into the second equation we have $y - 2 = 2$, so $y = 4$. Now that we know y and z, we can plug back into the first equation to determine $2x+12+1 = 4$, and thus $x = \frac{-9}{2}$.

We really like the types of systems as in the last example because they are almost trivial to solve: we just plug back into our equations and get one variable at a time. We would like it, then, if when given a more complicated system we were somehow able to determine an equivalent, but easy-to-solve, system. Since the systems are equivalent, solving the easy system tells us the solution to the complicated system.

The main question, then, is how do we determine if two systems are equivalent?

To do this, let's come up with a list of some simple operations that we can perform on a system of equations to come up with an equivalent system. What we are about to describe will work for any number of systems in any number of variables, so we'll state things very generally but then do some simpler examples.

Theorem 1.1.
If two rows in a linear system are exchanged, the newly obtained system is equivalent to the original one.

Proof.
The equations defining the system haven't been changed, just reordered. □

Example 1.12.

The following two systems are equivalent:

$$\begin{aligned} 2x + 3y - z &= 4 \\ x - y + 3z &= 9 \\ -x + 6y + 4z &= -2 \end{aligned} \qquad \begin{aligned} -x + 6y + 4z &= -2 \\ x - y + 3z &= 9 \\ 2x + 3y - z &= 4 \end{aligned}$$

Theorem 1.2.
If one equation in a linear system is modified by adding a multiple of another equation to it, the newly obtained system is equivalent to the original one.

Before proving this theorem in general, let's consider a very simple case: two variables and two equations. Say our system looks like

$$\begin{aligned} a_1x + a_2y &= b \\ \alpha_1x + \alpha_2y &= \beta \end{aligned}$$

Suppose the system is consistent and so there's some point (s_1, s_2) that satisfies the system: if we plug in $x = s_1$ and $y = s_2$, then both equations are solved simultaneously.

Now say that we modify the system by adding a multiple of the second equation to the first. That is, we will replace the first equation by adding c times the second equation to it, for some constant c. We then have the following system:

$$\begin{aligned} (a_1 + c\alpha_1)x + (a_2 + c\alpha_2)y &= b + c\beta \\ \alpha_1x + \alpha_2y &= \beta \end{aligned}$$

We claim that (s_1, s_2) is still a solution to this system. Since (s_1, s_2) satisfied the second equation before, and that second equation hasn't changed, all we need to do is verify that (s_1, s_2) satisfies the modified first equation, but this is easy:

$$\begin{aligned} &(a_1 + c\alpha_1)s_1 + (a_2 + c\alpha_2)s_2 \\ &= a_1s_1 + c\alpha_1s_1 + a_2s_2 + c\alpha_2s_2 \\ &= (a_1s_1 + a_2s_2) + c(\alpha_1s_1 + \alpha_2s_2) \\ &= b + c\beta \end{aligned}$$

A solution to the original system is thus a solution to this modified system as well. This shows that the solution set of the first system is a subset of the solution set of the second system. We still need to show that a solution to the modified system is also a solution to the original system, but the idea is basically the same as the above, so we will leave that as an exercise.

Exercise 1.3.
Show that if (t_1, t_2) is a solution to the system

$$\begin{aligned}(a_1 + c\alpha_1)x + (a_2 + c\alpha_2)y &= b + c\beta \\ \alpha_1 x + \alpha_2 y &= \beta,\end{aligned}$$

then it is also a solution to the system

$$\begin{aligned}a_1 x + a_2 y &= b \\ \alpha_1 x + \alpha_2 y &= \beta.\end{aligned}$$

Proving the general theorem is basically repeating the same argument above, just with more equations and variables.

Proof of Theorem 1.2.
Consider a linear system of m equations in n variables. Say two of the equations in this system are

$$a_1 x_1 + \cdots + a_n x_n = b \quad \text{and} \quad \alpha_1 x_1 + \cdots + \alpha_n x_n = \beta.$$

We want to leave the second equation alone, but replace the first equation with

$$(a_1 + c\alpha_1)x_1 + \cdots + (a_n + c\alpha_n)x_n = b + c\beta$$

for some constant c. The claim is that doing so doesn't change the set of solutions.

Let's call the set of solutions to the original system S, and the set of solutions to the modified system T. We want to show these two sets are the same: we want to show that $S = T$, which means we need to show that $S \subseteq T$ and $T \subseteq S$.

Let $(s_1, ..., s_n) \in S$ be a solution to the original system. We need to show this is also a solution to the modified system. Of the equations defining the systems, however, $m - 1$ of the equations are the same. So the only thing we need to check is that $(s_1, ..., s_n)$ is also a solution to

$$(a_1 + c\alpha_1)x_1 + \cdots + (a_n + c\alpha_n)x_n = b + c\beta$$

We simply plug in $(x_1, ..., x_n) = (s_1, ..., s_n)$ and verify that the equation holds:

$$\begin{aligned} &(a_1 + c\alpha_1)s_1 + \cdots + (a_n + c\alpha_n)s_n \\ =&a_1s_1 + c\alpha_1s_1 + \cdots + a_ns_n + c\alpha_ns_n \\ =&(a_1s_1 + \cdots + a_ns_n) + c(\alpha_1s_1 + \cdots + \alpha_ns_n) \\ =&b + c\beta \end{aligned}$$

This shows that $S \subseteq T$.

We leave the second part of the proof, that $T \subseteq S$, as an exercise. □

Exercise 1.4.
Finish the proof of Theorem 1.2.

Finally, there's one last operation that we will introduce that can be used to replace one system of equations with an equivalent one.

Theorem 1.3.
If each term (both the left- and right-hand sides) of one equation is multipled by a nonzero constant c, then the newly obtained system is equivalent to the original system.

Again, let's consider what's happening with two variables and two equations. If our original system was

$$a_1x + a_2y = b$$
$$\alpha_1x + \alpha_2y = \beta$$

then we claim that the following system is equivalent

$$ca_1x + ca_2y = cb$$
$$\alpha_1x + \alpha_2y = \beta$$

when c is any nonzero constant.

To prove this, again suppose that (s_1, s_2) is a system to the original system. We can easily verify that (s_1, s_2) solves the modified system. Of course, the second equation has remained the same, so (s_1, s_2) still satisfies it. For the first equation we have

$$ca_1s_1 + ca_2s_2$$
$$=c(a_1s_1 + a_2s_2)$$
$$=cb$$

To prove that a solution (t_1, t_2) to the modified system is also a solution to the original system we can perform the exact same procedure: just multiply through by $\frac{1}{c}$ to get the c's to cancel! In more variables and/or equations, the argument is exactly the same, so we will leave that as an exercise.

Exercise 1.5.
Prove Theorem 1.3.

1.4 A Procedure for Solving Linear Systems

We now want to use the three theorems above to develop a scheme for solving systems of linear equations, or determining that there is no solution. As we saw in an earlier example, it would be nice if the system we wanted to solve had the following sort of form:

$$ax + by = \alpha \qquad ax + by + cz = \alpha$$
$$cy = \beta \qquad dy + ez = \beta$$
$$fz \qquad = \gamma$$

In this situation it is super-easy to "work backwards," solving for one variable at a time, and then determining the others. Let's give systems of this form a special name so it's easier to refer to them: we will call a system like this is in ***echelon form***.

To solve a system that is not in echelon, let's try to replace the system with an equivalent system that *is* in echelon form. We'll do this by repeatedly applying our three theorems above, modifying the system a little bit at a time until it is in the form we'd like.

Example 1.13.
Solve the following system of equations.

$$\begin{aligned} x + 4y &= 3 \\ 2x - y &= 1 \end{aligned}$$

All we need to do to put the system in echelon form is get rid of the $2x$ in the second equation. We can do this by subtracting twice the first equation from the second. We will then replace the second equation with

$$\begin{aligned} 2x - y - 2(x + 4y) &= 1 - 2 \cdot 3 \\ \implies -9y &= -5 \end{aligned}$$

thus $y = \frac{5}{9}$, and pluggin this back into the first equation,

$$\begin{aligned} x + 4 \cdot \frac{5}{9} &= 3 \\ \implies x = 3 - \frac{20}{9} &= \frac{7}{9} \end{aligned}$$

and so the system has a unique solution, $(x, y) = (^5/_9, ^7/_9)$.

Example 1.14.

Solve the following system of equations.

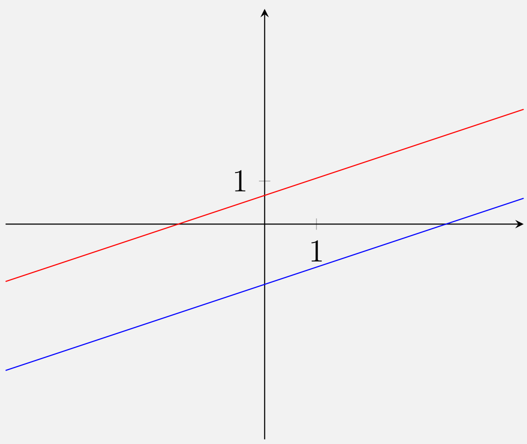
$$\begin{aligned} 2x - 5y &= 7 \\ -6x + 15y &= 10 \end{aligned}$$

We again try to put the system into echelon form by getting rid of the $-6x$ in the second equation. To do this, we add three times the first equation to the second:

$$\begin{aligned} &- 6x + 15 + 3(2x - 5y) = 7 + 3 \cdot 10 \\ \implies &0 = 37 \end{aligned}$$

Now we have a problem: zero is not equal to thirty-seven! What this means is that there is no solution to the system.

Geometrically, the lines determined by each of the equations above are parallel. A solution to the system would be where the two lines intersect, but two parallel lines never intersect, hence there is no solution.



Example 1.15.

Solve the following system:

$$\begin{aligned} 3x + 2y - z &= 3 \\ 12x - 4y + 2z &= 1 \\ 15x - 2y + z &= 4 \end{aligned}$$

To put this system into echelon form we need to first get rid of the x terms in the second and third equations. We will replace the second equation by subtracting four times the first equation, and we'll replace the third equation by subtracting five times the first equation.

The second equation then becomes,

$$\begin{aligned} & 12x - 4y + 2z - 4(3x + 2y - z) = 1 - 4 \cdot 3 \\ \Longrightarrow & -12y + 6z = -11 \end{aligned}$$

While the thid equation becomes

$$\begin{aligned} & 15x - 2y + z - 5(3x + 2y - z) = 4 - 5 \cdot 3 \\ \Longrightarrow & -12y + 6z = -11 \end{aligned}$$

Our system thus far is

$$\begin{aligned} 3x + 2y - z &= 3 \\ -12y + 6z &= -11 \\ -12y + 6z &= -11 \end{aligned}$$

To put the system in echelon form we need to get rid of the y term in the third equation, but doing this will of course get rid of all of the terms in the third equation. Thus the echelon form of the system is

$$\begin{aligned} 3x + 2y - z &= 3 \\ -12y + 6z &= -11 \end{aligned}$$

(If you want, there's an equation $0x + 0y + 0z = 0$ at the bottom of this.)

Let's notice that if we try to kill of y in the second equation by adding six times the first equation, we would also kill of z and our system would become

$$\begin{aligned} 3x + 2y - z &= 3 \\ 18x &= 7 \end{aligned}$$

So $x = {}^7/_{18}$: no matter what y and z happen to be, x must be ${}^7/_{18}$. Geometrically, this means all (x, y, z) solutions to our system must live in the plane $x = {}^7/_{18}$.

We could rewrite the first equation as

$$\frac{7}{6} + 2y - z = 3$$

and solving for y we would have

$$y = \frac{z}{2} + \frac{11}{12}.$$

This means that each solution to our system has the form

$$\left(\frac{7}{18}, \frac{z}{2} + \frac{11}{12}, z\right)$$

and z can take on any value: our set of solutions is a line in the plane $x = \frac{7}{18}$:

$$\left\{\left(\frac{7}{18}, \frac{z}{2} + \frac{11}{12}, z\right) \middle| z \in \mathbb{R}\right\}.$$

In the previous example, z is called a ***free variable*** because it can take on any value we wish. When we express the solution set in terms of free variables, such as above, we have a ***parametrization*** of the solution set.

Example 1.16.
Solve the following system of equations.

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

Let's first try to kill off the x in the second equation by subtracting the first equation from it (i.e., we we are applying Theorem 1.2 by adding -1 times the first equation to the second equation). The second equation is then replaced with

$$x - 2y + z - (x + y - 3z) = 3 - 4$$
$$\implies -3y + 4z = -1$$

Now our system looks like

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

We still need to kill of the $-3x$ in the third equation, so let's add three times the first equation to it. The third equation then becomes

$$-3x + y + 4z + 3(x + y - 3z) = 0 + 3 \cdot 4$$
$$\implies 4y - 5z = 12$$

So far we have replaced our original system with the following equivalent one:

$$\begin{aligned} x + y - 3z &= 4 \\ -3y + 4z &= -1 \\ 4y - 5z &= 12 \end{aligned}$$

We need to perform one last step to put the system in echelon form. Let's get rid of the $4y$ in the third equation by adding $\frac{4}{3}$ the second equation:

$$4y - 5z + \frac{4}{3}(-3y + 4z) = 12 + \frac{4}{3}(-1)$$
$$\implies -5z + 16/3z = 12 - \frac{4}{3}$$
$$\implies \frac{1}{3}z = \frac{32}{3}$$

We now have a system in echelon form that's equivalent to our original system:

$$\begin{aligned} x + y - 3z &= 4 \\ -3y + 4z &= -1 \\ \frac{1}{3}z &= \frac{32}{3} \end{aligned}$$

The system is already in echelon form, but let's kill off that $\frac{1}{3}$ in the third equation by using Theorem 1.3 to multiply the third equation by three:

$$\begin{aligned} x + y - 3z &= 4 \\ -3y + 4z &= -1 \\ z &= 32 \end{aligned}$$

This is a system we can easily solve by back-substitution. Plugging in $z = 32$ into the second equation gives us

$$-3y + 128 = -1$$

which tells us $y = 43$. Plugging $z = 32$ and $y = 43$ into the first equation gives us

$$x + 43 - 96 = 4$$

and so $x = 57$.

Since this system is equivalent to our original system, the solution to our original system is

$$\begin{aligned} x &= 57 \\ y &= 43 \\ z &= 32 \end{aligned}$$

1.5 Practice Problems

Solve each of the following systems by applying the three theorems above to replace the system with an equivalent system using the three theorems described above.

Problem 1.1.

$$\begin{aligned} 6x + 2y &= 1 \\ 3x - 4y &= 0 \end{aligned}$$

Problem 1.2.

$$\begin{aligned} x - y &= 3 \\ -2x + 4y &= 2 \end{aligned}$$

Problem 1.3.

$$\begin{aligned} 4x + y &= 5 \\ 8x + 2y &= 10 \\ 12x + 3y &= 15 \end{aligned}$$

Problem 1.4.

$$\begin{aligned} x - y + z &= 3 \\ 2x + 3y - z &= 4 \end{aligned}$$

Problem 1.5.

$$\begin{aligned} x - y + z &= 3 \\ 2x + 3y - z &= 4 \\ -x - 5y + 2z &= -7 \end{aligned}$$

Problem 1.6.

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

Matrices

2

No one can be told what The Matrix is: you have to see it for yourself.

MORPHEUS
The Matrix

Matrices are one efficient way of organizing the information in a system of linear equations, and as we will see later also have a variety of other uses.

2.1 Definitions and Examples

Matrices

A ***matrix*** is a rectangular table of numbers, usually written inbetween parentheses or square brackets. The ***size*** of a matrix is a pair of numbers telling us how many rows and columns the matrix has: if the matrix has m rows and n columns, we say the size of the matrix is $m \times n$, pronounced m *by* n.

Example 2.1.
The following matrices have respective sizes 2×4 and 5×3.

$$\begin{pmatrix} 4 & -7 & 0 & \pi \\ 1.5 & 2 & 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 \\ -1 & 2 & 14 \\ 3 & 7 & 7 \\ -2 & -2 & -2 \\ 5 & 4 & 2 \end{pmatrix}$$

Matrices associated to a linear system

One use of matrices is in encoding a system of linear equations. If we have a system of linear equations, all we really need to know about the system is what the coefficients are, and what the values on the right-hand side of the equations are: what we call the variables (x and y versus x_1 and x_2, for instance) doesn't really matter. If we record all of the coefficients of a system with m equations and n variables as an $m \times n$ matrix, we have the ***coefficient matrix*** of the system.

Example 2.2.
Consider the following system of three equations in four unknowns:

$$\begin{aligned} 6v + 3x - 2y + z &= 4 \\ 4v - x + y - 2z &= 3 \\ 2x + 4y + 4z &= 9 \\ v + z &= -1 \end{aligned}$$

The corresponding coefficient matrix of this system is

$$\begin{pmatrix} 6 & 3 & -2 & 1 \\ 4 & -1 & 1 & -2 \\ 0 & 2 & 4 & 4 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Notice that we pick up zeroes in the coefficient matrix when a variable is missing. The reason for this is that if a variable is missing, such as the missing v in the third equation of the system in Example 2.2, we can write it as 0 times that variable. The system in Example 2.2, for instance, may be written as

$$\begin{aligned} 6v + 3x - 2y + z &= 4 \\ 4v - x + y - 2z &= 3 \\ 0v + 2x + 4y + 4z &= 9 \\ v + 0x + 0y + z &= -1 \end{aligned}$$

Our goal will be to solve systems of linear equations by manipulating matrices. In doing so we of course also want to keep track of the values on the right-hand side of the equations. We will do this by just adding an extra column onto our coefficient matrix containing the right-hand sides. This gives us the ***augmented coefficient matrix*** of the system.

Example 2.3.
The augmented coefficient matrix of Example 2.2 is

$$\begin{pmatrix} 6 & 3 & -2 & 1 & 4 \\ 4 & -1 & 1 & -2 & 3 \\ 0 & 2 & 4 & 4 & 9 \\ 1 & 0 & 0 & 1 & -1 \end{pmatrix}$$

Remark.
Some people like to write a vertical bar in the augmented coefficient matrix to separate the coefficients of the left-hand sides of the equations from the values on the right-hand sides of the equations, such as

$$\left(\begin{array}{cccc|c} 6 & 3 & -2 & 1 & 4 \\ 4 & -1 & 1 & -2 & 3 \\ 0 & 2 & 4 & 4 & 9 \\ 1 & 0 & 0 & 1 & -1 \end{array}\right)$$

The addition of this vertical line is purely cosmetic. Our textbook does not use the bar, but it is fairly common. You are free to use the vertical bar if you'd like.

Elementary row operations

Recall from the last lecture that we had three different procedures that we could perform to a system of linear equations to obtain an equivalent system:

1. Swap two rows.
2. Replace one row with the sum of the original row and a multiple of another row.
3. Multiply every term in a row by the same non-zero constant.

We can perform these same three operations on the augmented coefficient matrix of a linear system and we obtain the augmented coefficient matrix of the equivalent linear system. Usually when we're performing these three operations to a matrix, we refer to them as the ***elementary row operations***. The process of using elementary row operations to turn one matrix into another is called ***row reduction***.

By repeatedly applying the elementary row operations to the augmented coefficient of a matrix, we can replace the turn our system of linear equations into an equivalent one which we can easily solve. In particular, we have a system that is easy to solve when we have put our matrix into *echelon form*. Before define the echelon form of a matrix, let's introduce one preliminary definition that will make the language a little easier.

The ***leading entry*** of a row in a matrix is the left-most non-zero element in that row.

We say that a matrix is in ***echelon form*** if the following three conditions are satisfied:

1. If the matrix has any rows consisting of only zeros, they occur at the bottom of the matrix.
2. The leading entry on each row in the matrix is to the right of the leading entry of the above rows.
3. All entries in the same column and below the leading entry in a row are zero.

Let's first see some examples of some things that are, and some things that are not, in echelon form.

Example 2.4.
The following matrices are all in echelon form:

$$\begin{pmatrix} 3 & -1 & 2 & 0 \\ 0 & 7 & 9 & -1 \\ 0 & 0 & 2 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 6 & 2 & 1 & 4 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

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

$$\begin{pmatrix} 6 & 7 & -4 & 3 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Example 2.5.
None of the following matrices are in echelon form:

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

$$\begin{pmatrix} 7 & 2 & -1 & 2 \\ 5 & 6 & 3 & 3 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 & 7 & 1 & 2 \\ 0 & 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

We can always put a matrix into echelon form by applying elementary row operations. One way to solve a system of linear equations, then, is to write out the augmented coefficient matrix, put it into echelon form, and then use ***back substitution*** (solving for the variables one at a time).

Example 2.6.
Put the following matrix into echelon form:

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{pmatrix}$$

The first thing we need to do is zero-out the entries below 1 in the first column. Let's subtract twice the first row from the second to obtain:

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & 0 & -3 & -6 \\ 4 & 5 & 4 & 2 \end{pmatrix}$$

Now subtract four times the first row from the third:

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & 0 & -3 & -6 \\ 0 & -3 & -12 & -18 \end{pmatrix}$$

Now let's swap the second and third rows,

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & -3 & -12 & -18 \\ 0 & 0 & -3 & -6 \end{pmatrix}$$

Writing out what we're doing in words is always okay to do, but it can get tedious sometimes, so let's introduce some notation to save ourselves some writing. When we perform a row operation on a matrix A to obtain a matrix A', let's draw an arrow from A to A' and label the arrow to describe which operation we are performing.

If we obtain A' from A by swapping row i and row j, we will write

$$A \xrightarrow{R_i \leftrightarrow R_j} A'$$

For example,

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & 0 & -3 & -2 \\ 0 & -3 & -4 & -18 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & -3 & -4 & -18 \\ 0 & 0 & -3 & -2 \end{pmatrix}$$

If we add c times row j to row i, we will write

$$A \xrightarrow{R_i + cR_j \rightarrow R_i} A'$$

E.g.,

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{pmatrix} \xrightarrow{R_2 - 2R_1} \begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & 0 & -3 & -2 \\ 4 & 5 & 4 & 2 \end{pmatrix}$$

If we multiply each element in row i by c, we will write

$$A \xrightarrow{cR_i \rightarrow R_i} A'.$$

For example,

$$\begin{pmatrix} 4 & 7 & 2 \\ 0 & 4 & 3 \\ -1 & 2 & 2 \\ 5 & -2 & 1 \end{pmatrix} \xrightarrow{-2R_3 \rightarrow R_3} \begin{pmatrix} 4 & 7 & 2 \\ 0 & 4 & 3 \\ 2 & -4 & -4 \\ 5 & -2 & 1 \end{pmatrix}$$

Notice that the echelon form of a matrix is not unique: if you give the same matrix to two people and ask them to put the matrix in echelon form, each person may give you back a different (but correct!) matrix in echelon form. The matrix in Example 2.6, for instance, could be put into echelon form in the following way:

Example 2.7.
Put the following matrix into echelon form:

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{pmatrix} \xrightarrow{R_3 - 2R_2 \to R_3} \begin{pmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 0 & -3 & -6 & -6 \end{pmatrix}$$

$$\xrightarrow{R_2 - 2R_1 \to R_2} \begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & 0 & -3 & -6 \\ 0 & -3 & -6 & -6 \end{pmatrix}$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & -3 & -6 & -6 \\ 0 & 0 & -3 & -6 \end{pmatrix}$$

In Example 2.6 and Example 2.7 we started with the same matrix, but produced two different matrices in echelon form because we performed two different sequences of elementary row operations.

Remark.
The above is something to consider if you compare answers to homework problems with another student. If you were both trying to put a matrix into echelon form, you may both come up with different, correct answers!

It would be nice if there was a way to modify echelon form so that we would always calculate the same matrix. This can be done if we modify the conditions of echelon form slightly to get *row-reduced echelon form*.

We say a matrix A is in ***row-reduced echelon form*** (abbreviated ***RREF***) if the following four conditions are satsified:

1. If the matrix has any rows consisting of only zeros, they occur at the bottom of the matrix.

2. The leading entry on each row in the matrix is to the right of the leading entry of the above rows.

3. All entries in the same column above and below the leading entry in a row are zero.

4. Every leading entry is a one.

So RREF is very similar to echelon form, but we'll make sure that leading entries are always equal to one, and that everything directly *above and below* a leading entry is zero.

Remark.
Some people simply say *reduced echelon form* where we have said *row-reduced echelon form*, but this is the same thing.

Let's take our two matrices in echelon form from Example 2.6 and Example 2.7 and convert them to RREF.

Example 2.8.
Convert the following matrix to RREF:

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & -3 & -12 & -18 \\ 0 & 0 & -3 & -6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & -3 & -12 & -18 \\ 0 & 0 & -3 & -6 \end{pmatrix} \xrightarrow{\frac{-1}{3}R_2 \to R_2} \begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & -3 & -6 \end{pmatrix}$$

$$\xrightarrow{R_1 - 2R_2 \to R_1} \begin{pmatrix} 1 & 0 & -4 & -7 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & -3 & -6 \end{pmatrix}$$

$$\xrightarrow{-\frac{1}{3}R_3 \to R_3} \begin{pmatrix} 1 & 0 & -4 & -7 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

$$\xrightarrow{R_2 - 4R_3 \to R_2} \begin{pmatrix} 1 & 0 & -4 & -7 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

$$\xrightarrow{R_1 + 4R_3 \to R_1} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

Example 2.9.
Convert the following matrix to RREF:

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & -3 & -6 & -6 \\ 0 & 0 & -3 & -6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & -3 & -6 & -6 \\ 0 & 0 & -3 & -6 \end{pmatrix} \xrightarrow{-\frac{1}{3}R_2 \to R_2} \begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & -3 & -6 \end{pmatrix}$$

$$\xrightarrow{-\frac{1}{3}R_3 \to R_3} \begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

$$\xrightarrow{R_1 - 2R_2 \to R_1} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

$$\xrightarrow{R_2 - 2R_3 \to R_2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

The main thing about main reason we prefer RREF over echelon form is that every matrix is equivalent to exactly one matrix in RREF.

Theorem 2.1.
Performing elementary row operations to put a matrix in row-reduced echelon form produces exactly one matrix.

The proof of this fact will be easier to explain after we talk about linear independence, so we will postpone the proof of this theorem for now.

Example 2.10.
Solve the following system of linear equations by putting the augmented coefficient matrix in RREF.

$$\begin{aligned} x + 2y + 4z &= 5 \\ 2x + 4y + 5z &= 4 \\ 4x + 5y + 4z &= 2 \end{aligned}$$

We have seen that the augmented coefficient matrix of this system can be put into the following matrix in RREF:

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

which is the augmented coefficient matrix of the equivalent system

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

and so the only solution to our system of equations is $(x, y, z) = (1, -2, 2)$.

We are now in a position to describe an algorithm for putting a matrix in RREF, but before presenting the algorithm we introduce one piece of terminology.

If a matrix is in RREF, then the location of the leading entries are called the ***pivot positions***; the columns containing pivot positions are called ***pivot columns***. More generally, if A can be reduced to a matrix A' in RREF, then the pivot positions and columns of A are defined to be the pivot positions and columns of A'.

By performing elementary row operations, we can always put a non-zero value in a pivot position. When we do this, the non-zero value we place in the pivot position is called a ***pivot***.

The algorithm for putting a matrix into RREF is as follows:

1. Starting from the top row of the matrix, the left-most position which *is not* in a column of all zeros will be a pivot position.

2. Swap rows if necessary so that the entry in the pivot position is non-zero.

3. Divide the row containing this pivot position by the new non-zero pivot value.

4. Add multiples of the row to the other rows so that we have only zeros above and below the pivot in the pivot column.

5. Repeat the process, but use the submatrix obtained by deleting everything to the left of and above the pivot position (including the

row and column containing the pivot position) to determine the next pivot position.

Example 2.11.
Put the following matrix in RREF:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 3 & 7 \\ 0 & 0 & 2 & 0 & 1 & 4 \\ 0 & 0 & 4 & 0 & 2 & 1 \\ 0 & 5 & 3 & 0 & 4 & 2 \end{pmatrix}$$

We need to work with our pivots one at a time, from the top left-most pivot down to the bottom right-most pivot. We will color code which pivot we are considering as follows: The pivot we are currently considering will be yellow, and the pivots we have finished working with will be pink.

We start with the top row. First finding the left-most entry which is not in an all-zero column.

$$\begin{pmatrix} 0 & \boxed{0} & 0 & 0 & 3 & 7 \\ 0 & 0 & 2 & 0 & 1 & 4 \\ 0 & 0 & 4 & 0 & 2 & 1 \\ 0 & 5 & 3 & 0 & 4 & 2 \end{pmatrix}$$

Now swap the top and bottom columns to put a 5 in the pivot position:

$$\begin{pmatrix} 0 & \boxed{5} & 3 & 0 & 4 & 2 \\ 0 & 0 & 2 & 0 & 1 & 4 \\ 0 & 0 & 4 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 & 7 \end{pmatrix}$$

Divide the first row by 5 to put a 1 into the pivot position:

$$\begin{pmatrix} 0 & \boxed{1} & \frac{3}{5} & 0 & \frac{4}{5} & \frac{2}{5} \\ 0 & 0 & 2 & 0 & 1 & 4 \\ 0 & 0 & 4 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 & 7 \end{pmatrix}$$

The matrix already has zeros below the pivot, so we move onto the pivot.

First find the left-most entry in the second row which is to the right of the previous pivot and *not* in a column of all zeros.

$$\begin{pmatrix} 0 & \boxed{1} & \frac{3}{5} & 0 & \frac{4}{5} & \frac{2}{5} \\ 0 & 0 & \boxed{2} & 0 & 1 & 4 \\ 0 & 0 & 4 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 & 7 \end{pmatrix}$$

Now divide the second row by 2 to put a 1 into the next pivot position:

$$\begin{pmatrix} 0 & \boxed{1} & \frac{3}{5} & 0 & \frac{4}{5} & \frac{2}{5} \\ 0 & 0 & \boxed{1} & 0 & \frac{1}{2} & 2 \\ 0 & 0 & 4 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 & 7 \end{pmatrix}$$

We need to zero out the non-zero entries in the pivot column above and below our pivot position. First we subtract $3/5$ the second row from the first:

$$\begin{pmatrix} 0 & \boxed{1} & 0 & 0 & \frac{1}{2} & -\frac{4}{5} \\ 0 & 0 & \boxed{1} & 0 & \frac{1}{2} & 2 \\ 0 & 0 & 4 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 & 7 \end{pmatrix}$$

Now subtract four times the second row from the third row:

$$\begin{pmatrix} 0 & \boxed{1} & 0 & 0 & \frac{1}{2} & -\frac{4}{5} \\ 0 & 0 & \boxed{1} & 0 & \frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 & -7 \\ 0 & 0 & 0 & 0 & 3 & 7 \end{pmatrix}$$

We move on to the third pivot. Since we have a zero in the pivot position, we need to swap the third and fourth rows to put the 3 into the pivot position:

$$\begin{pmatrix} 0 & \boxed{1} & 0 & 0 & \frac{1}{2} & -\frac{4}{5} \\ 0 & 0 & \boxed{1} & 0 & \frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & \boxed{3} & 7 \\ 0 & 0 & 0 & 0 & 0 & -7 \end{pmatrix}$$

Divide the third row by 3 to put a 1 into the pivot position:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \frac{1}{2} & -\frac{4}{5} \\ 0 & 0 & 1 & 0 & \frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & 1 & \frac{7}{3} \\ 0 & 0 & 0 & 0 & 0 & -7 \end{pmatrix}$$

Now zero out the entries above the third pivot. First subtract one-half the third row from the first:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & -\frac{59}{30} \\ 0 & 0 & 1 & 0 & \frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & 1 & \frac{7}{3} \\ 0 & 0 & 0 & 0 & 0 & -7 \end{pmatrix}$$

Finally subtract one-half the third row from the second:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & -\frac{59}{30} \\ 0 & 0 & 1 & 0 & 0 & \frac{5}{6} \\ 0 & 0 & 0 & 0 & 1 & \frac{7}{3} \\ 0 & 0 & 0 & 0 & 0 & -7 \end{pmatrix}$$

For the very last pivot we will simply divide by -7 to make the pivot a 1, and then zero out the entries above the pivot. Finally subtract one-half the third row from the second:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & -\frac{59}{30} \\ 0 & 0 & 1 & 0 & 0 & \frac{5}{6} \\ 0 & 0 & 0 & 0 & 1 & \frac{7}{3} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We now have the RREF of our original matrix:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Exercise 2.1.
Determine the number of pivot columns of the following matrix:

$$\begin{pmatrix} 3 & 4 & 0 & 2 \\ 4 & 0 & 0 & 1 \\ -1 & 0 & 0 & 2 \\ 6 & 0 & 0 & 7 \end{pmatrix}$$

Once we put an augmented coefficient matrix into RREF, it is then *very* easy to solve the corresponding system of linear equations.

Example 2.12.
Determine all solutions to the following system of equations:

$$\begin{aligned} 3x_5 &= 7 \\ 2x_3 + x_5 &= 4 \\ 4x_3 + 2x_5 &= 1 \\ 5x_2 + 3x_3 + 4x_5 &= 2 \end{aligned}$$

The augmented coefficient matrix of this system is precisely our matrix from before, so the RREF of that matrix tells us the following system is equivalent:

$$\begin{aligned} x_2 &= 0 \\ x_3 &= 0 \\ x_5 &= 0 \\ 0 &= 1 \end{aligned}$$

Because of this last equation, the system has no solutions!

Example 2.13.

Solve the following system of linear equations.

$$\begin{aligned} 3x + 7y + 17z &= 21 \\ 2y + 4z &= 6 \\ 4x + 10y + 24z &= 30 \end{aligned}$$

To solve the system, let's put the augmented coefficient matrix, which is

$$\left(\begin{array}{ccc|c} 3 & 7 & 17 & 21 \\ 0 & 2 & 4 & 6 \\ 4 & 10 & 24 & 30 \end{array}\right),$$

into RREF:

$$\begin{aligned}
\begin{pmatrix} 3 & 7 & 17 & 21 \\ 0 & 2 & 4 & 6 \\ 4 & 10 & 24 & 30 \end{pmatrix} &\xrightarrow{\frac{1}{3}R_1 \to R_1} \begin{pmatrix} 1 & \frac{7}{3} & \frac{17}{3} & 7 \\ 0 & 2 & 4 & 6 \\ 4 & 10 & 24 & 30 \end{pmatrix} \\
&\xrightarrow{R_3 - 4R_1 \to R_3} \begin{pmatrix} 1 & \frac{7}{3} & \frac{17}{3} & 7 \\ 0 & 2 & 4 & 6 \\ 0 & \frac{2}{3} & \frac{4}{3} & 2 \end{pmatrix} \\
&\xrightarrow{\frac{1}{2}R_2 \to R_2} \begin{pmatrix} 1 & \frac{7}{3} & \frac{17}{3} & 7 \\ 0 & 1 & 2 & 3 \\ 0 & \frac{2}{3} & \frac{4}{3} & 2 \end{pmatrix} \\
&\xrightarrow{R_1 - \frac{7}{3}R_2 \to R_1} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & \frac{2}{3} & \frac{4}{3} & 2 \end{pmatrix} \\
&\xrightarrow{R_3 - \frac{2}{3}R_2 \to R_3} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

The last matrix above is in RREF. This tells us that our original system of equations is equivalent to

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

Notice that we could solve both of these equations to express x and y in terms of z:

$$x = -z$$
$$y = 3 - 2z$$

That is, if we know what z is, then we instantly know what x and y must be. However, z could be anything! There are infinitely-many different choices for z, and each one gives us a different solution to our system. Thus there are infinitely-many solutions to the system, all of which have the form

$$(-z, 3 - 2z, z)$$

where z can be any real number.

Notice that in Example 2.13, not only did we determine that there were infinitely-many solutions to the system, but we said explicitly what the solutions had to look like. In such a situation, where we describe all of the solutions in terms of some variable, we say that we have given a ***parametrization*** of the solution set and call the variable that is allowed to change a ***free variable***.

Example 2.14.
Solve the following system of linear equations:

$$\begin{aligned} 5x_1 + 3x_2 - 8x_3 - 2x_4 &= 5 \\ 2x_1 + 4x_2 - 6x_3 + 2x_4 &= 2 \\ 2x_1 + 1x_2 - 3x_3 - x_4 &= 2 \\ 4x_1 + 3x_2 - 7x_3 - x_4 &= 4 \end{aligned}$$

We again put the augmented coefficient matrix into RREF to get an equivalent, easier-to-solve system:

$$\begin{pmatrix} 5 & 3 & -8 & -2 & 5 \\ 2 & 4 & -6 & 2 & 2 \\ 2 & 1 & -3 & -1 & 2 \\ 4 & 3 & -7 & -1 & 4 \end{pmatrix} \xrightarrow{\frac{1}{5}R_1 \to R_1} \begin{pmatrix} 1 & 3/5 & -8/5 & -2/5 & 1 \\ 2 & 4 & -6 & 2 & 2 \\ 2 & 1 & -3 & -1 & 2 \\ 4 & 3 & -7 & -1 & 4 \end{pmatrix}$$

$$\xrightarrow{R_2 - 2R_1 \to R_2} \begin{pmatrix} 1 & 3/5 & -8/5 & -2/5 & 1 \\ 0 & 14/5 & -14/5 & 14/5 & 0 \\ 2 & 1 & -3 & -1 & 2 \\ 4 & 3 & -7 & -1 & 4 \end{pmatrix}$$

$$\xrightarrow{R_3 - 2R_1 \to R_3} \begin{pmatrix} 1 & 3/5 & -8/5 & -2/5 & 1 \\ 0 & 14/5 & -14/5 & 14/5 & 0 \\ 0 & -1/5 & 1/5 & -1/5 & 0 \\ 4 & 3 & -7 & -1 & 4 \end{pmatrix}$$

$$\xrightarrow{R_4 - 4R_1 \to R_4} \begin{pmatrix} 1 & 3/5 & -8/5 & -2/5 & 1 \\ 0 & 14/5 & -14/5 & 14/5 & 0 \\ 0 & -1/5 & 1/5 & -1/5 & 0 \\ 0 & 3/5 & -3/5 & 3/5 & 0 \end{pmatrix}$$

$$\xrightarrow{\frac{5}{14}R_2 \to R_2} \begin{pmatrix} 1 & 3/5 & -8/5 & -2/5 & 1 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & -1/5 & 1/5 & -1/5 & 0 \\ 0 & 3/5 & -3/5 & 3/5 & 0 \end{pmatrix}$$

$$\xrightarrow{R_1 - \frac{3}{5}R_2 \to R_1} \begin{pmatrix} 1 & 0 & -1 & -1 & 1 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & -1/5 & 1/5 & -1/5 & 0 \\ 0 & 3/5 & -3/5 & 3/5 & 0 \end{pmatrix}$$

$$\xrightarrow{R_3 + \frac{1}{5}R_2 \to R_3} \begin{pmatrix} 1 & 0 & -1 & -1 & 1 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 3/5 & -3/5 & 3/5 & 0 \end{pmatrix}$$

$$\xrightarrow{R_4 - \frac{3}{5}R_2 \to R_4} \begin{pmatrix} 1 & 0 & -1 & -1 & 1 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

We now know that our original system of equations is equivalent to the following system:

$$\begin{aligned} x_1 - x_3 - x_4 &= 1 \\ x_2 - x_3 + x_4 &= 0 \end{aligned}$$

Solving the first two equations for x_1 and x_2, respectively, tells us that

$$\begin{aligned} x_1 &= 1 + x_3 + x_4 \\ x_2 &= x_3 - x_4 \end{aligned}$$

Here there are no restrictions on either x_3 or x_4, each of these can be any real number, and so we have two free variables. A parametrization of the solution set is

$$(1 + x_3 + x_4, x_3 - x_4, x_3, x_4).$$

2.2 Consistency and Inconsistency in Terms of RREF

After we put the augmented coefficient matrix of a system into RREF, we can quickly determine whether the system is consistent or not, and if it is consistent whether it has a unique solution or infinitely-many solutions. If the matrix in RREF has a row of the form

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \mid b \end{pmatrix}$$

where $b \neq 0$, then the system is inconsistent. The existence of such a row tells us that the system is equivalent to a system that has an equation of the form

$$0x_1 + 0x_2 + 0x_3 + \cdots + 0x_{n-1} + 0x_n = b$$

It doesn't matter what the other equations (or rows in the matrix) look like: there is no way to solve this one, so the system has no solutions.

If the matrix in RREF has the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & b_1 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & b_2 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & b_3 \\ & & & & \ddots & & & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & b_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & b_n \end{pmatrix}$$

then the system has a single unique solution, $(b_1, b_2, ..., b_n)$.

Notice that if we took the previous matrix in RREF and added rows of zeros to the bottom, then this doesn't change the solutions. When this happens some of the equations in the original system were "redundant."

Example 2.15.
Consider the following system:

$$\begin{aligned} x + y &= 1 \\ x - y &= 2 \\ 3x + y &= 4 \end{aligned}$$

The augmented coefficient matrix of the system is

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

The RREF of this matrix is

$$\begin{pmatrix} 1 & 0 & 3/2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \end{pmatrix}$$

The system thus has a unique solution of $(3/2, -1/2)$. The third equation of the original system doesn't give us any additional information about the system because it is twice the first equation plus the second: once you know what the solution to the two equations are, you instantly know what the solution to the third equation is as well, so there's no real need to have the third equation.

If the matrix in RREF has rows of zeros, but there is not a unique solution to the system, then the system has infinitely-many solutions. The number of free variables is determined by the number of rows of zeros. Each non-zero row gives us an equation relating the variables, and each zero row (up to the number of variables) gives us a free variable. Notice that the number of variables is one less than the number of columns in the augmented coefficient matrix: each variable gives us one column, plus we have one more column for the right-hand sides. This means that the number of free variables in our solution to a system is determined by the number of non-pivot columns (ignoring the right-most column corresponding to the right-hand sides of equations in our system).

In Example 2.13 we had one non-pivot column (ignoring the column corresponding to the RHS) and so one free variable; in Example 2.14 we had two non-pivot columns, so two free variables.

Exercise 2.2.
Determine the set of solutions to the following system by putting the augmented coefficient matrix into RREF. How many free variables are there?

$$\begin{aligned} 5x_1 + 3x_2 - 8x_3 - 2x_4 &= 5 \\ 2x_1 + 4x_2 - 6x_3 + 2x_4 &= 2 \\ 2x_1 + 1x_2 - 3x_3 - x_4 &= 2 \\ 4x_1 + 3x_2 - 7x_3 - x_4 &= 4 \\ -6x_1 - 3x_2 + 9x_3 + 3x_4 &= -6 \end{aligned}$$

2.3 Practice Problems

Problem 2.1.
Put the following matrices into echelon form, but not RREF.

(a)

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

(b)

$$\begin{pmatrix} 0 & 2 & -1 & 7 & 2 \\ 3 & 4 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 & -3 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Problem 2.2.
Put the following matrices in RREF:

(a)

$$\begin{pmatrix} -4 & -4 & -8 & -12 \\ 2 & 3 & 6 & 7 \\ 3 & 4 & 9 & 10 \end{pmatrix}$$

(b)

$$\begin{pmatrix} 2 & 3 & 7 & 11 \\ 1 & 1 & 3 & 4 \end{pmatrix}$$

(c)

$$\begin{pmatrix} 6 & 0 & 4 & 14 & 24 \\ 1 & 0 & 4 & 9 & 14 \\ 4 & 0 & 6 & 16 & 26 \\ 1 & 0 & 2 & 5 & 8 \end{pmatrix}$$

Problem 2.3.
Solve the following system of equations by writing the augmented coefficient matrix in RREF. If there are infinitely-many solutions, give a parametrization of the solution set.

(a)

$$\begin{aligned} 3y + 6z &= 12 \\ x + 2y + 3z &= \frac{19}{2} \\ 2x + 2y + 6z &= 13 \end{aligned}$$

(b)

$$\begin{aligned} 6x + 7y + 19z &= 18 \\ 3x + 2y + 8z &= 9 \\ x + 2y + 4z &= 3 \\ 5x + 6y + 16z &= 15 \end{aligned}$$

(c)

$$\begin{aligned} 3x + 5y + 11z &= 1 \\ 2x + 4y + 8z &= 1 \\ 4x + 7y + 15z &= 2 \end{aligned}$$

3 Vectors

Algebra is the offer made by the devil to the mathematician.

SIR MICHAEL ATIYAH

Vectors appear in many different areas of mathematics and sciences, and you have seen vectors before if you've taken a course in physics of multivariable calculus. In these classes vectors are usually described as quantities that have both a magnitude and a direction. We will see later in the course that many different types of quantities can be thought of as vectors, even things that don't obviously have a magnitude or direction (for exmaple, polynomials can be thought of as vectors). In this lecture we start with the basics though, first defining vectors as "arrows" in $\mathbb{R}^2$ and $\mathbb{R}^3$, and then generalizing vectors to $\mathbb{R}^n$ for any dimension n. We also see that some of the questions we are naturally lead to about vectors are really questions about systems of linear equations in disguise.

3.1 Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

We will see that vectors can be defined in any number of dimensions, but to get started we will consider vectors in two and three dimensions which will be familiar to anyone that has taken an introductory course in physics or multivariable calculus.

A vector is often described as a quantity which has both a direction and a magnitude. One physical example is force: every force has a direction (where the force is pushing from or pulling towards) and a magnitude (how strong the force is). Consider the gravitational force between the Earth and an object near the surface of the Earth. The object is being pulled "down" towards the Earth (this is the direction) and that object has some weight (this is the magnitude). If the object has more mass, then it will have a greater weight and gravity is pulling more strongly on the object. (Consider trying to hold a one kilogram object over your head versus a fifteen kilogram object. Gravity is pulling harder on the 15 kg object which is why that object feels heavier and harder to hold up.)

In two or three dimensions we represent these "direction together with a magnitude" quantities as arrows where the direction of the arrow is the direction of the vector, and the length of the vector represents the magnitude. We will focus on the two-dimensional case at first simply because it's slightly easier to draw pictures representing these quantities.

Two-dimensional vectors

A ***vector*** in $\mathbb{R}^2$ is simply an arrow in the plane: a line segment from some point (x_0, y_0) to another point (x_1, y_1) with an arrowhead at (x_1, y_1), as in Figure 3.1

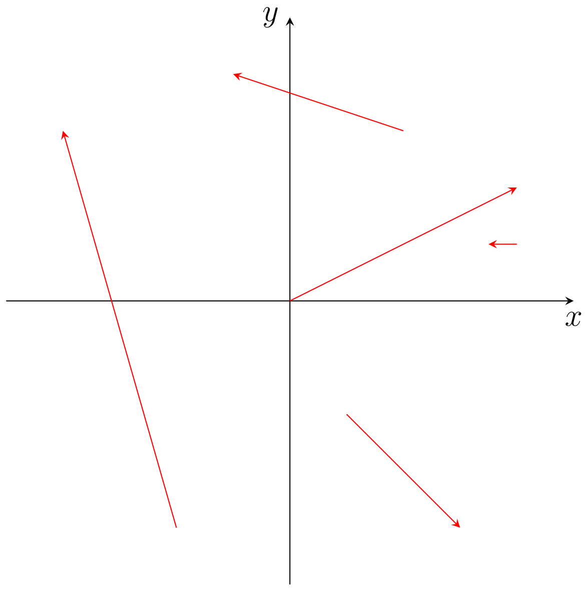


Figure 3.1: Vectors in $\mathbb{R}^2$.

One thing about vectors that may seem a little strange is that we *only* care about the direction the vector points in and its magnitude, and we *do not* care about where that vector is drawn in space. That is, given any vector we can move it around the plane or 3-space as much as we'd like and provided we don't stretch the vector (which would change its magnitude) or rotate it (which would change its direction) we still have the same vector. See Figure 3.2.

We usually give vectors a name, just like any other mathematical quantity, to make it easier to describe. Instead of just calling a vector v, however, it is common to write the letter in bold, $\mathbf{v}$, or to write an arrow over the letter, $\vec{v}$, to denote that this quantity is a vector. Almost everyone writes the arrow when they are writing vectors by hand on paper or a blackboard, and only some books (including our textbook) use the bold letters to denote vectors.

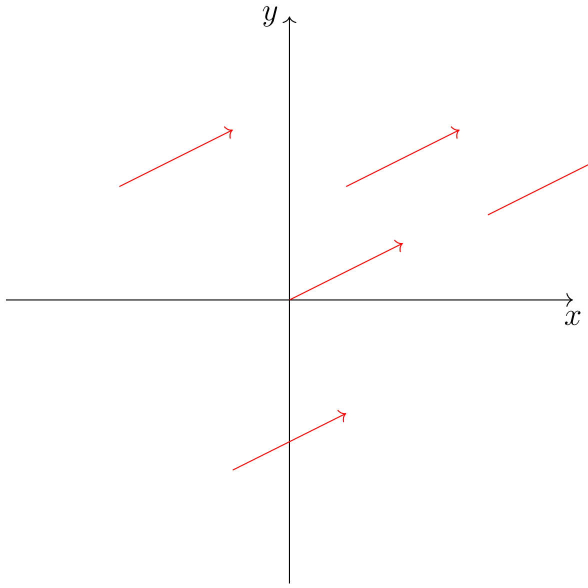


Figure 3.2: Each arrow in this picture represents the same vector since the direction and magnitude (length) is the same for each arrow.

We will sometimes call the beginning point of the arrow represent a vector (the part without an arrowhead) the ***tail*** of the vector, and the other point (the part with an arrowhead) the ***tip*** of the vector.

Given two vectors, $\vec{v}_1$ and $\vec{v}_2$, we can add the two vectors together to get a new vector, $\vec{v}_1 + \vec{v}_2$. There are a few ways we can describe this new vector. One way to describe $\vec{v}_1 + \vec{v}_2$ is to slide $\vec{v}_1$ and $\vec{v}_2$ around so that the tip of $\vec{v}_1$ is at the tail of $\vec{v}_2$. We then draw in an arrow from the tail of $\vec{v}_1$ to the tip of $\vec{v}_2$, and this new vector we've drawn is $\vec{v}_1 + \vec{v}_2$. This is called the ***triangle law*** for vector addition and is illustrated in Figure 3.3.

Another, equivalent way to describe vector addition is with the ***parallelogram law***. Here we make copies of $\vec{v}_1$ and $\vec{v}_2$ and slide them around to make a parallelogram, and then draw in the diagonal of this parallelogram connecting the corner that has two tails to the corner that has two tips. This diagonal vector is the sum $\vec{v}_1 + \vec{v}_2$. See Figure 3.4

Remark.
It is important that the diagonal vector drawn in the parallelogram law starts at the corner where two tails meet and goes to the corner where two tips meet. If you connect the corners in the wrong way,

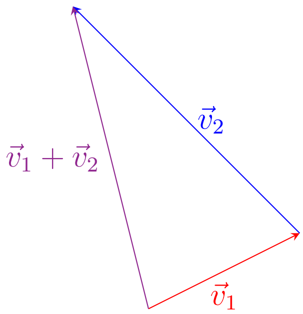


Figure 3.3: We can add two vectors by sliding them *tip-to-tail* and then completing the triangle.

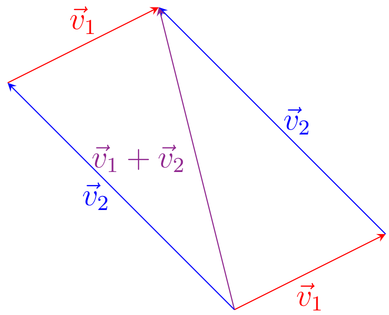


Figure 3.4: Vector addition can also be described by drawing in the diagonal of a parallelogram.

you will have the wrong vector.

If you've never seen vector addition before, it may seem a little bit like an odd thing to do, but let's consider one physical application. Suppose that two different forces act on an object: e.g., you and a friend are rearranging furniture in your dorm with one of you pushing a bed in one direction, and the other pushing the bed in another direction. Say you push the bed to the East and your friend pushes the bed to the North. Though you're applying two different forces, the net effect is the same as if you were to push the bed to the North-East. This is exactly what's happening when you add the two forces: the sum of two individual forces acting on an object is the ***net force*** that acts on that object.

There's another operation we can perform on vectors. Given a real number λ and a vector $\vec{v}$, we can define a new vector $\lambda\vec{v}$ by stretching the vector out by a factor of λ. For example, $2\vec{v}$ points in the same direction as $\vec{v}$ but is twice as long; $\frac{1}{3}\vec{v}$ points in the same direction as $\vec{v}$ but is one third as long.

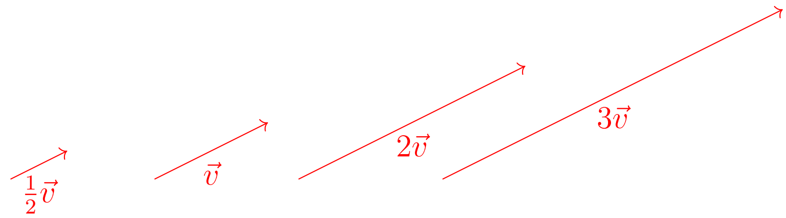


Figure 3.5: Multiplying a vector by a positive number stretches the vector.

If we multiply a vector $\vec{v}$ by a negative λ, then we stretch the vector out by a factor of $|\lambda|$, but make the vector point in the opposite direction.

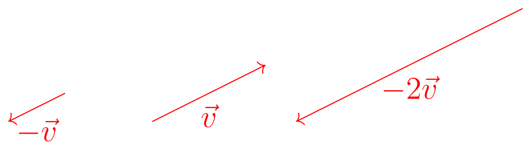


Figure 3.6: Multiplying a vector by a negative number stretches the vector and flips the vector's direction.

To distinguish "regular" numbers from vectors we sometimes call the numbers ***scalars*** because they scale vectors. The operation of multiplying a scalar and a vector is called ***scalar multiplication***.

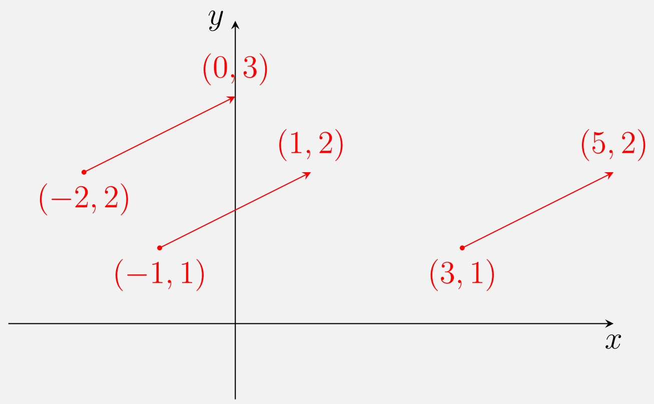
Notice that we could describe a vector in the plane by saying how much the x and y coordinates change when you walk from the tail of the vector to the tip. That is, if we've positioned the vector so that its tail is at the point (x_0, y_0) and its tip is at the point (x_1, y_1), then all we really need to know is the change in the x-coordinates, $x_1 - x_0$, and the change in the y-coordinates, $y_1 - y_0$. Regardless of where we've drawn the vector, if we don't stretch it or rotate it, we have the same change in x- and y-coordinates.

Example 3.1.
Suppose $\vec{v}$ is a vector which we've drawn in the plane so that its tail is at the point $(3, 1)$ and its tip is at the point $(5, 2)$. The change in

the x-coordinates is $5 - 3 = 2$ and the change in the y-coordinates is $2 - 1 = 1$.

If we were to have moved the vector around so that its tail was instead at $(-2, 2)$, then its tip would be at $(0, 3)$. Again we have the change in the x-coordinates is $0 - (-2) = 2$ and the change in y-coordinates is $3 - 2 = 1$.

If we placed the tail of the vector at $(-1, 1)$, then the tip would be at $(1, 2)$, and once again the change in x- and y-coordinates is 2 and 1, respectively.



So we could describe the vector simply by recording this change in x- and y-coordinates. There are several different ways we could record this, but two common ways would be to make a 2×1 matrix listing the change in x- and y-, or a 1×2 matrix:

$$\begin{pmatrix} x \\ y \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} x & y \end{pmatrix}$$

In the first situation, with the 2×1 matrix we say we have a ***column vector***, and the 1×2 matrix is called a ***row vector***. We will usually, but not always, use column vectors in this class. There's nothing magical about why we choose column vectors instead of row vectors, it's just a choice.

Notice that vector addition and scalar multiplication are very easy to express once we have coordinates like this:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}$$

$$\lambda \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}$$

Three-dimensional vectors

Vectors in $\mathbb{R}^3$ are completely analogous to vectors in $\mathbb{R}^2$: we they are arrows connecting a point at the tail of the vector to the tip, the direction of the vector is the direction of this arrow, and the length of the arrows is the magnitude. Adding three-dimensional vectors or doing scalar multiplication is exactly the same as adding two-dimensional vectors or doing scalar multiplication: we can use the parallelogram or triangle laws and stretch a vector out by a given amount. It's slightly harder to draw the pictures on a two-dimensional screen or piece of paper, but everything is defined exactly the same.

Just as we can represent a two-dimensional vector using two pieces of information, telling us the displacement in the x- and y-coordinates between the vector's tail and tip, we can do precisely the same thing in three dimensions and we simply have one more piece of information to deal with: the change in the z-coordinates.

Just as in the two-dimensional vectors can be represented as column or row vectors, so can three-dimensional vectors: A 3×1 matrix is a column vector in three dimensions, and a 1×3 matrix is a row vector in three dimensions.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} x & y & z \end{pmatrix}$$

and again, vector addition and scalar multiplication are easily expressed:

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix}$$

$$\lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \\ \lambda z \end{pmatrix}$$

3.2 Vectors in $\mathbb{R}^n$

We can define vectors in $\mathbb{R}^n$ completely analogously to how we defined row vectors (or column) vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$. If we're in $\mathbb{R}^n$ for $n \geq 4$ we can't really visualize the vectors as arrows anymore, but we can still define them algebraically.

A ***column vector*** in $\mathbb{R}^n$ is a $n \times 1$ matrix, and a ***row vector*** is a $1 \times n$ matrix. We will typically use column vectors and just say ***vector*** to mean column vectors – but this is just a convenient choice.

Given two vectors in $\mathbb{R}^n$, say

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

then we define their sum, $\vec{x} + \vec{y}$ by adding the components of the vectors together,

$$\vec{x} + \vec{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

We define scalar multiplication by multiplying each component of the vector by the same scalar,

$$\lambda \vec{x} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}$$

We will usually denote the vector $-1 \cdot \vec{v}$ as just $-\vec{v}$, and define "vector subtraction" to be vector addition, but with -1 times one of the vectors:

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

By $\vec{0}$ we always mean the vector of all zeros:

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Two vectors, $\vec{v}$ and $\vec{u}$, are equal precisely when their components are equal:

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

really means

$$\begin{aligned} v_1 &= u_1 \\ v_2 &= u_2 \\ &\vdots \\ v_n &= u_n \end{aligned}$$

3.3 Complex Vectors

Just as a real vector is an element of $\mathbb{R}^n$ – an n-tuple of real numbers – a ***complex vector*** is an element of $\mathbb{C}^n$, an n-tuple of complex numbers. Algebraically, vector addition and scalar multiplication are exactly like for vectors in $\mathbb{R}^n$, except we do allow complex numbers to be scalars.

If $(z_1, ..., z_n)$ and $(\zeta_1, ..., \zeta_n)$ are elements of $\mathbb{C}^n$, we define the vector addition of these two vectors to be

$$(z_1, ..., z_n) + (\zeta_1, ..., \zeta_n) = (z_1 + \zeta_1, ..., z_n + \zeta_n).$$

Given any complex number $\lambda \in \mathbb{C}$, we define scalar multiplication between λ and a vector $(z_1, ..., z_n) \in \mathbb{C}^n$ to be

$$\lambda \cdot (z_1, ..., z_n) = (\lambda z_1, ..., \lambda z_n).$$

For example, in $\mathbb{C}^3$ we have

$$(2 + 3i, i, 4) + (1 - 3i, 2, 1 + i) = (3, 2 + i, 5 + i)$$

and

$$i \cdot (2 + 3i, i, 4) = (-3 + 2i, -1, 4i).$$

We can't draw arrows for complex vectors except in the case of $\mathbb{C}^1$, but this is completely identical to the case of $\mathbb{R}^2$, with one important exception. If λ is a real number and $z = x + iy \in \mathbb{C} = \mathbb{C}^1$ is a complex number (thought of as an arrow from 0 to $x + iy$), then λz stretches the arrow out in exactly the same way as $\lambda(x, y)$ gets stretched out in $\mathbb{R}^2$. If λ is a complex number though, then multiplying by λ not only stretches the vector, but it can rotate it as well!.

Every complex number λ can be written in the form $re^{i\theta}$ where r and θ are real numbers, and we may assume $r > 0$. In real and imaginary components we may write this number as

$$re^{i\theta} = r\cos(\theta) + ir\sin(\theta).$$

It's also easy to write a number $a + ib$ in this form: $r = \sqrt{a^2 + b^2}$ and $\theta = \tan^{-1}(y/x)$.

When we multiply $z \in \mathbb{C}$ by $\lambda = re^{i\theta}$, the corresponding arrow gets stretched out by a factor of r, but rotated counter-clockwise by θ.

3.4 Properties of Vectors

Vector addition and scalar multiplication satisfy some of the basic algebraic properties that you would expect:

Proposition 3.1.
Let $\vec{u}$, $\vec{v}$, and $\vec{w}$ all be vectors in $\mathbb{R}^n$ (resp., $\mathbb{C}^n$) and let λ, μ be scalars in $\mathbb{R}$ (resp. $\mathbb{C}$). Then vector addition and scalar multiplication satisfy the following properties:

1. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$
2. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$
3. $\lambda(\mu\vec{v}) = (\lambda\mu)\vec{v}$
4. $\vec{v} + \vec{0} = \vec{v}$
5. $\vec{v} + (-\vec{v}) = \vec{0}$
6. $(\lambda + \mu)\vec{v} = \lambda\vec{v} + \mu\vec{v}$
7. $\lambda(\vec{u} + \vec{v}) = \lambda\vec{u} + \lambda\vec{v}$
8. $\lambda(\mu\vec{v}) = (\lambda\mu)\vec{v}$

9. $1 \cdot \vec{v} = \vec{v}$

Exercise 3.1.
Verify each of the properties in the previous proposition.

3.5 Vector Equations

Just we have equations involving numbers, we can have equations involving scalars. For example, consider the vectors

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \vec{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Now consider the equation

$$x\vec{v}_1 + y\vec{v}_2 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$

A solution to this vector equation would be a choice of scalars x and y making the equation hold.

Let's think about what would happen if we write out all of the details in the above vector equation

$$\begin{aligned} x\vec{v}_1 + y\vec{v}_2 &= \begin{pmatrix} 4 \\ 3 \end{pmatrix} \\ \implies x\begin{pmatrix} 1 \\ 1 \end{pmatrix} + y\begin{pmatrix} 1 \\ 2 \end{pmatrix} &= \begin{pmatrix} 4 \\ 3 \end{pmatrix} \\ \implies \begin{pmatrix} x \\ x \end{pmatrix} + \begin{pmatrix} y \\ 2y \end{pmatrix} &= \begin{pmatrix} 4 \\ 3 \end{pmatrix} \\ \implies \begin{pmatrix} x + y \\ x + 2y \end{pmatrix} &= \begin{pmatrix} 4 \\ 3 \end{pmatrix} \end{aligned}$$

Since two vectors are equal only when their components are equal, this means we really want to find x and y solving the system

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

So vector equations are really systems of equations in disguise! Furthermore, notice that the columns of the coefficient matrix for this system are exactly the original vectors in our vector equation!

3.6 Linear Combinations

We have two algebraic operations we can perform to vectors: scalar multiplication and vector addition. If we're given some vectors $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_m$ and we multiply each one by a some scalar – say we multiply $\vec{v}_i$ by λ_i – and then sum these vectors, we say that the resulting vector is a ***linear combination*** of the vectors $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_m$.

That is, a linear combination of $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_m$ is a vector that may be written as

$$\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \cdots + \lambda_m\vec{v}_m$$

for any choice of scalars λ_1, λ_2, ..., λ_m.

Example 3.2.
Consider the following two vectors in three-space:

$$\vec{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \quad \vec{v}_2 = \begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix}$$

One possible linear combination of these vectors is

$$7\vec{v}_1 - 4\vec{v}_2 = \begin{pmatrix} 7 \\ -7 \\ 14 \end{pmatrix} + \begin{pmatrix} -12 \\ -12 \\ 0 \end{pmatrix} = \begin{pmatrix} -5 \\ -19 \\ 14 \end{pmatrix}$$

Given $\vec{v}_1$, ..., $\vec{v}_m$ – all vectors of the same dimension, say n – the collection of all possible linear combinations of the vectors is called the ***span*** of the vectors and is denoted

$$\text{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_m\}.$$

In set-builder notation,

$$\text{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_m\} = \left\{\lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m \,\middle|\, \lambda_1, \lambda_2, ..., \lambda_m \in \mathbb{R}\right\}$$

Notice that in both $\mathbb{R}^2$ and $\mathbb{R}^3$, the span of a single, non-zero vector is a line through the origin. The span of two vectors could be plane (in $\mathbb{R}^2$ this would give us all possible vectors – the only plane sitting inside of $\mathbb{R}^2$ is the entire xy-plane), or it could be line. This second situation occurs when one vector is a multiple of another. That is, if our two vectors $\vec{v}_1$ and $\vec{v}_2$ have the property that $\vec{v}_2 = \mu\vec{v}_1$, then any linear combination of $\vec{v}_1$ and $\vec{v}_2$ is really just a multiple of $\vec{v}_1$:

$$\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 = \lambda_1\vec{v}_1 + \lambda_2\mu\vec{v}_1 = (\lambda_1 + \lambda_2\mu)\vec{v}_1.$$

In a situation like we say the vectors $\vec{v}_1$ and $\vec{v}_2$ are *linearly dependent*: meaning that one vector is a linear combination of another.

More generally, we say that a set of vectors $\{\vec{v}_1, ..., \vec{v}_m\}$ is ***linearly dependent*** if it's possible to write one vector as a linear combination of the others. If this can't be done – no vector is a linear combination of the others – then we say the set is ***linearly independent***.

Proposition 3.2.
A set of vectors $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_m\}$ is linearly independent if and only if the only scalars λ_1, λ_2, ..., λ_m making the following equation hold,

$$\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \cdots + \lambda_m\vec{v}_m = 0$$

are $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$.

Proof of Proposition 3.2.
If we could write

$$\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \cdots + \lambda_m\vec{v}_m = 0$$

for some non-zero choice of the λ_i, then we could pick one of the vectors $\vec{v}_i$ where $\lambda_i \neq 0$ and move it to the other side of the equation and divide by $-\lambda_i$ to write

$$\frac{\lambda_1}{-\lambda_i}\vec{v}_1 + \cdots + \frac{\lambda_{i-1}}{-\lambda_{i-1}}\vec{v}_{i-1} + \frac{\lambda_{i+1}}{-\lambda_{i+1}}\vec{v}_{i+1} + \cdots + \frac{\lambda_m}{-\lambda_i}\vec{v}_m = \vec{v}_i.$$

Thus if it is impossible to write one of the $\vec{v}_i$ as a linear combination of the other vectors (i.e., the vectors are linearly independent), then the only way to write $\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \cdots + \lambda_m\vec{v}_m = 0$ is if every λ_j was zero.

Conversely, if one of the vectors was a linear combination of the others,

$$\mu_i\vec{v}_1 + \cdots + \mu_{i-1}\vec{v}_{i-1} + \mu_{i+1}\vec{v}_{i+1} + \cdots + \mu_m\vec{v}_m = \vec{v}_i.$$

Then we can write

$$\mu_i\vec{v}_1 + \cdots + \mu_{i-1}\vec{v}_{i-1} - \vec{v}_i\mu_{i+1}\vec{v}_{i+1} + \cdots + \mu_m\vec{v}_m = 0.$$

Thus if we can not write $\lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m = 0$, then the vectors must be linearly independent. □

3.7 Practice Problems

Problem 3.1.
We saw that there was a graphical "triangle law" for vector addition. Come up with a similar law for vector subtraction.

Problem 3.2.
Find the values of x and y which solve the vector equation $x\vec{v} + y\vec{u} = \vec{w}$ where

$$\vec{v} = \begin{pmatrix} 3 \\ -2 \\ 8 \end{pmatrix} \quad \vec{u} = \begin{pmatrix} 5 \\ 0 \\ 9 \end{pmatrix} \quad \vec{w} = \begin{pmatrix} 2 \\ -3 \\ 8 \end{pmatrix}$$

Problem 3.3.
Show that every vector in $\mathbb{R}^2$ can be written as a linear combination of the vectors

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Problem 3.4.
Show that every vector in $\mathbb{R}^2$ can be written as a linear combination of the vectors

$$\begin{pmatrix} -1 \\ 3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 4 \\ -2 \end{pmatrix}$$

Problem 3.5.
Determine each of the following sets of vectors in $\mathbb{R}^3$ is linearly dependent or linearly independent:

(a) $\left\{ \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ -6 \\ 8 \end{pmatrix} \right\}$

(b) $\left\{ \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 6 \end{pmatrix} \right\}$

(c) $\left\{ \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ -6 \\ 8 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 6 \end{pmatrix} \right\}$

Problem 3.6.
Suppose you are given $m + 1$ vectors in $\mathbb{R}^n$, $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_m$, and $\vec{u}$. How can you determine if $\vec{u}$ is in the span of $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_m$?

Problem 3.7.
Is the vector

$$\vec{u} = \begin{pmatrix} 4 \\ 7 \\ 2 \\ -3 \end{pmatrix}$$

in the span of

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \\ -2 \end{pmatrix}, \right\}$$

If so, how can $\vec{u}$ be written as a linear combination of these vectors? If not, explain why $\vec{u}$ *can not* be written as a linear combination of these vectors.

The equation $Ax = b$ 4

The further a mathematical theory is developed, the more harmoniously and uniformly does its construction proceed, and unsuspected relations are disclosed between hitherto separated branches of the science.

DAVID HILBERT

In the last lecture we introduced vectors and saw that there were two algebraic operations that could be performed on vectors: vector addition and scalar multiplication. In general we can not multiply two vectors, but we can actually define the product of a matrix and a vector – at least if the sizes of the matrix and vector agree in a particular way. We will also see that this gives us a very concise way of expressing a system of a linear equations which will pave the way to later showing that properties of a linear system's coefficient matrix are directly related to the solutions of the system.

4.1 Products of Matrices and Vectors

Suppose that $\vec{a}_1$, $\vec{a}_2$, ..., $\vec{a}_n$ are vectors in $\mathbb{R}^m$. In the last lecture we considered linear combinations of vectors which were scalar multiples of the vectors added together:

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n.$$

Notice that these scalars, x_1, x_2, ..., x_n, that we multiply each vector by could be regarded as the components of some n-dimensional vector which we might call $\vec{x}$:

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

We could also think of each vector $\vec{a}_1$, ..., $\vec{a}_n$, as forming the columns of some matrix A:

$$A = \begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{pmatrix}$$

Example 4.1.
Suppose we have four three-dimensional vectors,

$$\vec{a}_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \quad \vec{a}_2 = \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} \quad \vec{a}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \vec{a}_4 = \begin{pmatrix} 4 \\ 2 \\ -7 \end{pmatrix}$$

and we considered the linear combination

$$5\vec{a}_1 - 3\vec{a}_2 + 2\vec{a}_3 + 2\vec{a}_4.$$

Then our vector $\vec{x}$ would be

$$\vec{x} = \begin{pmatrix} 5 \\ -3 \\ 2 \\ 2 \end{pmatrix}$$

and our matrix A would be

$$A = \begin{pmatrix} 2 & -1 & 1 & 4 \\ 1 & 0 & 1 & 2 \\ 0 & 3 & 1 & -7 \end{pmatrix}.$$

In general, we will define the ***product*** of an $m \times n$ matrix A with an n-dimensional vectors $\vec{x}$ as the linear combination of the columns of A with scalars given by the components of $\vec{x}$.

Example 4.2.
If A is the matrix

$$A = \begin{pmatrix} 3 & 4 & 0 \\ 2 & 1 & -1 \\ -5 & 7 & 2 \end{pmatrix}$$

and $\vec{x}$ is the vector

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

then the product $A\vec{x}$ is the the linear combination

$$-\begin{pmatrix}3\\2\\-5\end{pmatrix}+2\begin{pmatrix}4\\1\\7\end{pmatrix}+3\begin{pmatrix}0\\-1\\2\end{pmatrix}=\begin{pmatrix}-3+8+0\\-2+2-3\\5+14+6\end{pmatrix}=\begin{pmatrix}5\\-3\\25\end{pmatrix}$$

Example 4.3.
If A is the matrix

$$A=\begin{pmatrix}4&6&-5&3\\0&2&1&4\end{pmatrix}$$

and $\vec{x}$ is the matrix

$$\vec{x}=\begin{pmatrix}8\\-3\\5\\2\end{pmatrix}$$

then the product $A\vec{x}$ is the linear combination

$$\begin{aligned}&8\begin{pmatrix}4\\0\end{pmatrix}-3\begin{pmatrix}6\\2\end{pmatrix}+5\begin{pmatrix}-5\\1\end{pmatrix}+2\begin{pmatrix}3\\4\end{pmatrix}\\=&\begin{pmatrix}32-18-25+6\\0-6+5+8\end{pmatrix}\\=&\begin{pmatrix}-5\\7\end{pmatrix}\end{aligned}$$

Remark.
In order for this definition of the product of a matrix and a vector to make sense, it is absolutely essential that the number of columns of the matrix equals the number of rows in the vector (the dimension of the vector).

Example 4.4.

$$\begin{pmatrix} 3 & 2 & 4 \\ 1 & -2 & 3 \\ 0 & 4 & 4 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 6+0+12 \\ 2+0+9 \\ 0+0+12 \\ 2+0-3 \end{pmatrix}$$
$$= \begin{pmatrix} 18 \\ 11 \\ 12 \\ -1 \end{pmatrix}$$

4.2 The Matrix Equation $A\vec{x} = \vec{b}$

If $\vec{b}$ is some particular vector n-dimensional vector, we may want to know if there is a solution to the vector equation

$$x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n = \vec{b}$$

which can be more easily concisely written as

$$A\vec{x} = \vec{b}$$

where A is the matrix whose columns are given by $\vec{a}_1$, ..., $\vec{a}_n$, and $\vec{x}$ is the vector containing the variables x_1, ..., x_n.

Example 4.5.
Asking for x_1, x_2, and x_3 solving the vector equation

$$x_1 \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 4 \\ 7 \\ 7 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

is the same as asking if there is a vector

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

such that

$$\begin{pmatrix} 1 & 2 & 4 \\ -1 & 3 & 7 \\ 2 & 0 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

Notice that this is really just a system of linear equations with variables x_1, x_2, x_3 with coefficient matrix A and augmented coefficient matrix

$$\left(\begin{array}{c|c} A & \vec{b} \end{array} \right)$$

Thus solving systems of linear equations and solving the matrix equation $A\vec{x} = \vec{b}$ are two sides of the same coin.

Example 4.6.
If A and $\vec{b}$ are

$$A = \begin{pmatrix} 1 & 2 & 4 \\ -1 & 3 & 7 \\ 2 & 0 & 7 \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

Then finding a vector $\vec{x}$ solving $A\vec{x} = b$ is the same as finding a solution (x_1, x_2, x_3) to the system

$$\begin{aligned} x_1 + 2x_2 + 4x_3 &= 3 \\ -x_1 + 3x_2 + 7x_3 &= 1 \\ 2x_1 + x_3 &= 2 \end{aligned}$$

That solving systems of linear equations and solving matrix equations $A\vec{x} = b$ are really the same thing leads to the following theorem.

Proposition 4.1.
Let A be an $m \times n$ matrix. Then the system of linear equations with augmented coefficient matrix $\left(A \mid \vec{b}\right)$ has a solution if and only if $\vec{b}$ is in the span of the columns of A: $\vec{a}_1$, $\vec{a}_2$, ..., $\vec{a}_n$.

Notice that nothing deep is going on in this proposition: we're just translating the language of systems of linear equations to the language of matrix equations.

Exercise 4.1.
Prove Proposition 4.1.

Remark.
Sometimes the hardest part of solving a mathematical problem is determining the right way to express it: some problems seem easier or more difficult depending on the language you use to describe them. We are in the process of taking the ideas we described at the start of the semester (systems of linear equations) and converting them into another language (matrices and vectors) because, as we will see, it is actually a lot easier to think about many problems in terms of matrices and vectors. This may sound strange at first, especially if you're learning about matrices and vectors for the first time, but using the language matrices will actually make many problems much easier to think about and ultimately solve.

Example 4.7.

Is there a solution to the following matrix equation?

$$\begin{pmatrix} 2 & 3 & 4 \\ 6 & 18 & 24 \\ 2 & 3 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ -12 \\ -13 \end{pmatrix}$$

By the definition of the product of a matrix and a vector, this really means we want to find x_1, x_2, and x_3 such that

$$x_1 \begin{pmatrix} 2 \\ 6 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} 3 \\ 18 \\ 3 \end{pmatrix} + x_3 \begin{pmatrix} 4 \\ 24 \\ 9 \end{pmatrix} = \begin{pmatrix} 2 \\ -12 \\ -13 \end{pmatrix}$$

But if we do the scalar multiplication and vector addition we can rewrite the left-hand side of this equation to obtain

$$\begin{pmatrix} 2x_1 + 3x_2 + 4x_3 \\ 6x_1 + 18x_2 + 24x_3 \\ 2x_1 + 3x_2 + 9x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ -12 \\ -13 \end{pmatrix}$$

Equating components of the vectors, this is really a system of equations,

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 2 \\ 6x_1 + 18x_2 + 24x_3 &= -12 \\ 2x_1 + 3x_2 + 9x_3 &= -13 \end{aligned}$$

We know how to solve a system like this, though: we write down the augmented coefficient matrix (which we could have easily read off from the original matrix equation),

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 2 \\ 6 & 18 & 24 & -12 \\ 2 & 3 & 9 & -13 \end{array}\right)$$

then proceed to put the matrix in RREF, which gives us

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

This tells us the system of equations is equivalent to

$$x_1 = 4$$
$$x_2 = 2$$
$$x_3 = -3$$

and so we have a unique solution to the system.

In terms of the vector equation, we have

$$4\begin{pmatrix}2\\6\\2\end{pmatrix} + 2\begin{pmatrix}3\\18\\3\end{pmatrix} - 3\begin{pmatrix}4\\24\\9\end{pmatrix} = \begin{pmatrix}2\\-12\\-13\end{pmatrix}$$

And so the vector solving our original matrix equation is

$$\vec{x} = \begin{pmatrix}4\\2\\-3\end{pmatrix}.$$

That is,

$$\begin{pmatrix}2 & 3 & 4\\6 & 18 & 24\\2 & 3 & 9\end{pmatrix}\begin{pmatrix}4\\2\\-3\end{pmatrix} = \begin{pmatrix}2\\-12\\-13\end{pmatrix}$$

4.3 Existence of Solutions

We have seen that systems of linear equations sometimes have a unique solution, sometimes have no solution, and sometimes have infinitely-many solutions. Whether a solution exists or not depends less on the right-hand side of the equations of the system, and more about the coefficients of the system. In the language of matrices and vectors, solving $A\vec{x} = \vec{b}$ depends more on what A is than on what $\vec{b}$ is. In particular, we have the following theorem:

Theorem 4.2.
Suppose that A is an $m \times n$ matrix and $\vec{x}$ and $\vec{b}$ are m-dimensional vectors. Then the following are equivalent:

(a) *The equation $A\vec{x} = \vec{b}$ has a solution* for every *choice of $\vec{b}$.*

(b) *Every m-dimensional vector $\vec{b}$ is a linear combination of the columns of A.*

(c) *The columns of A span $\mathbb{R}^m$.*

(d) *A has a pivot position in every row.*

Remark.
The theorem above uses the phrase *the following are equivalent*. This means that if one of the statements is true, then all of the statements are true; if one of the statements is false, then all of the statements are false. This is really a shorthand for several *if and only if* statements. When we say "the following are equivalent: (a) ... (b) ... (c) ... (d) ..." what we really means is that statement (a) happens if and only if statement (b) happens if and only if statement (c) happens if and only if statement (d) happens.

We've seen before that "if and only if" statements are really two statements: there's actually two things to prove. If you want to prove "(a) if and only if (b)" then you need to show that statement (a) implies statement (b) and also that statement (b) implies statement (a). Thus it might seem like for the above we need to show twelve different things: (a) implies (b), (b) implies (a), (a) implies (c), (c) implies (a), (a) implies (d), (d) implies (a), (b) implies (c), (c) implies (b), and so on.

It would be completely correct to show all of these implications, but luckily there's an easier way. We can instead show that (a) implies (b), (b) implies (c), (c) implies (d), and finally (d) implies (a). If we show this then everything else can be deduced. For example, if we show the four implications above, then the fact that (c) implies (a), for instance comes for free: we know (c) implies (d) and also that (d) implies (a), hence (c) implies (a) as well.

In hand-written notes, *the following are equivalent* is often abbreviated *TFAE*.

Proof of Theorem 4.2.

$(a) \implies (b)$

Because of the way we have defined the product of a matrix and a vector, saying $A\vec{x} = \vec{b}$ means exactly that $\vec{b}$ is a linear combination of the columns of A. Hence if $A\vec{x} = \vec{b}$ has a solution for every $\vec{b}$, then it must be the case that every $\vec{b}$ can be written as a linear combination of the columns of A.

$(b) \implies (c)$

The span of a set of vectors is exactly the set of all possible linear combinations of those vectors. So if every vector in $\mathbb{R}^m$ can be written as a linear combination of the columns of A, then the span of the columns of A is all of $\mathbb{R}^m$.

$(c) \implies (d)$

We will prove the contrapositive: if (d) does not occur, then (c) can't occur either.

So suppose that there was some row that *did not* have a pivot. This means precisely that the row-reduced echelon form of A has a row of all zeros (otherwise we would have a leading entry of 1 which would be our pivot). We could then find choices of $\vec{b}$ so that the row-reduced echelon form of $\left(A \mid \vec{b} \right)$ has a row of all zeros followed by a 1. Thus the system has no solution which means $\vec{b}$ can't be written as a linear combination of the columns of A.

We've proven the contrapositive "if not (d), then not (c)" which is logically equivalent to the original statement "if (c), then (d)."

$(d) \implies (a)$

Finally, suppose that A has a pivot position in every row. Then the row-reduced echelon form of A has no rows of all zeros, and we can solve any system $A\vec{x} = \vec{b}$.

□

4.4 Properties of Ax

Algebraic properties

We have defined a new algebraic operation: multiplying an $m \times n$ matrix A with an m-dimensional vector $\vec{x}$. Anytime we introduce a new operation, it's natural to ask what kind of algebraic properties that operation satisfies. The following two properties are absolutely fundamental and will form the basis for what's to come when we define linear transformations.

Theorem 4.3.
If A is an $m \times n$ matrix, then for every pair of n-dimensional vectors $\vec{x}$ and $\vec{y}$, and every scalar $\lambda \in \mathbb{R}$, we have the following properties:

(a) $A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y}$

(b) $A(\lambda \cdot \vec{x}) = \lambda \cdot A\vec{x}$

Proof.
We do a direct proof, and simply verify these properties hold for any arbitrary $m \times n$ matrix A, arbitrary m-dimensional vectors $\vec{x}$ and $\vec{y}$, and arbitrary scalar λ.

We may suppose that A has the form

$$A = \begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{pmatrix}$$

and that

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

(a)

$$
\begin{aligned}
A(\vec{x}+\vec{y}) &= A\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}\right) \\
&= A\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_m+y_m \end{pmatrix} \\
&= (x_1+y_1)\vec{a}_1 + (x_2+y_2)\vec{a}_2 + \cdots + (x_m+y_m)\vec{a}_m \\
&= x_1\vec{a}_1 + y_1\vec{a}_1 + x_2\vec{a}_2 + y_2\vec{a}_2 + \cdots + x_m\vec{a}_m + y_m\vec{a}_m \\
&= x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_m\vec{a}_m + y_1\vec{a}_1 + y_2\vec{a}_2 + \cdots + y_m\vec{a}_m \\
&= A\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} + A\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \\
&= A\vec{x} + A\vec{y}
\end{aligned}
$$

(b)

$$
\begin{aligned}
A(\lambda\cdot\vec{x}) &= A\begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_m \end{pmatrix} \\
&= \lambda x_1\vec{a}_1 + \lambda x_2\vec{a}_2 + \cdots + \lambda x_m\vec{a}_m \\
&= \lambda\cdot(x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_m\vec{a}_m) \\
&= \lambda\cdot A\vec{x}
\end{aligned}
$$

□

Computational properties

There is an alternative way to think about the product $A\vec{x}$ that is sometimes handy. Notice that if A is an $m\times n$ matrix and $\vec{x}$ is an m-dimensional vector, then the product $A\vec{x}$ is also an n-dimensional vector: it's a linear combination of n-dimensional vectors (the columns of A). This vector can be generated one element at a time by walking across each row of the matrix A, while simultaneously going down the column vector $\vec{x}$ element by element, multiplying the elements and adding them up.

Example 4.8.
Consider the product

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

To get the first element in the product we look at the first row of the matrix, multiplying the entry in the j-th column by the j-th entry in the vector.

For the first entry we have

$$1 \cdot 2 + 4 \cdot 0 + (-2) \cdot (-2) + 1 \cdot 1 = 7.$$

We get the second entry,

$$1 \cdot 3 + 4 \cdot 1 + (-2) \cdot 0 + 1 \cdot 2 = 9.$$

For the third entry,

$$1 \cdot 2 + 4 \cdot (-2) + (-2) \cdot 1 + 1 \cdot 1 = -7.$$

And thus the product is

$$\begin{pmatrix} 7 \\ 9 \\ -7 \end{pmatrix}.$$

4.5 Practice Problems

Problem 4.1.
Find the values of $\vec{x}$ solving the system $A\vec{x} = \vec{b}$ where A and $\vec{b}$ are given in each problem below.

(a)
$$A = \begin{pmatrix} 1 & 3 & -4 \\ 1 & 5 & 2 \\ -3 & -7 & 6 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} -2 \\ 4 \\ 12 \end{pmatrix}$$

(b)
$$A = \begin{pmatrix} 1 & 3 & -4 \\ 1 & 5 & 2 \\ -3 & -7 & 6 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} -1 \\ 2 \\ 6 \end{pmatrix}$$

(c)
$$A = \begin{pmatrix} 1 & 2 & -1 \\ -3 & -4 & 2 \\ 5 & 2 & 3 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$$

(d)
$$A = \begin{pmatrix} 3 & -5 \\ -2 & 6 \\ 1 & 1 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} 0 \\ 4 \\ 4 \end{pmatrix}$$

(e)
$$A = \begin{pmatrix} 2 & 5 & -1 \\ 0 & 1 & -1 \\ 1 & 2 & 0 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} 4 \\ -1 \\ 4 \end{pmatrix}$$

Problem 4.2.
Do the columns of the following matrix span $\mathbb{R}^4$?

$$\begin{pmatrix} 1 & 4 & 1 & 2 \\ 0 & 1 & 3 & -4 \\ 0 & 2 & 6 & 7 \\ 2 & 9 & 5 & -7 \end{pmatrix}$$

Problem 4.3.
Do the columns of the following matrix span $\mathbb{R}^4$?

$$\begin{pmatrix} 1 & 3 & 0 & 3 \\ -1 & -1 & -1 & 1 \\ 0 & -4 & 2 & -8 \\ 2 & 0 & 3 & -1 \end{pmatrix}$$

Problem 4.4.
Do the vectors below span $\mathbb{R}^3$?

$$\vec{u} = \begin{pmatrix} 0 \\ 0 \\ -3 \end{pmatrix} \quad \vec{v} = \begin{pmatrix} 0 \\ -3 \\ 9 \end{pmatrix} \quad \vec{w} = \begin{pmatrix} 4 \\ -2 \\ -6 \end{pmatrix}$$

5 Solution Sets

I have not failed. I've just found 10,000 ways that won't work.

THOMAS EDISON

Recall that a system of equations may have zero solutions, a single unique solution, or infinitely-many solutions. Regardless of how many solutions we have, we say that the collection of all possible solutions to a system of n-variables (the collection of all points $(x_1, x_2, ..., x_n)$ in $\mathbb{R}^n$ whose coordinates simultaneously satisfy all of the equations in the system) the ***solution set*** of the system.

Remark.
If a system has no solutions, then the solution set is the empty set $\emptyset$.

5.1 Homogeneous Systems

A system of linear equations where the right-hand sides are all zeroes,

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= 0 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= 0 \\ &\vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= 0 \end{aligned}$$

is called a ***homogeneous system***. As a matrix equation, we have the coefficient matrix A whose columns are given by

$$\vec{a}_j = \begin{pmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{pmatrix}$$

and we are interested in finding the vector

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

so that $A\vec{x} = \vec{0}$.

Notice that homogeneous systems *always* have a solution: at the very least we can take $x_1 = x_2 = \cdots = x_n = 0$ – in terms of vectors, $\vec{x} = \vec{0}$. This is called the ***trivial solution*** to the homogeneous system. If there are other solutions, we call them ***non-trivial solutions***.

Notice that there are only two options for a homogeneous system: there is either only the trivial solution, or there are infinitely-many solutions.

Exercise 5.1.
Suppose that there is an $\vec{x} \neq \vec{0}$ solving $A\vec{x} = 0$. Show that there must in be infinitely-many other, non-trivial solutions.

Recall that when a linear system has infinitely-many solutions there is some way to parametrize the solution set: there is some way to explicitly describe all of the solutions as a function of the free variables.

In particular, suppose the solution set is spanned by vectors $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_d$. Then the solution set to the system $A\vec{x} = \vec{0}$ is parametrized by the linear combinations of the $\vec{v}_i$.

Example 5.1.
Parametrize all of the solutions to

$$\begin{pmatrix} 1 & -2 \\ -3 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Solving this matrix equation is the same as solving the system of linear equations

$$\begin{aligned} x_1 - 2x_2 &= 0 \\ -3x_2 + 6x_2 &= 0 \end{aligned}$$

If we put the augmented coefficient matrix of this system in RREF we have

$$\left(\begin{array}{cc|c} 1 & -2 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

which means that our system is equivalent to

$$x_1 - 2x_2 = 0$$

so our solutions satisfy

$$x_1 = 2x_2.$$

That is, our vectors satisfying the matrix equation have the form

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_2 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

So our solutions live in the span of this one vector; the solutions are parametrized by

$$\vec{x} = t \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

where t can be any real number.

Notice that we could have instead rewritten the equation

$$x_1 - 2x_2 = 0$$

as

$$x_2 = \frac{1}{2}x_1.$$

and said that our solutions have the form

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ 1/2 x_2 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$$

and our solutions are also parametrized by

$$\vec{x} = s \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$$

There is no contradicition here: we simply have parametrized the set of all possible solutions in two different ways:

$$\vec{x} = t\begin{pmatrix}2\\1\end{pmatrix} = s\begin{pmatrix}1\\1/2\end{pmatrix}.$$

we could parametrize this in many other different ways:

$$\begin{aligned}\vec{x} =& t\begin{pmatrix}2\\1\end{pmatrix} = s\begin{pmatrix}1\\1/2\end{pmatrix}\\ =& \tau\begin{pmatrix}-14\\-7\end{pmatrix}\\ =& \sigma\begin{pmatrix}528\\264\end{pmatrix}.\end{aligned}$$

The important thing here is the collection of all possible things we can get as scalar multiples of these vectors, and they all describe the same set:

$$\begin{aligned}&\operatorname{span}\left(\left\{\begin{pmatrix}2\\1\end{pmatrix}\right\}\right)\\ =&\operatorname{span}\left(\left\{\begin{pmatrix}1\\1/2\end{pmatrix}\right\}\right)\\ =&\operatorname{span}\left(\left\{\begin{pmatrix}-14\\-7\end{pmatrix}\right\}\right)\\ =&\operatorname{span}\left(\left\{\begin{pmatrix}528\\264\end{pmatrix}\right\}\right).\end{aligned}$$

For example, the vector $\begin{pmatrix}8\\4\end{pmatrix}$ is a solution to the system; it's each of the spans above,

$$\begin{pmatrix}8\\4\end{pmatrix} = 4\begin{pmatrix}2\\1\end{pmatrix}$$
$$= 8\begin{pmatrix}1\\1/2\end{pmatrix}$$
$$= -\frac{4}{7}\begin{pmatrix}-14\\-7\end{pmatrix}$$
$$= \frac{1}{66}\begin{pmatrix}528\\264\end{pmatrix}$$

Example 5.2.
Parametrize all of the solutions to $A\vec{x} = \vec{0}$ where

$$A = \begin{pmatrix}3 & 1 & -5 & -1\\0 & 1 & 1 & -1\\2 & 1 & -3 & -1\\1 & 2 & 0 & -2\end{pmatrix}$$

We are trying to describe all of the values of $\vec{x}$ that make the following equaiton hold:

$$\begin{pmatrix}3 & 1 & -5 & -1\\0 & 1 & 1 & -1\\2 & 1 & -3 & -1\\1 & 2 & 0 & -2\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix}$$

If we put the augmented coefficient matrix of this system into RREF we have

$$\left(\begin{array}{cccc|c}1 & 0 & 0 & -2 & 0\\0 & 1 & 1 & -1 & 0\\0 & 0 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 0\end{array}\right)$$

This means our system is equivalent to

$$\begin{aligned} x_1 - 2x_4 &= 0 \\ x_2 + x_3 - x_4 &= 0 \end{aligned}$$

Thus

$$\begin{aligned} x_1 &= 2x_4 \\ x_2 &= -x_3 + x_4 \end{aligned}$$

and x_3 and x_4 are free variables. So, our vectors $\vec{x}$ solving the equation look like

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2x_4 \\ -x_3 + x_4 \\ x_3 \\ x_4 \end{pmatrix}$$

which we can write as

$$\vec{x} = x_3 \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

So the solutions to our equation are precisely the vectors in

$$\text{span}\left(\left\{\begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}\right\}\right).$$

That is, our solutions are parametrized by

$$\vec{x} = s \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

where s and t can be any real numbers.

Now, this particular parametrization comes from the fact that we took x_3 and x_4 to be free variables above. We could have instead taken x_1 and x_2 to be free variables. If we rewrite the equations

$$\begin{aligned} x_1 &= 2x_4 \\ x_2 &= -x_3 + x_4 \end{aligned}$$

so that x_3 and x_4 are functions of x_1 and x_2, then we have

$$\begin{aligned} x_4 &= \frac{1}{2}x_1 \\ x_3 &= x_4 - x_2 = \frac{1}{2}x_1 - x_2 \end{aligned}$$

which tells us that the solutions look like

$$\begin{aligned} \vec{x} &= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ x_2 \\ \frac{1}{2}x_1 - x_2 \\ \frac{1}{2}x_1 \end{pmatrix} \\ &= x_1 \begin{pmatrix} 1 \\ 0 \\ 1/2 \\ 1/2 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \end{aligned}$$

So the set of solutions to $A\vec{x} = \vec{0}$ is

$$\operatorname{span}\left(\left\{\begin{pmatrix} 1 \\ 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}\right\}\right)$$

That is, every vector $\vec{x}$ has the form

$$\vec{x} = \sigma \begin{pmatrix} 1 \\ 0 \\ 1/2 \\ 1/2 \end{pmatrix} + \tau \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$$

Thus we have said that the set of solutions to $A\vec{x} = \vec{0}$ can be described in two different ways:

$$\begin{aligned} \vec{x} &= s \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix} \\ &= \sigma \begin{pmatrix} 1 \\ 0 \\ 1/2 \\ 1/2 \end{pmatrix} + \tau \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \end{aligned}$$

Again there's no contradiction here: if you write any one vector in the solution set as

$$\vec{x} = s \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

you can write the same vector as

$$\vec{x} = \sigma \begin{pmatrix} 1 \\ 0 \\ 1/2 \\ 1/2 \end{pmatrix} + \tau \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$$

by taking $\sigma = 2t$ and $\tau = -s$.

For example, the vector

$$\begin{pmatrix} 4 \\ -1 \\ 3 \\ 2 \end{pmatrix}$$

is a solution to the system. In terms of one parametrization we get this vector by taking $s = 3$ and $t = 2$:

$$\begin{pmatrix} 4 \\ -1 \\ 3 \\ 2 \end{pmatrix} = 3 \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

In terms of the other parametrization we take $\sigma = 4$ and $\tau = -3$:

$$\begin{pmatrix} 4 \\ -1 \\ 3 \\ 2 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 0 \\ 1/2 \\ 1/2 \end{pmatrix} - 3 \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}.$$

Example 5.3.
Parametrize the set of vectors solving the equation $A\vec{x} = \vec{0}$ where A is the matrix

$$A = \begin{pmatrix} 4 & 2 & 9 & 3 \\ 3 & 2 & 7 & 2 \\ 3 & 1 & 7 & 1 \\ 2 & 4 & 5 & 3 \end{pmatrix}$$

Notice that the RREF of A is

$$\begin{pmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

When we solve the system $A\vec{x} = \vec{0}$, the RREF of the augmented coefficient matrix is thus

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 7 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

which means that

$$\begin{aligned} x_1 + 7x_4 &= 0 \\ x_2 + x_4 &= 0 \\ x_3 - 3x_4 &= 0 \end{aligned}$$

Thus

$$\begin{aligned} x_1 &= -7x_4 \\ x_2 &= -x_4 \\ x_3 &= 3x_4 \end{aligned}$$

and x_4 can be whatever we want.

That is, the solutions to the system $A\vec{x} = \vec{0}$ have the form

$$\begin{pmatrix} -7x_4 \\ -x_4 \\ 3x_4 \\ x_4 \end{pmatrix}$$

which we could write as

$$x_4 \begin{pmatrix} -7 \\ -1 \\ 3 \\ 1. \end{pmatrix}$$

So all the solutions to $A\vec{x} = \vec{0}$ are scalar multiples of one vector.

Another way to think about this is to notice that our original equation,

$$\begin{pmatrix} 4 & 2 & 9 & 3 \\ 3 & 2 & 7 & 2 \\ 3 & 1 & 7 & 1 \\ 2 & 4 & 5 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

is equivalent to

$$\begin{pmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

But notice that when we multiply the matrix and vector on the left-hand side we get

$$\begin{pmatrix} x_1 + 7x_4 \\ x_2 + x_4 \\ x_3 - 3x_4 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

We could rewrite this as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} 7x_4 \\ x_4 \\ -3x_4 \\ -x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

which we could further write as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + x_4 \begin{pmatrix} 7 \\ 1 \\ -3 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

but this implies

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_4 \begin{pmatrix} -7 \\ -1 \\ 3 \\ 1 \end{pmatrix}.$$

Example 5.4.

Parametrize the solutions of

$$\begin{pmatrix} 5 & 2 & -2 & 15 & 10 \\ 4 & 1 & 0 & 5 & 10 \\ 3 & 1 & 1 & 5 & -25 \\ 6 & 2 & -3 & 15 & 45 \\ 1 & 0 & -1 & 0 & 35 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Notice that the RREF of $\left(A \mid \vec{0}\right)$ is

$$\left(\begin{array}{ccccc|c} 1 & 0 & 0 & -1 & 16 & 0 \\ 0 & 1 & 0 & 9 & -54 & 0 \\ 0 & 0 & 1 & -1 & -19 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

This means that our original system is equivalent to

$$\begin{pmatrix} 1 & 0 & 0 & -1 & 16 \\ 0 & 1 & 0 & 9 & -54 \\ 0 & 0 & 1 & -1 & -19 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Multiplying the matrix and vector on the left gives

$$\begin{pmatrix} x_1 - x_4 + 16x_5 \\ x_2 + 9x_4 - 54x_5 \\ x_3 - x_4 - 19x_5 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

We can rewrite this as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 9 \\ -1 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} 16 \\ -54 \\ -19 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Let's go one step furth and write this as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_4 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 9 \\ -1 \\ -1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} 16 \\ -54 \\ -19 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

which then implies

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = x_4 \begin{pmatrix} 1 \\ -9 \\ 1 \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -16 \\ 54 \\ 19 \\ 0 \\ 1 \end{pmatrix}$$

So we may parametrize the set of all solutions as

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = s \begin{pmatrix} 1 \\ -9 \\ 1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -16 \\ 54 \\ 19 \\ 0 \\ 1 \end{pmatrix}$$

At this point you've probably noticed that there's some sort of relationship between the solutions of a homogeneous system and the columns of our matrix in RREF.

Example 5.5.
In Example 5.1 we saw that our matrix in RREF was

$$\begin{pmatrix} 1 & -2 \\ 0 & 0 \end{pmatrix}$$

Notice that one possible parametrization of the solution set of Example 5.1 was the collection of all scalar mutliples of $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

In Example 5.2 our matrix in RREF was

$$\begin{pmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and we could parametrize the solution set as linear combinations of

$$\begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

In Example 5.3 the matrix in RREF was

$$\begin{pmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and the solution set could be parametrized as scalar multiples of

$$\begin{pmatrix} -7 \\ -1 \\ 3 \\ 1 \end{pmatrix}.$$

In Example 5.4 our matrix in RREF was

$$\left(\begin{array}{ccccc|c} 1 & 0 & 0 & -1 & 16 & 0 \\ 0 & 1 & 0 & 9 & -54 & 0 \\ 0 & 0 & 1 & -1 & -19 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

and the solution set could be parametrized as linear combinations of

$$\begin{pmatrix} 1 \\ -9 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -16 \\ 54 \\ 19 \\ 0 \\ 1 \end{pmatrix}.$$

Notice that in each example after we put the matrix into RREF some of the columns of the matrix wind up telling us how to parametrize the solution set.

Let's look at one more example before we make the relationship between the columns in RREF and the solution set precise.

Example 5.6.
Parametrize solutions to $A\vec{x} = \vec{0}$ where

$$A = \begin{pmatrix} 4 & 12 & 2 & 6 & 10 \\ 1 & 3 & 1 & 3 & 3 \\ 3 & 9 & 2 & 6 & 8 \end{pmatrix}$$

In RREF this matrix becomes

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

So solving $A\vec{x} = \vec{0}$ is the same as solving

$$\begin{aligned} x_1 + 3x_2 + 2x_5 =& 0 \\ x_3 + 3x_4 + x_5 =& 0 \end{aligned}$$

We could write this as

$$\begin{aligned} x_1 &= -3x_2 - 2x_5 \\ x_3 &= -3x_4 - x_5 \end{aligned}$$

So the solutions have the form

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -3x_2 - 2x_5 \\ x_2 \\ -3x_4 - x_5 \\ x_4 \\ x_5 \end{pmatrix} = x_2 \begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -2 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

The relationship between the columns of our matrix and the solution set is elucidated by the following:

Proposition 5.1.
The number of free variables in the solution set of $A\vec{x} = \vec{0}$ *is precisely the number of columns of the matrix that do not contain pivot.*

This proposition is actually a corollary of a very important result called the *rank-nullity theorem* so we will wait and prove this proposition after we prove the rank-nullity theorem.

5.2 Non-Homogeneous Equations

All of the examples of parametrizations we have seen thus far have been for homogeneous systems, but parametrizing solutions to non-homogeneous systems is almost identical.

Theorem 5.2.
Suppose that $A\vec{x} = \vec{b}$ *has a solution* $\vec{y}$*. Then the solutions to* $A\vec{x} = \vec{b}$ *all have the form* $\vec{y} + \vec{h}$ *where* $\vec{h}$ *is a solution to the homogeneous equation* $A\vec{x} = \vec{0}$*.*

Proof.
It's clear that $\vec{y} + \vec{h}$ is a solution to $A\vec{x} = \vec{b}$:

$$A\left(\vec{y} + \vec{h}\right) = A\vec{y} + A\vec{h} = \vec{b} + \vec{0} = \vec{b}.$$

Now suppose that $\vec{y'}$ is any other solution to $A\vec{x} = \vec{b}$ and let $\vec{h'} = \vec{y'} - \vec{y}$. Notice that

$$A\vec{h'} = A(\vec{y'} - \vec{y}) = A\vec{y'} - A\vec{y} = \vec{b} - \vec{b} = \vec{0}.$$

We again have that $\vec{y'} = \vec{y} + \vec{h'}$ where $\vec{h'}$ solves the homogeneous equation. □

This theorem tells us that if we can parametrize $A\vec{x} = \vec{0}$ then we can just as easily parametrize $A\vec{x} = \vec{b}$ by taking any one solution $\vec{y}$ of $A\vec{x} = \vec{b}$, called a ***particular solution*** and adding to it the solutions to the homogeneous equation $A\vec{x} = \vec{0}$.

Example 5.7.
Parametrize all of the solutions to

$$\begin{pmatrix} 1 & -2 \\ -3 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ -15 \end{pmatrix}$$

Once we have one solution to this system, such as

$$\vec{x} = \begin{pmatrix} 9 \\ 2 \end{pmatrix},$$

then all of the other solutions to the equation have the form

$$\begin{pmatrix} 9 \\ 2 \end{pmatrix} + t \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

where t can be any real number because we know from Example 5.1 that solutions to the homogeneous equation with the same matrix have the form $t \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Choosing a different particular solution doesn't really change anything. For example, if we had instead used

$$\vec{x} = \begin{pmatrix} 13 \\ 4 \end{pmatrix}$$

then we still have that all of the solutions to our equation have the form

$$\begin{pmatrix} 13 \\ 4 \end{pmatrix} + \tau \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

We could have also used a different parametrization for the homogeneous part of the solutions and write all the solutions as

$$\begin{pmatrix} 13 \\ 4 \end{pmatrix} + s \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}.$$

For example, $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ can be obtained from the first equation by taking $t = -4$, from the second by taking $\tau = -6$, and from the third by taking $s = -12$:

$$\begin{aligned} \begin{pmatrix} 1 \\ -2 \end{pmatrix} &= \begin{pmatrix} 9 \\ 2 \end{pmatrix} - 4 \begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 13 \\ 4 \end{pmatrix} - 6 \begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 13 \\ 4 \end{pmatrix} - 12 \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}. \end{aligned}$$

Exercise 5.2.
Parametrize the solutions to $A\vec{x} = \vec{b}$ where A and $\vec{b}$ are as described below:

(a)

$$A = \begin{pmatrix} 3 & 1 & -5 & -1 \\ 0 & 1 & 1 & -1 \\ 2 & 1 & -3 & -1 \\ 1 & 2 & 0 & -2 \end{pmatrix} \quad \text{and} \quad \vec{b} = \begin{pmatrix} 2 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

(b)

$$A = \begin{pmatrix} 4 & 2 & 9 & 3 \\ 3 & 2 & 7 & 2 \\ 3 & 1 & 7 & 1 \\ 2 & 4 & 5 & 3 \end{pmatrix} \quad \text{and} \quad \vec{b} = \begin{pmatrix} 2 \\ 0 \\ -2 \\ 2 \end{pmatrix}$$

(c)

$$A = \begin{pmatrix} 5 & 2 & -2 & 15 & 10 \\ 4 & 1 & 0 & 5 & 10 \\ 3 & 1 & 1 & 5 & -25 \\ 6 & 2 & -3 & 15 & 45 \\ 1 & 0 & -1 & 0 & 35 \end{pmatrix} \quad \text{and} \quad \vec{b} = \begin{pmatrix} 30 \\ 27 \\ -6 \\ 63 \\ 33 \end{pmatrix}$$

(d)

$$A = \begin{pmatrix} 4 & 12 & 2 & 6 & 10 \\ 1 & 3 & 1 & 3 & 3 \\ 3 & 9 & 2 & 6 & 8 \end{pmatrix} \quad \text{and} \quad \vec{b} = \begin{pmatrix} 12 \\ 6 \\ 12 \end{pmatrix}$$

When we already know how to parametrize the solutions set of the homogeneous equation $A\vec{x} = \vec{0}$ and we have some particular solution to $A\vec{x} = \vec{b}$, then it becomes easy to parametrize all of the solutions to $A\vec{x} = \vec{b}$. Usually we won't have a particular solution to start with, but the above discussion tells us how to parametrize solution sets.

5.3 Practice Problems

Problem 5.1.
Parametrize the set of solutions to the following system of linear equations:

$$\begin{aligned} 2x_1 + 2x_2 + 4x_3 =& 0 \\ -4x_1 - 4x_2 - 8x_3 =& 0 \\ -3x_2 - 3x_3 =& 0 \end{aligned}$$

Problem 5.2.
Parametrize the set of solutions to the following system of linear equations:

$$\begin{aligned} x_1 + 2x_2 - 3x_3 =& 0 \\ 2x_1 + x_2 - 3x_2 =& 0 \\ -x_1 + x_2 =& 0 \end{aligned}$$

Problem 5.3.
Parametrize the set of solutions to the equation $A\vec{x} = 0$ where A is the matrix below:

$$A = \begin{pmatrix} 1 & -2 & 3 & -6 & 5 & 0 \\ 0 & 0 & 0 & 1 & 4 & -6 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Problem 5.4.
Suppose the solution set to a certain system of linear equations can be described as

$$\begin{aligned} x_1 =& 5x_4 \\ x_2 =& 3 - 2x_4 \\ x_3 =& 2 + 5x_4 \end{aligned}$$

and x_4 is a free variable. Use vectors to describe this set as a line in $\mathbb{R}^4$.

6

Matrix Algebra

Algebra is the metaphysics of arithmetic.

JOHN RAY

In this lecture we discuss the various types of operations that can be performed on matrices, and the algebra of these operations.

6.1 Linear Transformations

Many of the operations we perform on matrices have an interpretation in terms of functions, and understanding that interpretation helps to motivate why some of the constructions below are things we should care about.

We will say that a map $T : \mathbb{R}^n \to \mathbb{R}^m$ (that is, a map that takes n-dimensional vectors and converts them into m-dimensional vectors) is a ***linear transformation*** if it satisfies the following two axioms:

(i) For every pair of n-dimensional vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, T satisfies the following equation:

$$T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}).$$

(ii) For every n-dimensional vector $\vec{v}$ and every scalar $\lambda \in \mathbb{R}$, we have

$$T(\lambda\vec{v}) = \lambda T(\vec{v}).$$

Recall that there are two basic operations we can perform on vectors: vector addition and scalar multiplication. Linear transformations are precisely the maps that "respect" these two operations.

Example 6.1.
Consider the following which takes two-dimensional vectors and

transforms them into three-dimensional vectors:

$$T : \mathbb{R}^2 \to \mathbb{R}^3$$
$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ -x \\ x+y \end{pmatrix}$$

This map takes $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and turns it into the vector

$$T \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix}$$

and it takes the vector $\begin{pmatrix} -2 \\ -6 \end{pmatrix}$ and turns it into

$$T \begin{pmatrix} -2 \\ -6 \end{pmatrix} = \begin{pmatrix} -6 \\ 2 \\ -8 \end{pmatrix}$$

Example 6.2.
The following map takes four-dimensional vectors and turns them into two-dimensional vectors:

$$T : \mathbb{R}^4 \to \mathbb{R}^2$$
$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 - 2x_4 \\ 4x_3 + x_2 \end{pmatrix}$$

Here are some examples of what this function does:

$$T\begin{pmatrix}1\\2\\3\\4\end{pmatrix} = \begin{pmatrix}-7\\14\end{pmatrix}$$
$$T\begin{pmatrix}7\\2\\0\\-1\end{pmatrix} = \begin{pmatrix}9\\2\end{pmatrix}$$
$$T\begin{pmatrix}8\\4\\3\\3\end{pmatrix} = \begin{pmatrix}2\\16\end{pmatrix}$$

Notice that these two different maps are actually given by matrices:

Example 6.3.
The map $T : \mathbb{R}^2 \to \mathbb{R}^3$ from Example 6.1 is given by multiplying a vector $\begin{pmatrix}x\\y\end{pmatrix}$ with the matrix

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

For example,

$$\begin{pmatrix}0 & 1\\-1 & 0\\1 & 1\end{pmatrix}\begin{pmatrix}1\\3\end{pmatrix} = 1\begin{pmatrix}0\\-1\\1\end{pmatrix} + 3\begin{pmatrix}1\\0\\1\end{pmatrix} = \begin{pmatrix}3\\-1\\4\end{pmatrix}.$$

Example 6.4.
The map $T : \mathbb{R}^4 \to \mathbb{R}^2$ from Example 6.2 is given by multiplying a

four-dimensional vector with the matrix

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

For example,

$$\begin{pmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 4 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} -7 \\ 14 \end{pmatrix}$$

In fact, *every matrix* determines such a map: Every $m \times n$ matrix defines a map from $\mathbb{R}^n$ to $\mathbb{R}^m$ by matrix multiplication:

$$\vec{x} \mapsto A\vec{x}.$$

By the properties of multiplication between matrices and vectors, we see that such a map is always a linear transformation. In fact, it turns out that every linear transformation is determined by a matrix in this way.

6.2 The Matrix of a Linear Transformation

We said above that every matrix determines a linear transformation. It turns out, however, that *every* linear transformation is determined by a matrix. That is, for every linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$, there is some $m \times n$ matrix A such that $T(\vec{v}) = A\vec{v}$.

To see this, let's notice that every n-dimensional vector can be written as a linear combination of the vectors

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \quad \cdots \quad \vec{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

That is, $\vec{e}_i$ is the vector that has all zeros except for a 1 in the i-th row.

Exercise 6.1.

(a) Show that the set of n-dimensional vectors $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ is linearly independent.

(b) Show that every n-dimensional vector can be written as a linear combination of the $\vec{e}_i$.

Now suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and so $T(\vec{e}_i)$ is some m-dimensional vector – let's call it $\vec{a}_i$. Now consider the matrix A whose columns are given by these vectors,

$$A = \begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{pmatrix}.$$

This matrix represents our linear transformation: if

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

is any vector, then we have

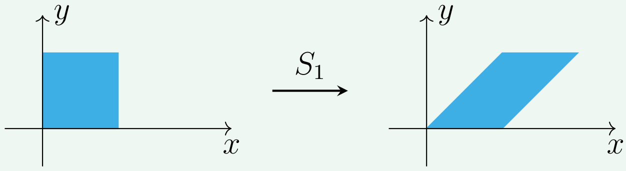
$$\begin{aligned} & T(\vec{x}) \\ =& T\left(x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n\right) \\ =& x_1 T(\vec{e}_1) + x_2 T(\vec{e}_2) + \cdots + x_n T(\vec{e}_n) \\ =& x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n \end{aligned}$$

But notice that

$$\begin{aligned} & x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n \\ =& \begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ =& A\vec{x} \end{aligned}$$

Thus every linear transformation is determined by some matrix!

Exercise 6.2.
Consider the map $S_m : \mathbb{R}^2 \to \mathbb{R}^2$ which *shears* the plane horizontally by a factor of m. That is, a given vector $\begin{pmatrix} x \\ y \end{pmatrix}$ gets sent to $\begin{pmatrix} x + my \\ y \end{pmatrix}$.



(a) Show that each S_m is a linear transformation.

(b) Determine the matrix representing S_m.

Exercise 6.3.
Consider the map $R : \mathbb{R}^2 \to \mathbb{R}^2$ which reflects across the x-axis:

$$R \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}$$

(a) Show that R is a linear transformation.

(b) Determine the matrix representing R.

6.3 Matrix Addition and Scalar Multiplication

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^n \to \mathbb{R}^m$ be the corresponding linear transformations. We can produce a new linear transformation $T + S : \mathbb{R}^n \to \mathbb{R}^m$ by adding vectors. That is, given a vector $\vec{v} \in \mathbb{R}^n$ we consider the map

$$\vec{v} \mapsto T(\vec{v}) + S(\vec{v}).$$

This gives us a new map which we denote $T+S$. It's easy to see that if T and S are linear transformations, then so is $T+S$:

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

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

Since $T+S$ is a linear transformation, there must be some matrix representing it. Before we determine what this matrix must be, let's suppose that T is given by the matrix A with columns $\vec{a}_1$, $\vec{a}_2$, ..., $\vec{a}_n$, and S is determined by the matrix B with columns $\vec{b}_1$, $\vec{b}_2$, ..., $\vec{b}_n$.

We know that the columns of the matrix representing $T+S$ are given by the vectors $(T+S)(\vec{e}_i)$, where $\vec{e}_i$ is the n-dimensional vector that consists entirely of zero except for a 1 in the i-th component.

Notice

$$\begin{aligned}(T+S)(\vec{e}_i) =& T(\vec{e}_i)+S(\vec{e}_i)\\ =& a_i+b_i\end{aligned}$$

That is, the columns of the matrix representing $T+S$ are determined by adding the columns of the matrices representing T and S.

Example 6.5.
Suppose T and S are the linear transformations corresponding to the matrices

$$A=\begin{pmatrix}1 & 0 & 2\\ 3 & -1 & 4\\ 5 & 2 & 2\\ 1 & -1 & -1\end{pmatrix} \quad B=\begin{pmatrix}-1 & 3 & 3\\ 2 & 2 & -2\\ 1 & 1 & -3\\ 4 & 1 & 2\end{pmatrix}$$

Then the matrix corresponding to $T+S$ is

$$\begin{pmatrix} 0 & 3 & 5 \\ 5 & 1 & 2 \\ 6 & 3 & -1 \\ 5 & 0 & 1 \end{pmatrix}$$

The matrix we get by adding the columns of a matrix A with the columns of a matrix B like this is denoted $A+B$. Notice that since we add vectors (the columns of the matrices) component-by-component, we add matrices component-by-component as well – this also means that addition of matrices only makes sense if the matrices are the same size.

Example 6.6.

$$\begin{pmatrix} 2 & 3 & 4 & 2 \\ 1 & 0 & 2 & -1 \\ 3 & 4 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & -1 & 1 & 0 \\ 2 & 2 & 7 & 13 \\ 4 & 2 & 9 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 5 & 2 \\ 3 & 2 & 9 & 12 \\ 7 & 6 & 9 & 2 \end{pmatrix}$$

Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ and a scalar μ we can define a new map by multiplying the outputs of T with μ: $\vec{v} \mapsto \mu T(\vec{v})$. This map is denoted μT and is also a linear transformation:

$$\begin{aligned} \mu T(\vec{v}+\vec{w}) &= \mu(T(\vec{v}) + T(\vec{w})) \\ &= \mu T(\vec{v}) + \mu T(\vec{w}) \end{aligned}$$

$$\mu T(\lambda \vec{v}) = \mu \lambda T(\vec{v})$$

Since μT is a linear transformation it is represented by some matrix whose columns are $\mu T(\vec{e}_i)$. If T is represented by matrix A with columns $\vec{a}_i = T(\vec{e}_i)$, then μT is represented by the matrix with columns $\mu \vec{a}_i$. That is, the matrix representing μT is simply the matrix representing T, but with every entry multiplied by μ.

Example 6.7.
If $T : \mathbb{R}^2 \to \mathbb{R}^4$ is given by the matrix

$$\begin{pmatrix} 4 & 2 \\ 6 & -3 \\ 7 & 8 \\ 2 & 0 \end{pmatrix},$$

then the matrix representing $-5T$ is given by

$$\begin{pmatrix} -20 & -10 \\ -30 & 15 \\ -35 & -40 \\ -10 & 0 \end{pmatrix}.$$

The matrix obtained by multiplying each entry of a matrix A by μ is denoted μA.

Example 6.8.

$$3 \begin{pmatrix} 2 & 7 & 1 & 0 \\ 4 & 2 & -2 & 3 \\ 1 & 1 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 6 & 21 & 3 & 0 \\ 12 & 6 & -6 & 9 \\ 3 & 3 & 12 & 6 \end{pmatrix}$$

So we have two operations we can perform on matrices: matrix addition and scalar multiplication, corresponding to performing vector addition and scalar multiplication with the corresponding linear transformations.

Anytime we introduce an algebraic operation such as this, we'd like to know what properties the operation may satisfy; and if we have multiple operations, we want to know how the different operations interact with one another.

In what's to follow we will use 0 to mean the zero matrix: the matrix of all zeros. It will usually be clear from context when we write 0 whether we're referring to the scalar number zero, or the zero matrix.

Theorem 6.1.
Let A, B, and C be $m \times n$ matrices and let λ and μ be scalars. Then matrix addition and scalar multiplication satisfy the following properties:

(i) $A + B = B + A$

(ii) $(A + B) + C = A + (B + C)$

(iii) $A + 0 = A$

(iv) $A - A = A + (-A) = 0$

(v) $\lambda(A + B) = \lambda A + \lambda B$

(vi) $(\lambda + \mu)A = \lambda A + \mu A$

(vii) $(\lambda\mu)A = \lambda(\mu A)$

Exercise 6.4.
Prove Theorem 6.1

6.4 Matrix Multiplication

If $f : A \to B$ and $g : B \to C$ is a map, then their ***composition*** is a map from A to C given by

$$a \mapsto g(f(a))$$

and denoted $g \circ f : A \to C$.

Consider linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^p$. Their composition, $S \circ T$, is a map from $\mathbb{R}^n$ to $\mathbb{R}^p$ which takes vectors in $\mathbb{R}^n$ and maps them into $\mathbb{R}^p$ according to the rule

$$\vec{v} \mapsto S(T(\vec{v})).$$

Notice that this is a linear transformation:

$$\begin{aligned}(S \circ T)(\vec{v} + \vec{w}) =& S(T(\vec{v} + \vec{w})) \\ =& S(T(\vec{v}) + T(\vec{w})) \\ =& S(T(\vec{v})) + S(T(\vec{w})) \\ =& (S \circ T)(\vec{v}) + (S \circ T)(\vec{w})\end{aligned}$$

$$\begin{aligned}(S \circ T)(\lambda\vec{v}) =& S(T(\lambda\vec{v})) \\ =& S(\lambda T(\vec{v})) \\ =& \lambda S(T(\vec{v})) \\ =& \lambda(S \circ T)(\vec{v}).\end{aligned}$$

Since $S \circ T : \mathbb{R}^n \to \mathbb{R}^p$ is a linear transformation it must be represented by some $p \times n$ matrix, the columns of which are given by $S(T(\vec{e}_i))$. To determine what these columns look like, suppose that T is represented by the $m \times n$ matrix A and S is represented by the $p \times m$ matrix B. Suppose the columns of A are $\vec{a}_1$, $\vec{a}_2$, ..., $\vec{a}_n$ and the columns of B are $\vec{b}_1$, $\vec{b}_2$, ..., $\vec{b}_m$. Then $S(T(\vec{e}_i)) = S(\vec{a}_i)$. Applying S corresponds to multiplying by the matrix B, however, so $S(\vec{a}_i) = B\vec{a}_i$. That is, the matrix representing $S \circ T$ has the form

$$\begin{pmatrix} B\vec{a}_1 & B\vec{a}_2 & \cdots & B\vec{a}_n \end{pmatrix}$$

If we suppose that $\vec{a}_i$ has the form

$$\vec{a}_i = \begin{pmatrix} a_{1,i} \\ a_{2,i} \\ \vdots \\ a_{m_i} \end{pmatrix}$$

then the i-th column of the matrix above is

$$B\vec{a}_i = a_{1,i}\vec{b}_1 + a_{2,i}\vec{b}_2 + \cdots + a_{m,i}\vec{b}_i$$

Supposing that the column $\vec{b}_j$ has the form

$$\vec{b}_j = \begin{pmatrix} b_{1,j} \\ b_{2,j} \\ \vdots \\ b_{p,j} \end{pmatrix}$$

We have

$$B\vec{a_i} = a_{1,i}\begin{pmatrix} b_{1,1} \\ b_{2,1} \\ \vdots \\ b_{p,1} \end{pmatrix} + a_{2,i}\begin{pmatrix} b_{1,2} \\ b_{2,2} \\ \vdots \\ b_{p,2} \end{pmatrix} + \cdots + a_{m,i}\begin{pmatrix} b_{1,m} \\ b_{2,m} \\ \vdots \\ b_{p,m} \end{pmatrix}$$

$$= \begin{pmatrix} a_{1,i}b_{1,1} + a_{2,i}b_{1,2} + \cdots + a_{m,i}b_{1,m} \\ a_{1,i}b_{2,1} + a_{2,i}b_{2,2} + \cdots + a_{m,i}b_{2,m} \\ \vdots a_{1,i}b_{p,1} + a_{2,i}b_{p,2} + \cdots + a_{m,i}b_{p,m} \end{pmatrix}$$

$$= \begin{pmatrix} b_{1,1}a_{1,i} + b_{1,2}a_{2,i} + \cdots + b_{1,m}a_{m,i} \\ b_{2,1}a_{1,i} + b_{2,2}a_{2,i} + \cdots + b_{2,m}a_{m,i} \\ \vdots \\ b_{p,1}a_{1,i} + b_{p,2}a_{2,i} + \cdots + b_{p,m}a_{m,i} \end{pmatrix}$$

Putting this all together, the entry in the i-th row and j-th column of our $p \times n$ matrix is

$$\sum_{k=1}^{m} b_{i,k}a_{k,j}.$$

This matrix is called the ***product*** of the matrices B and A and is denoted BA.

Just to reiterate: given two matrices A and B where A has size $m \times n$ and B has size $n \times p$, we can define the product AB which is a $m \times p$ matrix whose entry in the i-th row and j-th column is

$$\sum_{k=1}^{n} a_{i,k}b_{k,j}$$

This matrix corresponds to the composition of the linear transformations determined by A and B; applying B first and then A.

Notice that if A is $1 \times n$ and B is $n \times 1$, then this multiplication is easy to do:

$$\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1b_1 + a_2b_2 + a_3b_3 + \cdots + a_nb_n \end{pmatrix}$$

Example 6.9.

$$\begin{pmatrix} 3 & 7 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 4 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \cdot 2 + 7 \cdot (-1) + 2 \cdot 4 + 1 \cdot 0 \end{pmatrix} = \begin{pmatrix} 7 \end{pmatrix}$$

Since the product of a row vector and a column vector like this is always a 1×1 matrix it's just a single number, and so we usually think of this as being a scalar instead of a vector.

Remark.
If you've taken multivariable calculus, you might notice that multiplying a row vector and a column vector like this is the same as taking the dot product of two vectors.

The entry in the i-th row, j-th column of the product AB is obtained by multiplying the i-th row of A with the j-th column of B. This observation greatly simplifies the calculation of the product of two matrices.

Example 6.10.
Let A and B be the matrices below, and compute the product AB.

$$A = \begin{pmatrix} 4 & 6 & 3 \\ 0 & 1 & -1 \\ 3 & 2 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 3 & 1 \\ 1 & 2 \\ 0 & 2 \end{pmatrix}$$

The first row, first column of AB will be the product of the first row of A with the first column of B:

$$\begin{pmatrix} 4 & 6 & 3 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} = 4 \cdot 3 + 6 \cdot 1 + 3 \cdot 0 = 18$$

The first row, second column of AB will be the product of the first row of A and the second column of B:

$$\begin{pmatrix} 4 & 6 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = 4 \cdot 1 + 6 \cdot 2 + 3 \cdot 2 = 22$$

The second row, first column of AB is the product of the second row of A and the first column of B:

$$\begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} = 0 \cdot 3 + 1 \cdot 1 + (-1) \cdot 0 = 1$$

Continuing like this we can compute each entry of AB:

$$AB = \begin{pmatrix} 18 & 22 \\ 1 & 0 \\ 11 & 11 \end{pmatrix}$$

Example 6.11.
Suppose $T : \mathbb{R}^3 \to \mathbb{R}^2$ and $S : \mathbb{R}^2 \to \mathbb{R}^4$ are given by

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + z \\ y - z \end{pmatrix}$$

$$S \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2y \\ x + y \\ x - y \\ 3x \end{pmatrix}.$$

What does the composition $S \circ T$ do? What is the corresponding matrix?

Our map $S \circ T$ will take convert three-dimensional vectors into four-dimensional vectors in the following way:

$$\begin{aligned} S \circ T \begin{pmatrix} x \\ y \\ z \end{pmatrix} =& S \left(T \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right) \\ =& S \begin{pmatrix} x+z \\ y-z \end{pmatrix} \\ =& \begin{pmatrix} 2(y-z) \\ x+z+y-z \\ x+z-(y-z) \\ 3(x+z) \end{pmatrix} \\ =& \begin{pmatrix} 2y-2z \\ x+y \\ x-y+2z \\ 3x+3z \end{pmatrix} \end{aligned}$$

We could compute the matrix of $S \circ T$ in two different ways: by multiplying the matrices of S and T, or by computing $S \circ T(\vec{e}_i)$. We'll compute the matrix both ways.

First notice that the matrix of T is

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

The matrix of S is

$$A = \begin{pmatrix} 0 & 2 \\ 1 & 1 \\ 1 & -1 \\ 3 & 0 \end{pmatrix}.$$

The matrix of $S \circ T$ is thus

$$\begin{aligned} AB &= \begin{pmatrix} 0 & 2 \\ 1 & 1 \\ 1 & -1 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 0\cdot 1 + 2\cdot 0 & 0\cdot 0 + 2\cdot 1 & 0\cdot 1 + 2\cdot(-1) \\ 1\cdot 1 + 1\cdot 0 & 1\cdot 0 + 1\cdot 1 & 1\cdot 1 + 1\cdot(-1) \\ 1\cdot 1 + (-1)\cdot 0 & 1\cdot 0 + (-1)\cdot 1 & 1\cdot 1 + (-1)\cdot -1 \\ 3\cdot 1 + 0\cdot 0 & 3\cdot 0 + 0\cdot 1 & 3\cdot 1 + 0\cdot(-1) \end{pmatrix} \\ &= \begin{pmatrix} 0 & 2 & -2 \\ 1 & 1 & 0 \\ 1 & -1 & 2 \\ 3 & 0 & 3 \end{pmatrix} \end{aligned}$$

Just to confirm this is correct, notice

$$S \circ T(\vec{e}_1) = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 3 \end{pmatrix}$$

$$S \circ T(\vec{e}_2) = \begin{pmatrix} 2 \\ 1 \\ -1 \\ 0 \end{pmatrix}$$

$$S \circ T(\vec{e}_3) = \begin{pmatrix} -2 \\ 0 \\ 2 \\ 3 \end{pmatrix}$$

6.5 Properties of Matrix Multiplication

Before mentioning some of the algebraic properties that matrix multiplication satisfies, we mention some things about matrix multiplication which are very different when compared to the usual multiplication of real numbers that you're used to.

Notice that unlike multiplication of real numbers, multiplication of matrices is not commutative. That is, $AB \neq BA$ in general;

Example 6.12.
Suppose A and B are the matrices below:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & -1 \\ 4 & 2 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 1 & 2 \\ 3 & -1 & -2 \end{pmatrix}$$

Then

$$AB = \begin{pmatrix} 13 & -2 & -1 \\ -3 & 1 & 2 \\ 13 & -3 & 6 \end{pmatrix}$$

$$BA = \begin{pmatrix} 6 & 6 & 8 \\ 9 & 6 & 4 \\ -5 & 2 & 8 \end{pmatrix}$$

Also unlike normal multiplication of numbers, we can have two non-zero matrices that multiply to the zero matrix.

Example 6.13.

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

$$\begin{pmatrix} 1 & 2 & 0 & -2 & 4 \\ 0 & 1 & 0 & -1 & 1 \\ 2 & 1 & 1 & 1 & 7 \\ 0 & 2 & 0 & -2 & 2 \\ 3 & 2 & 3 & 4 & 14 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ 1 & -1 \\ 2 & 6 \\ 0 & -2 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

In general we can't "divide" matrices either. For example, if x, y and z are real numbers and $xy = xz$, then as long as $x \neq 0$ we can divide out the x's to conclude $y = z$. This is not the case for matrices.

Example 6.14.
Let A, B, and C be the matrices below.

$$A = \begin{pmatrix} 1 & 3 & 5 \\ 0 & 2 & 2 \\ 2 & 1 & 5 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 2 & 2 \\ 3 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \quad C = \begin{pmatrix} -1 & 0 & -2 \\ 2 & 0 & -3 \\ 2 & 1 & 3 \end{pmatrix}$$

Notice that

$$AB = AC = \begin{pmatrix} 15 & 5 & 4 \\ 8 & 2 & 0 \\ 10 & 5 & 8 \end{pmatrix}$$

even though $B \neq C$.

Now that we've seen some of the properties that matrix multiplication *doesn't* satisfy, let's mention some of the properties that *are* satisfied.

Theorem 6.2.
Let A, B, and C be matrices of the appropriate sizes so that products and sums below are defined, and let λ be a scalar.

(i) $A(BC) = (AB)C$

(ii) $A(B + C) = AB + AC$

(iii) $(A + B)C = AC + BC$

(iv) $\lambda(AB) = (\lambda A)B = A(\lambda B)$

The proofs of each property above are straight-forward, but slightly tedious, so we leave them as an exercise.

Exercise 6.5.
Prove Theorem 6.2.

6.6 The Transpose

Given any $m \times n$ matrix A, we can define an $n \times m$ matrix called the *transpose* of A and denoted A^T by swapping the rows and columns of A.

Example 6.15.

$$A = \begin{pmatrix} 1 & 2 & 7 & -3 & 2 & 4 \\ 4 & -2 & 1 & 1 & 0 & 2 \\ 3 & 1 & 2 & 2 & 2 & 2 \end{pmatrix}$$

$$A^T = \begin{pmatrix} 1 & 4 & 3 \\ 2 & -2 & 1 \\ 7 & 1 & 2 \\ -3 & 1 & 2 \\ 2 & 0 & 2 \\ 4 & 2 & 2 \end{pmatrix}$$

Notice that this operation turns row vectors into column vectors and vice versa:

Example 6.16.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 5 \\ 6 \\ 7 \\ 8 \end{pmatrix}^T = \begin{pmatrix} 5 & 6 & 7 & 8 \end{pmatrix}.$$

Remark.
If $\vec{v}$ is a column vector (an $n \times 1$ matrix), then the transpose $\vec{v}^T$ is a row vector (a $1 \times n$ matrix). The dot product of two vectors $\vec{u}$ and $\vec{v}$ can then be written as $\vec{u}^T\vec{v}$ where we perform matrix multiplication on the $1 \times n$ and $n \times 1$ vector to get a single number.

As the transpose is defined by exchanging the roles of columns and rows, if $\vec{a}_i$ is the i-th column of A, then $\vec{a}_i^T$ is the i-th row of A^T. Similarly, if $\vec{\alpha}_j$ is the j-th row of A (here $\vec{\alpha}_j$ is a row vector), then $\vec{\alpha}_j^T$ (now a column vector) is the j-th column of A.

Theorem 6.3.
If A is any matrix, then $(A^T)^T = A$.

Proof.

Suppose the columns of A are $\vec{a}_1$, $\vec{a}_2$, ..., $\vec{a}_n$:

$$A = \begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{pmatrix}$$

Then the transposes of those columns give the rows of A^T:

$$A^T = \begin{pmatrix} \vec{a}_1^T \\ \vec{a}_2^T \\ \vdots \\ \vec{a}_n^T \end{pmatrix}.$$

If we take the transpose yet again, then we turn these rows back into the original columns of A:

$$(A^T)^T = \begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{pmatrix} = A$$

□

Remark.
Notice that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is any linear transformation, say with corresponding matrix A, then the transpose of A determines a linear transformation $\mathbb{R}^m \to \mathbb{R}^n$ (in the reverse order of the original transformation T).

We now have four different operations we can perform on matrices: matrix addition, scalar multiplication, matrix multiplication, and now the transpose. It's reasonable to ask how our new operation, transpose, gets along with the previous operations.

Theorem 6.4.

Let A and B be matrices of the appropriate sizes so that the operations below are defined, and let λ be a scalar. We then have the following:

(i) $(A + B)^T = A^T + B^T$

(ii) $(\lambda A)^T = \lambda(A^T)$

(iii) $(AB)^T = B^T A^T$.

Proof of Theorem 8.3 (i).
We will only prove part (i) and leave the proofs of the other properties as exercises.

Suppose that A and B are both $m \times n$ matrices so that their sum is defined. Suppose that a_{ij} is the entry in the i-th row, j-th column of A and b_{ij} is the entry in the i-th row, j-th column of B. For simplicity let's refer to $A + B$ as C and say $c_{ij} = a_{ij} + b_{ij}$ is the entry in the i-th row, j-th column of $C = A + B$.

Notice that since the transpose reverses the roles of rows and columns, the entry in the i-th row, j-th column of A^T is a_{ji} (note i and j are in the reverse order), and similarly for B^T and C^T. Thus the i-th row, j-th column of C has entry c_{ji} which by definition is $a_{ji} + b_{ji}$, but this is the sum of what's in the i-th row and j-th column of A^T and B^T. Thus $C^T = A^T + B^T$. □

Exercise 6.6.
Prove parts (ii) and (iii) of Theorem 8.3.

6.7 Inverses

If $f : A \to B$ is a map, any map $g : B \to A$ which satisfies

$$g(f(a)) = a \text{ for every } a \in A \tag{6.1}$$

$$f(g(b)) = b \text{ for every } b \in B \tag{6.2}$$

is called an ***inverse*** of f. Two important properties of inverses are given by the following theorems:

Theorem 6.5.
A map $f : A \to B$ has an inverse if and only if f is a bijection.

Proof.
Suppose that f has an inverse: we suppose there is some map $g : B \to A$ satisfying the two equations above, and we need to show that f must be both surjective and injective. For surjectivity, let $b \in B$ and notice that there exists some element $a \in A$ that f sends to b: namely, take $a = g(b)$. By the second equation in the definition of an inverse we then have

$$f(a) = f(g(b)) = b$$

and so f is surjective.

For injectivity, suppose that there are elements $a, a' \in A$ such that $f(a) = f(a')$. If we then apply g to $f(a)$ and $f(a')$, however, we have

$$a = g(f(a)) = g(f(a')) = a',$$

and so f is injective.

Now we show the converse: suppose that f is bijective, and we want to show that f has an inverse. We define a map $g : B \to A$ by declaring that for each $b \in B$, $g(b) = a$ where $a \in A$ is the element that f sends to b. Such an a must exist since f is surjective, and a is unique because f is injective. So g is a well-defined map. Now we just need to verify that $f \circ g$ and $g \circ f$ satisfy the defining properties of an inverse, but this is almost obvious because of the way we defined g. By definition, $g(f(a))$ is the element in A which f sends to $f(a)$, but that is simply a and so $g(f(a)) = a$. For the second equation, notice that $g(b)$ is the element of A that f sends to b, so $f(g(b)) = b$. □

Theorem 6.6.
If f has an inverse (i.e., if f is bijective), then its inverse is unique. That is, there is only one map $g : B \to A$ satisfying Equations (6.1) and (6.2).

Proof.
To see this, suppose there were two different maps, say g_1 and g_2, satisfying the equations. We will show that g_1 and g_2 must in fact be the same map. Notice that

$$f(g_1(b)) = b = f(g_2(b)),$$

but f is injective so $g_1(b) = g_2(b)$. □

Since inverses are unique, we are justified in saying *the* inverse of a map instead of *an* inverse of a map. We adopt the notation f^{-1} to denote the inverse of f. Notice that this *is not* f raised to the negative first power; this is not one over f. (In fact, since we're just talking about sets that don't necessarily have a notation of any sort of "arithmetic" with their elements, this is a non-issue.)

Notice that Equations (6.1) and (6.2) imply the following:

$$f(a) = b \implies f^{-1}(b) = a$$
$$f^{-1}(b) = a \implies f(a) = b.$$

To see this, simply apply f^{-1} to both sides of $f(a) = b$ and then use Equation (6.1); and similarly apply f to both sides of $f^{-1}(b) = a$ and use Equation (6.2). This can be stated more simply as $f(a) = b$ if and only if $f^{-1}(b) = a$, which we can write symbolically as $f(a) = b \iff f^{-1}(b) = a$.

We can simplify Equations (6.1) and (6.2) by introducing the ***identity map***. For every set A, the identity map is a function from A to itself which fixes every element: that is, $a \mapsto a$ for every $a \in A$. The identity map is denoted id_A or id if the set A is clear from context.

Lemma 6.7.
Given any map $f : A \to B$, composing f with the identity map does not change f:

$$f \circ id_A = f = id_B \circ f.$$

Proof.
For every $a \in A$,

$$f(\mathrm{id}_A(a)) = f(a), \text{ and}$$
$$\mathrm{id}_B(f(a)) = f(a)$$

□

Equations (6.1) and (6.2) can then be expressed more tersely as

$$f \circ f^{-1} = f^{-1} \circ f = \mathrm{id}.$$

6.8 Inverse of a Matrix

Now suppose that $T : \mathbb{R}^n \to \mathbb{R}^n$ is a bijective linear map (i.e., the columns of the corresponding $n \times n$ matrix are linearly independent and span $\mathbb{R}^n$), and so has some inverse T^{-1}.

Lemma 6.8.
If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a bijective linear transformation, then its inverse T^{-1} is also linear.

Proof.
By definition, $T^{-1}(\vec{v_1}) = \vec{u_1}$ if $T(\vec{u_1}) = \vec{v_1}$. Also consider vectors $\vec{v_2}$ and $\vec{u_2}$ with $T^{-1}(\vec{v_2}) = \vec{u_2} \iff T(\vec{u_2}) = \vec{v_2}$. We need to show that $T^{-1}(\vec{v_1} + \vec{v_2}) = \vec{u_1} + \vec{u_2}$, but notice

$$T(\vec{u_1} + \vec{u_2}) = T(\vec{u_1}) + T(\vec{u_2}) = \vec{v_1} + \vec{v_2}.$$

Since T^{-1} is the inverse of T, $T^{-1}(\vec{v_1}+\vec{v_2})$ is the element of $\mathbb{R}^n$ which T takes to $\vec{v_1}+\vec{v_2}$, but we have just shown that $\vec{u_1}+\vec{u_2}$ is that element, and so

$$T^{-1}(\vec{v_1} + \vec{v_2}) = T^{-1}(\vec{v_1}) + T^{-1}(\vec{v_2}).$$

Similarly, suppose $T^{-1}(\vec{v}) = \vec{u}$. We need to show that $T^{-1}(\lambda\vec{v}) = \lambda\vec{u}$, but this must be the case as

$$T(\lambda\vec{u}) = \lambda T(\vec{u}) = \lambda\vec{v}.$$

Thus T^{-1} is linear. □

So if T is a linear bijection, then so is its inverse T^{-1}. Thus there is some matrix that corresponds to T^{-1}. To figure out what this matrix is, let's consider some properties of this matrix. First we need to know about the identity transformation and identity transformation.

The ***identity transformation*** is a map $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ which leaves vectors alone. That is, $\mathrm{id}(\vec{v}) = \vec{v}$. Notice that if $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear bijection, then Equations (6.1) and (6.2) can be rewritten as

$$T \circ T^{-1} = \mathrm{id} = T^{-1} \circ T.$$

Since id is a composition of linear maps, id is linear. (It's also very easy to check that id is a linear transformation directly.)

The matrix of the identity transformation is called the ***identity matrix*** and is denoted by I:

$$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

That is, the identity matrix for $\mathbb{R}^n$ is a square, $n \times n$ matrix that has 1's on the diagonal, and zeros everywhere else.

Sometimes we will write I_n to mean the $n \times n$ identity matrix, and sometimes we will just write I if the dimension is clear from context.

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Just as composing a map with the identity map doesn't change the map, multiplying a matrix with the identity matrix doesn't change the matrix:

$$AI = A = IA.$$

Say $T : \mathbb{R}^n \to \mathbb{R}^n$ is a bijective linear transformation with corresponding matrix A, and let A^{-1} denote the matrix of T. Since $T \circ T^{-1} = T^{-1} \circ T =$ id and since matrix multiplication corresponds to composition of linear transformations, we know

$$AA^{-1} = A^{-1}A = I.$$

Our goal is to determine what A^{-1} is given A. To do this we introduce elementary matrices.

6.9 Elementary Matrices

An ***elementary matrix*** is a matrix produced by performing an elementary row operation to the identity matrix. That is, an elementary matrix is a square matrix which is given by taking the identity matrix I and performing one of the following operations to it:

(i) Swap two rows.

(ii) Add a multiple of one row to another.

(iii) Multiply everything in one row by a constant.

Example 6.17.
The following are some 3×3 elementary matrices.

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} -4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Theorem 6.9.
If E is an elementary matrix, then the product EA is the same as performing the corresponding elementary row operation on A.

Before proving Theorem 8.6, let's recall a fact that was left as a practice problem in the previous lecture's notes. Suppose that A is any $m \times n$ matrix and B is any $n \times p$ matrix. Then the rows of AB are linear combinations of the rows of B where the scalars each row is multiplied by are determined by the entries in each row of A. For example, if the rows of B are the row vectors $\vec{r}_1$, $\vec{r}_2$, ..., $\vec{r}_n$ and the k-th row of A has the form

$$\begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_n \end{pmatrix},$$

then the k-th row of AB will be

$$\lambda_1 \vec{r}_1 + \lambda_2 \vec{r}_2 + \cdots + \lambda_n \vec{r}_n.$$

With this fact in hand, we can easily prove Theorem 8.6.

Proof of Theorem 8.6.
Let A be any $m \times n$ matrix whose rows we will suppose are the row vectors $\vec{r}_1$, $\vec{r}_2$, ..., $\vec{r}_m$. Let E be an elementary $m \times m$ matrix.

There are three cases to consider corresponding to the three elementary row operations.

- Suppose E is obtained from I_m by swapping two rows, say rows i and j:

$$I \xrightarrow{R_i \leftrightarrow R_j} E.$$

This means that the i-th row of E is $\vec{e}_j^T$, the j-th row of E is $\vec{e}_i^T$, and for any k that is not i or j, the k-th row of E is $\vec{e}_k^T$.

If k is not i or j, then the k-th row of EA is

$$\begin{aligned}&0 \cdot \vec{r}_1 + 0 \cdot \vec{r}_2 + \cdots + 0 \cdot \vec{r}_{k-1} + 1 \cdot \vec{r}_k + 0 \cdot \vec{r}_{k+1} + \cdots + 0 \cdot \vec{r}_m \\ =&\vec{r}_k\end{aligned}$$

So the k-th row of A remains the same. The i-th row of EA, however is

$$0 \cdot \vec{r}_1 + 0 \cdot \vec{r}_2 + \cdots + 0 \cdot \vec{r}_{j-1} + 1 \cdot \vec{r}_j + 0 \cdot \vec{r}_{j+1} + \cdots + 0 \cdot \vec{r}_m = \vec{r}_j$$

and the k-th row of EA is

$$0 \cdot \vec{r}_1 + 0 \cdot \vec{r}_2 + \cdots + 0 \cdot \vec{r}_{i-1} + 1 \cdot \vec{r}_i + 0 \cdot \vec{r}_{i+1} + \cdots + 0 \cdot \vec{r}_m = \vec{r}_i.$$

That is, every row of A is unchanged except for the i-th and j-th rows which are swapped.

- Suppose E is obtained from I_m by adding c times the i-th row to the j-th row. Then all rows of E, except for the j-th row, are all zeros except for a 1 in the k-th position of the k-th row. Thus every row of EA, except the j-th row, is the same as the corresponding row of A. The j-th row of E is all zeros except for a 1 in the j-th position and a c in the i-th position. Hence the j-th row of EA is the j-th row of A plus c times the i-th row of A.
- Left as an exercise.

□

Exercise 6.7.
Suppose that E is obtained from I by multiplying the i-th row of I by c. Show that EA is obtained from A by multiplying the i-th row of A by c.

The above theorem about elementary row operations will be combined with the following observation to obtain a method for determining inverse matrices.

Lemma 6.10.
If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear bijection, then the corresponding matrix A becomes the identity matrix when put into RREF.

Proof.
Since T is surjective, its matrix has a pivot in every row. However, since T is injective, it also has a pivot in every column. The only matrix in RREF with pivots in every row and column is the identity matrix. □

This lemma tells us that there is some sequence of elementary row operations that we can perform to A to get the identity matrix I. Each of these elementary row operations corresponds to multiplication by some elementary matrix. So there is some collection of elementary matrices, E_1, E_2, E_3, ..., E_q such that

$$E_q E_{q-1} \cdots E_2 E_1 A = I$$

The product $E_q E_{q-1} \cdots E_2 E_1$ is thus A^{-1}:

$$A^{-1} = E_q E_{q-1} \cdots E_2 E_1.$$

Example 6.18.
Let A be the following 2×2 matrix

$$\begin{pmatrix} 2 & 4 \\ 1 & 6 \end{pmatrix}$$

We can put this matrix into RREF with the following sequence of elementary row operations:

$$\begin{aligned}\begin{pmatrix}2 & 4\\1 & 6\end{pmatrix} &\xrightarrow{\frac{1}{2}R_1\to R_1}\begin{pmatrix}1 & 2\\1 & 6\end{pmatrix}\\ &\xrightarrow{R_2-R_1\to R_2}\begin{pmatrix}1 & 2\\0 & 4\end{pmatrix}\\ &\xrightarrow{\frac{1}{4}R_2\to R_2}\begin{pmatrix}1 & 2\\0 & 1\end{pmatrix}\\ &\xrightarrow{R_1-2R_2\to R_1}\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\end{aligned}$$

This corresponds to multiplying A by the following elementary matrices:

$$\begin{aligned}E_1 &= \begin{pmatrix}1/2 & 0\\0 & 1\end{pmatrix}\\ E_2 &= \begin{pmatrix}1 & 0\\-1 & 1\end{pmatrix}\\ E_3 &= \begin{pmatrix}1 & 0\\0 & 1/4\end{pmatrix}\\ E_4 &= \begin{pmatrix}1 & -2\\0 & 1\end{pmatrix}\end{aligned}$$

Now we multiply $E_4E_3E_2E_1$ to get A^{-1}:

$$A^{-1} = E_4E_3E_2E_1 = \begin{pmatrix}3/4 & -1/2\\-1/8 & 1/4\end{pmatrix}.$$

We can easily check that this really is the inverse of A: i.e., that $A^{-1}A = I$:

$$\begin{aligned}A^{-1}A &= \begin{pmatrix}3/4 & -1/2\\-1/8 & 1/4\end{pmatrix}\begin{pmatrix}2 & 4\\1 & 6\end{pmatrix}\\ &= \begin{pmatrix}3/4\cdot 2 + (-1/2)\cdot 1 & 3/4\cdot 4 + (-1/2)\cdot 6\\ -1/8\cdot 2 + 1/4\cdot 1 & -1/8\cdot 4 + 1/4\cdot 6\end{pmatrix}\\ &= \begin{pmatrix}3/2 - 1/2 & 3-3\\ -1/4+1/4 & -1/2+3/2\end{pmatrix}\\ &= \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\end{aligned}$$

There is a little trick we can use to make obtaining A^{-1} slightly easier. Suppose that A is $n \times n$ and consider the $n \times 2n$ matrix obtained by augmenting A with the $n \times n$ identity matrix:

$$\left(A \mid I\right).$$

We then start performing the elementary row operations that put A into RREF (this is the same as multiplying by E_1, then E_2, then E_3 and so on). Eventually, through some sequence of elementary row operations the above matrix becomes

$$\left(I \mid A^{-1}\right)$$

Example 6.19.
We compute the inverse of the matrix

$$A = \begin{pmatrix} 2 & 4 \\ 1 & 6 \end{pmatrix}$$

from the previous example using this trick.

$$\begin{aligned}
&\left(A \mid I\right) \\
&= \left(\begin{array}{cc|cc} 2 & 4 & 1 & 0 \\ 1 & 6 & 0 & 1 \end{array}\right) \\
\xrightarrow{\frac{1}{2}R_1 \to R_1} &\left(\begin{array}{cc|cc} 1 & 2 & 1/2 & 0 \\ 1 & 6 & 0 & 1 \end{array}\right) \\
\xrightarrow{R_2 - R_1 \to R_2} &\left(\begin{array}{cc|cc} 1 & 2 & 1/2 & 0 \\ 0 & 4 & -1/2 & 1 \end{array}\right) \\
\xrightarrow{\frac{1}{4}R_2 \to R_2} &\left(\begin{array}{cc|cc} 1 & 2 & 1/2 & 0 \\ 0 & 1 & -1/8 & 1/4 \end{array}\right) \\
\xrightarrow{R_1 - 2R_2 \to R_1} &\left(\begin{array}{cc|cc} 1 & 0 & 3/4 & -1/2 \\ 0 & 1 & -1/8 & 1/4 \end{array}\right)
\end{aligned}$$

The inverse matrix A^{-1} is now the right-hand side of this augmented matrix,

$$A^{-1} = \begin{pmatrix} 3/4 & -1/2 \\ -1/8 & 1/4 \end{pmatrix},$$

which agrees with our previous calculation.

This trick for computing the inverse works for any square matrix which is invertible, regardless of the size, though the work certainly gets more tedious as we consider larger and larger matrices.

Example 6.20.
Compute the inverse of the following matrix:

$$A = \begin{pmatrix} -2/3 & 1 & 0 \\ 1 & -1 & 0 \\ 4 & -6 & 1 \end{pmatrix}$$

$$\begin{aligned}
(A \mid I) &\xrightarrow{R_1 \leftrightarrow R_2} \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 0 & 1 & 0 \\ -2/3 & 1 & 0 & 1 & 0 & 0 \\ 4 & -6 & 1 & 0 & 0 & 1 \end{array}\right) \\
&\xrightarrow{R_2 + \frac{2}{3}R_1 \to R_2} \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 1/3 & 0 & 1 & 2/3 & 0 \\ 4 & -6 & 1 & 0 & 0 & 1 \end{array}\right) \\
&\xrightarrow{R_3 - 4R_1 \to R_3} \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 1/3 & 0 & 1 & 2/3 & 0 \\ 0 & -2 & 1 & 0 & -4 & 1 \end{array}\right) \\
&\xrightarrow{3R_2 \to R_2} \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 3 & 2 & 0 \\ 0 & -2 & 1 & 0 & -4 & 1 \end{array}\right) \\
&\xrightarrow{R_1 + R_2 \to R_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & 3 & 0 \\ 0 & 1 & 0 & 3 & 2 & 0 \\ 0 & -2 & 1 & 0 & -4 & 1 \end{array}\right) \\
&\xrightarrow{R_3 + 2R_2 \to R_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & 3 & 0 \\ 0 & 1 & 0 & 3 & 2 & 0 \\ 0 & 0 & 1 & 6 & 0 & 1 \end{array}\right)
\end{aligned}$$

Hence

$$A^{-1} = \begin{pmatrix} 3 & 3 & 0 \\ 3 & 2 & 0 \\ 6 & 0 & 1 \end{pmatrix}$$

Notice that this only tells us A^{-1} if the RREF of A is the identity matrix. If the RREF of A *is not* the identity matrix, then A does not have an inverse.

In the special case of 2×2 matrices there is a quick and easy formula for the derivative.

Theorem 6.11.
If A is a 2×2 matrix of the form

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then A is invertible if and only if $ad - bc \neq 0$, and the inverse of A is

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Proof.
First suppose that A is invertible, so the RREF of A is the identity

matrix. We then proceed to calculate A^{-1} as above:

$$
\begin{aligned}
& \left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right) \\
\xrightarrow{\frac{1}{a}R_1 \to R_1} & \left(\begin{array}{cc|cc} 1 & b/a & 1/a & 0 \\ c & d & 0 & 1 \end{array}\right) \\
\xrightarrow{R_2 - cR_1 \to R_2} & \left(\begin{array}{cc|cc} 1 & b/a & 1/a & 0 \\ 0 & (ad-bc)/a & -c/a & 1 \end{array}\right) \\
\xrightarrow{\frac{a}{ad-bc}R_2 \to R_2} & \left(\begin{array}{cc|cc} 1 & b/a & 1/a & 0 \\ 0 & 1 & -c/(ad-bc) & a/(ad-bc) \end{array}\right) \\
\xrightarrow{R_1 - \frac{b}{a}R_2 \to R_1} & \left(\begin{array}{cc|cc} 1 & 0 & 1/a + bc/a(ad-bc) & -b/(ad-bc) \\ 0 & 1 & -c/(ad-bc) & a/(ad-bc) \end{array}\right) \\
= & \left(\begin{array}{cc|cc} 1 & 0 & (ad-bc+bc)/a(ad-bc) & -b/a \\ 0 & 1 & -c/(ad-c) & a/(ad-bc) \end{array}\right) \\
= & \left(\begin{array}{cc|cc} 1 & 0 & d/(ad-bc) & -b/a \\ 0 & 1 & -c/(ad-bc) & a/(ad-bc) \end{array}\right)
\end{aligned}
$$

Thus

$$
A^{-1} = \begin{pmatrix} d/(ad-bc) & -b/a \\ -c/(ad-bc) & a/(ad-bc) \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}
$$

This of course implies that $ad - bc \neq 0$ since if it were the above expression for A^{-1} would be undefined.

Now suppose that $ad - bc \neq 0$. The calculation above shows us that the RREF of A is the identity, and further tells us what A^{-1} is, so A must be invertible. □

6.10 Properties of Inverses

Now that we know how to compute the inverse of a matrix, we turn our attention to some properties that inverses satisfy.

Lemma 6.12.
If A is an invertible matrix and if B and C are matrices satisfying that

$AB = AC$, then $B = C$.

Proof.

$$\begin{aligned} &AB = AC \\ \implies &A^{-1}AB = A^{-1}AC \\ \implies &IB = IC \\ \implies &B = C \end{aligned}$$

□

As we saw in the last lecture, the above lemma is false if A is not invertible.

Lemma 6.13.
If A and B are invertible $n \times n$ matrices and if $AB = BA = I$, then $A^{-1} = B$ and $B^{-1} = A$.

Exercise 6.8.
Prove Lemma 6.13.

Theorem 6.14.

(i) *If A is an invertible matrix, then so is A^{-1} and $(A^{-1})^{-1} = A$.*

(ii) *If A and B are both invertible matrices of the same size, then their product AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.*

(iii) If A is an invertible matrix, then so is A^T and $\left(A^T\right)^{-1} = (A^{-1})^T$.

Proof.

(i) For the moment let C denote the matrix $(A^{-1})^{-1}$. Then $A^{-1}C = CA^{-1} = I$. But we know that $A^{-1}A = AA^{-1} = I$ and so by Lemma 6.13, $C = A$.

(ii) Simply notice that

$$\begin{aligned} B^{-1}A^{-1}AB &= B^{-1}(A^{-1}A)B \\ &= B^{-1}IB \\ &= B^{-1}B \\ &= I \end{aligned}$$

(iii) Recall that $(AB)^T = B^T A^T$ and notice that $I^T = I$

$$\begin{aligned} &AA^{-1} = I \\ \implies &\left(AA^{-1}\right)^T = I^T = I \\ \implies &\left(A^{-1}\right)^T A^T = I \end{aligned}$$

but this implies that the inverse of A^T is $(A^{-1})^T$.

□

The following theorem tells us there are several different ways to think about invertible matrices. Some of the items in the theorem we have already seen, but we list them in this theorem so that we will have a single theorem to refer to that characterizes invertible matrices.

Theorem 6.15.
Let A be an $n \times n$ square matrix. The following are equivalent:

(a) *A is an invertible matrix.*

(b) *The RREF of A is the identity matrix.*

(c) *A has n pivots.*

(d) *A has a pivot in every row.*

(e) *A has a pivot in every column.*

(f) *The only solution to the homogeneous equation $A\vec{x} = \vec{0}$ is the trivial solution.*

(g) *The columns of A are linearly independent.*

(h) *The columns of A span $\mathbb{R}^n$.*

(i) *The equation $A\vec{x} = \vec{b}$ has one solution for every $b \in \mathbb{R}^n$.*

(j) *The linear transformation $\vec{x} \mapsto A\vec{x}$ is injective.*

(k) *The linear transformation $\vec{x} \mapsto A\vec{x}$ is surjective.*

(l) *There is an $n \times n$ matrix B so that $AB = I$.*

(m) *There is an $n \times n$ matrix B so that $BA = I$.*

(n) *A^T is invertible.*

Proof.
We have actually already shown everything in this theorem, but sometimes using the language of systems of linear equations or linear transformations instead of matrices. See pg. 112 of Lay for details. □

6.11 Practice Problems

Problem 6.1.
Let A, B, C, D, E, and F be the matrices given below.

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & -1 \\ 1 & 1 & -3 \end{pmatrix} \qquad B = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 2 & 3 & 3 & -1 \\ -4 & 0 & 0 & 1 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 2 & -3 \\ 3 & 4 & 5 \\ 1 & 2 & 3 \end{pmatrix} \qquad D = \begin{pmatrix} -1 & 2 \\ 4 & 1 \\ 0 & 1 \end{pmatrix}$$

$$E = \begin{pmatrix} 5 & 2 \\ 1 & 4 \\ 3 & -3 \end{pmatrix} \qquad F = \begin{pmatrix} 4 & 0 & 0 & 1 \\ 2 & 1 & 3 & 1 \\ 1 & -1 & -1 & 2 \end{pmatrix}$$

For each of the problems below, first determine if the operations are defined. If so, compute the described matrix.

(a) $AB + F$

(b) $2D - 3E$

(c) AC

(d) $4AD + E$

(e) FBC

(f) BCF

(g) $AF - DE$

Problem 6.2.
Compute the transposes of the matrices A through F mentioned in Problem 1.

Problem 6.3.
Suppose that A is the following matrix

$$A = \begin{pmatrix} 1 & -3 \\ -3 & 5 \end{pmatrix},$$

and B is some 2×2 matrix such that

$$AB = \begin{pmatrix} -3 & -1 \\ 1 & 17 \end{pmatrix}.$$

Determine B.

Problem 6.4.
Let $\vec{r}$ be an m-dimensional row vector,

$$\vec{r} = \begin{pmatrix} r_1 & r_2 & \cdots & r_m \end{pmatrix},$$

and let A be an $m \times n$ matrix. Describe the product $\vec{r}A$ (this will be an n-dimensional row vector) as a linear combination of the rows of A.

Problem 6.5.
Let A be an $m \times n$ matrix and B an $n \times p$ matrix.

(a) Show that each column of AB is a linear combination of the columns of A. (Hint: Suppose the columns of B are $\vec{b}_1$, $\vec{b}_2$, ..., $\vec{b}_p$. Show that the i-th column of the product AB is given by $A\vec{b}_i$.)

(b) Show that each row of AB is a linear combination of the rows of B. (Hint: Suppose the rows of A are row vectors $\vec{a}_1$, $\vec{a}_2$, ..., $\vec{a}_m$. Show that the i-th row of AB is given by $\vec{a}_i B$.)

Problem 6.6.
Suppose A is the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 9 \\ 6 & 7 & 13 \end{pmatrix}$$

Find a 3×3 matrix B *which is not all zeros* so that AB is the zero matrix. (Hint: Use 5a to help you determine the matrix.)

Problem 6.7.
For each matrix below, determine if the matrix is invertible. If so, calculate the inverse.

(a) $\begin{pmatrix} 5 & 2 \\ 3 & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & 3 \\ -3 & 4 \end{pmatrix}$

(c) $\begin{pmatrix} 6 & 8 \\ 3 & 4 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & -3 & -6 \\ 0 & 4 & 3 \\ -3 & 6 & 0 \end{pmatrix}$

(e) $\begin{pmatrix} 3 & 4 & 7 & 4 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & 2 & 8 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

Problem 6.8.
Can a square matrix with two identical columns be invertible? Explain why not, or give an example of an invertible matrix with two identical columns.

Problem 6.9.
Can a matrix with a column of all zeros be invertible? Explain why not, or give an example of an invertible matrix with a column of all zeros.

Problem 6.10.
Show that if A is an $n \times n$ matrix that is not invertible, then it must have a column that can be written as a linear combination of the other columns.

Problem 6.11.
Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the following linear transformation:

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x - y \\ x + 3y \end{pmatrix}.$$

Show that T is an isomorphism and find its inverse.

7 Special Matrices and Factorizations

Mathematics, rightly viewed, possesss not only truth, but supreme beauty.

BERTRAND RUSSELL

In this lecture we introduce some terminology and mention some special types of matrices, as well as discuss a useful method of factoring some matrices.

7.1 Square, Diagonal, and Symmetric Matrices

We say a matrix A is a ***square matrix*** if it has just as many rows as columns. Some properties we are interested in, such as invertibility, require we have a square matrix.

A square matrix A is called a ***diagonal matrix*** if its only non-zero entries are on the ***diagonal*** of the matrix.

Example 7.1.
The following matrices are diagonal.

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

A square matrix is called ***symmetric*** if it equals its transpose: $A^T = A$. That is, we can reflect all of the entries in the matrix about the diagonal and get back the same matrix.

Example 7.2.

The following matrix is symmetric:

$$\begin{pmatrix} 1 & 2 & 7 & -4 \\ 2 & 0 & 3 & 1 \\ 7 & 3 & 5 & -1 \\ -4 & 1 & -1 & 2 \end{pmatrix}$$

7.2 Partitioned Matrices

For some types of computational problems it is necessary or convenient to take a large matrix and split it up into smaller submatrices. When we fill a matrix with small submatrices, we say the original, large matrix is a ***partitioned matrix***. Equivalently, we tkae our large matrix and draw a series of vertical and horizontal lines to separate the matrix into regions. We then imagine the larger matrix as being a matrix of matrices.

Example 7.3.
Suppose A, B, C, and D are the matrices below.

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 3 & 0 & 1 \\ -1 & 2 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix}$$

$$C = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -2 & -1 \end{pmatrix} \qquad D = \begin{pmatrix} 3 \\ -7 \end{pmatrix}$$

Then we can put these matrices together to get a larger, partitioned matrix:

$$P = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

This simply means the matrix

$$\begin{pmatrix} 4 & 2 & 1 & 0 \\ 3 & 0 & 1 & 2 \\ -1 & 2 & 2 & -1 \\ 2 & 1 & 1 & 3 \\ 0 & -2 & -1 & -7 \end{pmatrix}$$

Our original matrices A, B, C, and D are submatrices of P obtained by diving P as indicated below.

$$\left(\begin{array}{ccc|c} 4 & 2 & 1 & 0 \\ 3 & 0 & 1 & 2 \\ -1 & 2 & 2 & -1 \\ \hline 2 & 1 & 1 & 3 \\ 0 & -2 & -1 & -7 \end{array}\right)$$

One practical reason why you might want to partition a matrix is to make it easier to do computations. In particular, if you were a computer programmer writing some program that had to add, subtract, and multiply very large matrices it may be computationally advantage (i.e., speed up how quickly your program runs) if you can chop your matrices up into smaller pieces and do computations on those. If the matrices you were interested in were *very* large, you might not even be able to store the entire matrix in memory at one time, and but could still be able to store smaller submatrices by partitioning.

Scalar multiplication of partitioned matrices is as simple as you could hope it would be: If A is a matrix partitioned as indicated below,

$$A = \left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1p} \\ \hline A_{21} & A_{22} & \cdots & A_{2p} \\ \hline & & \ddots & \\ \hline A_{q1} & A_{q2} & \cdots & A_{qp} \end{array}\right)$$

then λA is

$$\lambda A = \left(\begin{array}{c|c|c|c} \lambda A_{11} & \lambda A_{12} & \cdots & \lambda A_{1p} \\ \hline \lambda A_{21} & \lambda A_{22} & \cdots & \lambda A_{2p} \\ \hline & & \ddots & \\ \hline \lambda A_{q1} & \lambda A_{q2} & \cdots & \lambda A_{qp} \end{array}\right)$$

For matrix addition we have to be slightly more careful. If A and B are partitioned matrices of the same size *which are partitioned in the same way*, meaning they are broken into submatrices of the same size in the same way, then we can add A and B by adding their submatrices: If

$$A = \left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1p} \\ \hline A_{21} & A_{22} & \cdots & A_{2p} \\ \hline & & \ddots & \\ \hline A_{q1} & A_{q2} & \cdots & A_{qp} \end{array}\right) \qquad B = \left(\begin{array}{c|c|c|c} B_{11} & B_{12} & \cdots & B_{1p} \\ \hline B_{21} & B_{22} & \cdots & B_{2p} \\ \hline & & \ddots & \\ \hline B_{q1} & B_{q2} & \cdots & B_{qp} \end{array}\right)$$

then

$$A+B=\left(\begin{array}{c|c|c|c} A_{11}+B_{11} & A_{12}+B_{12} & \cdots & A_{1p}+B_{1p} \\ \hline A_{21}+B_{21} & A_{22}+B_{22} & \cdots & A_{2p}+B_{2p} \\ \hline & & \ddots & \\ \hline A_{q1}+B_{q1} & A_{q2}+B_{q2} & \cdots & A_{qp}+B_{qp} \end{array}\right).$$

For multiplication we have to be very careful. We can multiply two partitioned matrices by multiplying their submatrices, but the matrices have to partitioned in such a way that the products of submatrices are defined. That is, if

$$A=\left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1p} \\ \hline A_{21} & A_{22} & \cdots & A_{2p} \\ \hline & & \ddots & \\ \hline A_{q1} & A_{q2} & \cdots & A_{qp} \end{array}\right) \qquad B=\left(\begin{array}{c|c|c|c} B_{11} & B_{12} & \cdots & B_{1t} \\ \hline B_{21} & B_{22} & \cdots & B_{2t} \\ \hline & & \ddots & \\ \hline B_{p1} & B_{p2} & \cdots & B_{pt} \end{array}\right)$$

then the product AB is given by

$$\begin{aligned} AB \\ =\left(\begin{array}{c|c|c|c} A_{11}B_{11}+A_{12}B_{21}+\cdots+A_{1p}B_{p1} & A_{11}B_{12}+A_{12}B_{22}+\cdots+A_{1p}B_{p2} & \cdots & A_{11}B_{1t}+A_{12}B_{2t}+\cdots+A_{1p}B_{pt} \\ \hline A_{21}B_{11}+A_{22}B_{21}+\cdots+A_{2p}B_{p1} & A_{21}B_{12}+A_{22}B_{22}+\cdots+A_{2p}B_{p2} & \cdots & A_{21}B_{1t}+A_{22}B_{2t}+\cdots+A_{2p}B_{pt} \\ \hline & & \ddots & \\ \hline A_{q1}B_{11}+A_{q2}B_{21}+\cdots+A_{qp}B_{p1} & A_{q1}B_{12}+A_{q2}B_{22}+\cdots+A_{qp}B_{p2} & \cdots & A_{q1}B_{1t}+A_{q2}B_{2t}+\cdots+A_{qp}B_{pt} \end{array}\right) \end{aligned}$$

provided that all products $A_{ij}B_{k\ell}$ are defined. If all of these products are defined, then we say partitioned matrices A and B are ***conformable*** to multiplication.

Example 7.4.
Suppose A and B are the matrices partitioned by

$$A=\left(\begin{array}{cc|c} 1 & 3 & 2 \\ 0 & 1 & -1 \\ \hline 4 & 2 & 1 \end{array}\right) \qquad B=\left(\begin{array}{cc} -2 & 2 \\ 1 & 2 \\ \hline 4 & 1 \end{array}\right)$$

Here

$$A_{11}=\begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} \qquad A_{12}=\begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

$$A_{21}=\begin{pmatrix} 4 & 2 \end{pmatrix} \qquad A_{22}=\begin{pmatrix} 1 \end{pmatrix}$$

7.3 Factorization

Recall that *factoring* is the process of breaking an object up into pieces which multiply together to give the original object. For example, every natural number has a prime factorization: we can write any natural number as a product of prime numbers. We can also factor polynomials, writing a larger, complicated polynomial as a product of simpler ones. Similarly, we can factor matrices and write a given matrix as a product of simpler ones. Having such a factorization has important computational applications as certain operations, such as solving large systems of equations, become computationally easier if we have a factorization.

7.4 Triangular Matrices

We will try to factor a matrix into special types of matrices called *triangular matrices*. A ***lower triangular matrix*** is a square matrix whose only non-zero entries occur on or below the diagonal.

Example 7.5.
The following are lower triangular matrices.

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 2 & 4 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ -1 & 3 & 2 & 0 & 0 \\ 4 & 1 & 7 & 3 & 0 \\ 2 & 3 & 4 & 1 & 3 \end{pmatrix}$$

An ***upper triangular matrix*** is a square matrix whose only non-zero entries occur on or above the diagonal.

Example 7.6.

Each of the following matrices are upper triangular.

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 1 \\ 0 & 0 & 6 \end{pmatrix} \qquad \begin{pmatrix} 1 & 3 & 4 & 7 \\ 0 & 2 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

Triangular matrices have two nice properties that we can take advantage of. Because about "half" the entries in a triangular matrix are zeros, multiplying triangular matrices can be done a lot quicker than multiplying non-triangular matrices.

In particular, if you multiply two $n \times n$ matrices together, each entry of the matrix requires that you perform n multiplications and $n - 1$ additions, so each entry requires you do $2n - 1$ arithmetic operations. But there are n^2 entries in the matrix, so you have to perform $2n^3 - n^2$ operations. If you multiplied triangular matrices, however, you can cut your work down considerably because you know many of those multiplications will yield zeros, so you don't need to bother doing all of the work.

Say you wanted to multiply a lower triangular matrix L with an upper triangular matrix U, both of which are $n \times n$. The entry in the i-th row and j-th column of the product is

$$\sum_{k=1}^{n} \ell_{ik} u_{kj},$$

but $\ell_{ik} = 0$ if $k > i$, and $u_{kj} = 0$ if $k > j$. So instead of looking at all of the entries in each row and each column, we can stop once we reach a point where we're just adding up zeros. That is, the entry in the i-th row and j-th column of the product LU is

$$\sum_{k=1}^{\min(i,j)} \ell_{ik} u_{kj}.$$

Adding up the total number of operations we have to perform, we have to do $\frac{2}{3}n^3 - n^2 - \frac{2}{3}n$ operations, so we do about one third of the work to multiply a lower and upper triangular matrix than we have to do to multiply two general $n \times n$ matrices.

The other nice thing about triangular matrices is that, if they represent the coefficient matrix of a system of equations we're interested in, then we can perform back substitution to quickly solve the system. That is, if U is an upper triangular matrix, then we can solve the equation $U\vec{x} = \vec{b}$ by solving for each component of $\vec{x}$ one at a time from the bottom to the top. (For a lower triangular matrix L, we solve $L\vec{x} = \vec{b}$ one component at a time from the top to the bottom.) This is reminiscent of how we had solved systems of equations earlier in the semester by putting a matrix into echelon form.

Proposition 7.1.
If L_1 and L_2 are lower triangular $n \times n$ matrices, then their product $L_1 L_2$ is also lower triangular.

Proof.
Let the entry in the i-th row and j-th column of L_1 be denoted ℓ_{ij}, and the corresponding entry in L_2 will be denoted λ_{ij}. The entries of $L_1 L_2$ are thus

$$\sum_{k=1}^{n} \ell_{ik}\lambda_{kj}$$

Notice $\ell_{ik} = 0$ if $k > i$, and $\lambda_{kj} = 0$ if $j > k$. If $j > i$ (that is, we're above the diagonal), then each λ_{1j}, λ_{2j}, ..., $\lambda_{j-1,j}$ equals zero, and each $\ell_{i,j}$, $\ell_{i,j+1}$, ..., $\ell_{i,n}$ equals zero as well. So every term above the diagonal is zero, and we have a lower triangular matrix. □

Exercise 7.1.
Let U_1 and U_2 be two upper diagonal matrices, both of which are $n \times n$. Show that their product $U_1 U_2$ is also an upper triangular matrix.

Proposition 7.2.
A lower triangular matrix is invertible if and only if it does not contain any zeros on the diagonal.

Proof.
Suppose L is invertible, so there is a matrix L^{-1} with $LL^{-1} = I$. This means the j-th column of I equals L times the j-th column of L^{-1}. In particular, we can write the j-th column of I, which is all zeroes except for a 1 in the j-th row, as a linear combination of the columns of L. Notice that if the first entry in the j-th column of L^{-1} *is not* a zero, then we will have whatever occurs in the first row of the first column of L (which is non-zero by the exercise below). Similarly for the second, third, and fourth entries, and so on. Hence the first $j-1$ entries of the j-th column of L^{-1} must be zero. □

Exercise 7.2.
Prove the second half of Proposition 7.2: show that if L is a lower triangular matrix and does not have zeros on its diagonal, then L is invertible.

Proposition 7.3.
If L is a lower triangular matrix and is invertible, then so is L^{-1}.

Proof.
Since L is invertible, its RREF is the identity matrix. To get the identity we do row operations, and the product of the corresponding elementary matrices is the inverse. But since L is lower triangular,

we only need to perform row operations which divide the diagonal entries to make them 1 (each such elementary row matrix is diagonal) and then zero out the entries below the diagonal (each of these elementary matrices is lower triangular). Hence L^{-1} is a product of lower triangular matrices, and must be lower triangular. □

7.5 The LU Factorization

We will now describe a procedure for taking a matrix A and factoring it as $A = LU$ where L is lower triangular and U is upper triangular. For simplicity, first suppose that A can be put into echelon form *without* swapping two rows. That is, we perform some sequence of elementary row operations to turn A into an upper triangular matrix. Each of these elementary row operations corresponds to multiplying A on the left by an elementary matrix, and so we have

$$E_pE_{p-1}\cdots E_2E_1A = U.$$

Since we're only putting A into echelon form, not row-reduced echelon form, and since we are assuming no rows are swapped, each elementary E_i is a lower triangular matrix, so the product $E_pE_{p-1}\cdots E_2E_1$ is lower triangular. Call this matrix M for the moment. We then have $MA = U$ where M is lower triangular and A is upper triangular. Multiplying both sides by M^{-1} we have $A = M^{-1}U$; notice that M^{-1} is lower triangular since M is lower triangular. Letting $L = M^{-1}$, we have the ***LU factorization***, $A = LU$ where L is a lower triangular matrix and U is an upper triangular matrix.

Example 7.7.
Find the LU factorization of the following matrix:

$$A = \begin{pmatrix} 3 & 2 & 1 & 4 \\ 6 & 5 & 3 & 10 \\ 3 & 12 & 4 & 5 \\ 0 & 3 & 3 & 15 \end{pmatrix}$$

First we perform elementary row operations to put A into echelon form, keeping track of the operations:

$$\begin{pmatrix} 3 & 2 & 1 & 4 \\ 6 & 5 & 3 & 10 \\ 3 & 12 & 4 & 5 \\ 0 & 3 & 3 & 15 \end{pmatrix} \xrightarrow{R_2 - 2R_1 \to R_2} \begin{pmatrix} 3 & 2 & 1 & 4 \\ 0 & 1 & 1 & 2 \\ 3 & 12 & 4 & 5 \\ 0 & 3 & 3 & 15 \end{pmatrix}$$

$$\xrightarrow{R_3 - R_1 \to R_3} \begin{pmatrix} 3 & 2 & 1 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 10 & 3 & 1 \\ 0 & 3 & 3 & 15 \end{pmatrix}$$

$$\xrightarrow{R_3 - 10R_2 \to R_3} \begin{pmatrix} 3 & 2 & 1 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -7 & -19 \\ 0 & 3 & 3 & 15 \end{pmatrix}$$

$$\xrightarrow{R_4 - 3R_2 \to R_4} \begin{pmatrix} 3 & 2 & 1 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -7 & -19 \\ 0 & 0 & 0 & 9 \end{pmatrix}$$

The matrix we have left over is our U. To get L we need to multiply the elementary matrices corresponding to the above row operations, and then take the inverse of the product.

$$E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$E_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$E_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -10 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$E_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -3 & 0 & 1 \end{pmatrix}$$

Now we multiply all of these matrices together,

$$E_4E_3E_2E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 19 & -10 & 1 & 0 \\ 6 & -3 & 0 & 1 \end{pmatrix}$$

And our L will be the inverse of this matrix:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 19 & -10 & 1 & 0 \\ 6 & -3 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1 & 10 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{pmatrix}$$

And this is our LU-factorization of the original matrix:

$$A = \begin{pmatrix} 3 & 2 & 1 & 4 \\ 6 & 5 & 3 & 10 \\ 3 & 12 & 4 & 5 \\ 0 & 3 & 3 & 15 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1 & 10 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 & 1 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -7 & -19 \\ 0 & 0 & 0 & 9 \end{pmatrix} = LU$$

The above discussion was heavily contingent on our being able to put A into echelon form *without* exchanging any two rows. In general, however, we will need to exchange rows to put a given matrix in echelon form. Thus our LU factorization requires a slight change. What we can do is first perform a series of row operations, basically doing all of the necessary exchanges, to turn A into a matrix which can then be put into echelon form without any additional row exchanges. Let's suppose we have to perform row exchanges corresponding to elementary matrices E_1, E_2, ..., E_k, and then let P be their product, $P = E_k \cdots E_1$. So PA is a matrix where we can perform the LU factorization above. This is sometimes called the ***LUP factorization*** of A, but it means we have

$$PA = LU.$$

or

$$A = P^{-1}LU.$$

It can be shown that for every matrix P formed by multiplying elementary matrices which correspond to exchanging rows, such as the P above, $P^{-1} = P$, so we can rewrite the above as

$$A = PLU.$$

7.6 LU Factorization for non-square matrices

We can define an LU factorization for matrices A which are not square similarly to how we defined the LU factorization for square matrices. Instead of taking U to be a square matrix, however, we simply take U to be an echelon form of A. The matrix L *is* still square, however. Otherwise everything is exactly the same.

7.7 Practice Problems

Problem 7.1.
Suppose that the following partitioned matrix is given:

$$\left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline A & I & 0 \\ \hline B & D & I \end{array}\right)$$

and the inverse of this matrix is the block matrix

$$\left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline P & I & 0 \\ \hline Q & R & I \end{array}\right).$$

Write P, Q, and R in terms of A, B, and D.

Problem 7.2.
A ***block upper triangular matrix*** is a partitioned matrix where all the squares below the diagonal are zero matrices. For example,

$$\left(\begin{array}{ccc|cc} 4 & 3 & 2 & 1 & 4 \\ 2 & 1 & 1 & 3 & 1 \\ 1 & 0 & 0 & 2 & 1 \\ \hline 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 2 & 4 \end{array}\right).$$

Notice that the "larger" matrix may not be triangular, but when we think of this matrix as being divided into submatrices,

$$\left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array}\right).$$

this partitioned matrix is triangular.

Show that an upper block triangular matrix whose diagonal submatrices are all squares will be invertible if and only if each square matrix on the diagonal is invertible.

Problem 7.3.
Compute the inverse of the following matrix by first rewriting the matrix as a block upper triangular matrix.

$$\begin{pmatrix} 2 & 4 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 2 & 5 \end{pmatrix}$$

Problem 7.4.
Compute the LU factorization of the following matrices.

(a)
$$\begin{pmatrix} 2 & 6 \\ 4 & 7 \end{pmatrix}$$

(b)
$$\begin{pmatrix} 2 & -4 & 2 \\ -4 & 5 & 2 \\ 6 & -9 & 1 \end{pmatrix}$$

(c)
$$\begin{pmatrix} 1 & -2 & -2 & -3 \\ 3 & -9 & 0 & -9 \\ -1 & 2 & 4 & 7 \\ -3 & -6 & 26 & 2 \end{pmatrix}$$

Problem 7.5.
When a matrix A has an LU factorization, it is often computationally easier to solve the system $A\vec{x} = \vec{b}$ by using the LU factorization. That is, instead of solving $A\vec{x} = \vec{b}$ we solve $LU\vec{x} = \vec{b}$. If we let $U\vec{x} = \vec{y}$, then this becomes $L\vec{y} = \vec{b}$.

That is, if $A = LU$ and we want to solve $A\vec{x} = \vec{b}$, we first solve $L\vec{y} = \vec{b}$, and then solve $U\vec{x} = \vec{y}$. Notice that as L and U are triangular, each of these equations is easily solved for one variable at a time.

Use your LU factorization of the matrix in problem (10.4c) to solve the following system

$$\begin{pmatrix} 1 & -2 & -2 & -3 \\ 3 & -9 & 0 & -9 \\ -1 & 2 & 4 & 7 \\ -3 & -6 & 26 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \\ 0 \\ 3 \end{pmatrix}$$

Part II

Determinants, Eigenvectors, and Eigenvalues

8

Determinants

8.1 Introduction

We had seen earlier that a 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

was invertible if and only if $ad - bc \neq 0$. In this lecture we extend this idea to general $n \times n$ matrices. That is, we will see that there is a way to associate a number to an $n \times n$ matrix in such a way that the matrix will be invertible if and only if the associated number is non-zero. The formula for calculating this value, which is called the *determinant* of the matrix, is given recursively. That is, we define the determinant of an $n \times n$ matrix in terms of determinants of $(n-1) \times (n-1)$ matrices.

We will see that determinants actually tell us a lot more information than simply whether a linear transformation is invertible or not, but this is one easy-to-appreciate reason for studying determinants.

Instead of giving the formula for computing a determinant directly, we will first describe some properties we want this determinant to have and then show there is only one possible way to associate a number to a matrix such that these properties are satisfied.

8.2 Properties of Determinants

Given an $n \times n$ matrix A, we will associate to A a number called the ***determinant*** of A and denoted $\det(A)$. We will first give some geometric properties of determinants, and then give some algebraic properties.

Geometric properties

One way to interpret the determinant is as a way of describing the "size" of subsets of $\mathbb{R}^n$, and how a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ changes the size of a subsets.

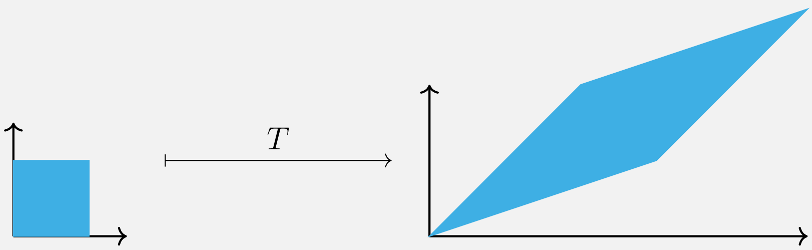
Example 8.1.
In $\mathbb{R}^1$, a linear transformation $T : \mathbb{R}^1 \to \mathbb{R}^1$ is the same thing as multiplication by some fixed number which we'll call A. For example,

$T(x) = 3x$ is a linear transformation of $\mathbb{R}^1$. Given any interval, say $I = [2, 4]$ in $\mathbb{R}^1$, the image of I under T is another interval. In this case, $T(I) = [6, 12]$. Notice that the length of $T(I)$ is 6 while the length of I is 2; so applying T stretched out the interval by a factor of 3.

Example 8.2.
Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix

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

This map takes the unit square and converts it into some parallelogram of area 4.



Notice that the determinant of this matrix (which we had previously defined for 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ as $ad - bc$) is 4.

In general, if we have a set $S \subseteq \mathbb{R}^2$ and we apply a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ to each point of S to get a new set $T(S)$. The absolute value of $\det(T)$ is the ratio of the areas of S and $T(S)$:

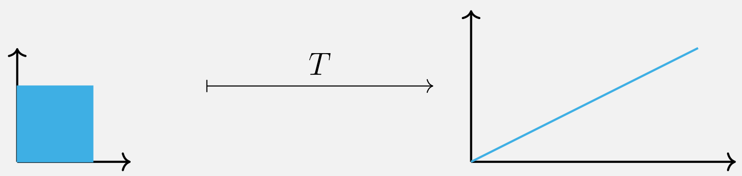
$$\text{area}(T(S)) = |\det(T)| \cdot \text{area}(S).$$

If it happened to be that our linear transformation was not invertible, then intuitively we should expect the linear transformation to "collapse" along certain directions. (If $T(v) = T(w)$, then $T(v - w) = \vec{0}$ and everything parallel to $v - w$ gets sent to the zero vector.) In terms of sizes of regions, this means we can take a set of positive size and send it to something with zero size, and so the determinant of a non-invertible transformation should be zero. In general, if $v \in \ker(T)$, then all multiples of v are in the kernel of T as well, so T collapses all vectors parallel to v.

Example 8.3.
Consider the linear transformation in $\mathbb{R}^2$ given by the matrix

$$\begin{pmatrix} 2 & 1 \\ 1 & 1/2 \end{pmatrix}$$

Notice that $\begin{pmatrix} 1 & -2 \end{pmatrix}^T$ is in the kernel of this map, and so everything parallel to this vector collapses to zero.



Since the area of the image on the right is zero, the determinant of the linear transformation should be zero.

The same thing holds in higher dimensions: if $S \subseteq \mathbb{R}^3$ and $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation, then

$$\text{volume}(T(S)) = |\det(T)| \cdot \text{volume}(S).$$

It is possible for the determinant of a matrix to be negative, so it's important that we use absolute values when we discuss areas or volumes since we don't want to talk about negative area or negative volume.

This idea extends to higher-dimensional spaces as well.

Remark.
Here the "size" of a set depends on what dimension we're talking about. In one dimension, "size" means the arclength of a subset of the real line. In two dimensions, "size" means the area of a subset of the plane. In three dimensions, "size" means volume. To define

size in higher dimensions it's helpful if you know some calculus. In particular, we can define the "size" of an set S in $\mathbb{R}^n$ as the integral

$$\underset{S}{\iint \cdots \iint} 1\, dx_1\, dx_2 \, \cdots dx_{n-1}\, dx_n$$

This notion of size is sometimes called *n-dimensional hypervolume*.

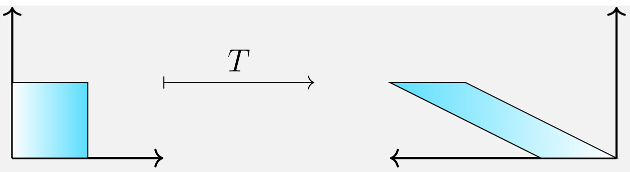
Besides telling us how the size of a set changes, we want the determinant to also tell us if a linear transformation "reverses" a set. This is simplest to describe in $\mathbb{R}^1$ and $\mathbb{R}^2$, but the idea extends to higher dimensions.

Example 8.4.
If $T : \mathbb{R}^1 \to \mathbb{R}^1$ is the linear transformation given by $T(x) = -2x$, then T not only stretches subsets by a factor of two, but it also reverses them. That is, given an interval $I = [a, b]$, its image $T(I) = [-2b, -2a]$ has the "opposite" left- and right-hand sides compared to I: the left-hand side of I became the right-hand side of $T(I)$, and the right-hand side of I became the left-hand side of $T(I)$. In a situation such as this we say that T is ***orientation reversing***. We want the determinant of T to tell us if the map is orientation reversing, $\det(T) < 0$, or ***orientation preserving***, $\det(T) > 0$.

Example 8.5.
Consider a map $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix

$$\begin{pmatrix} -1 & 2 \\ 0 & 1 \end{pmatrix}$$

This map shears sets to the right, which doesn't change the area of the set, but then flips the set over.



Since this map "reverses" a set we again expect its determinant to be negative, which it is: here the map has determinant -1.

Algebraic properties

In order to give some properties that this determinant will satisfy, it will be helpful to think of $\det$ as a function which takes n vectors, all of which are n-dimensional, and converts them into a single real number. That is, we think of $\det$ as a function

$$\det : \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ times}} \to \mathbb{R}$$

When we write $\det(A)$ what we will really mean is

$$\det(a_1, a_2, ..., a_n)$$

where a_1, ..., a_n are the columns of the $n \times n$ matrix A.

By "algebraic properties" of the determinant we mean the rules the determinant should obey when we modify the arguments of the function. There are only three algebraic properties we need to uniquely determine the determinant:

1. **Linearity**
   Our function $\det$ should be linear in each component. That is, for the i-th argument of $\det$ we should have

$$\det(—, v + w, —) = \det(—, v, —) + \det(—, w, —)$$

   and

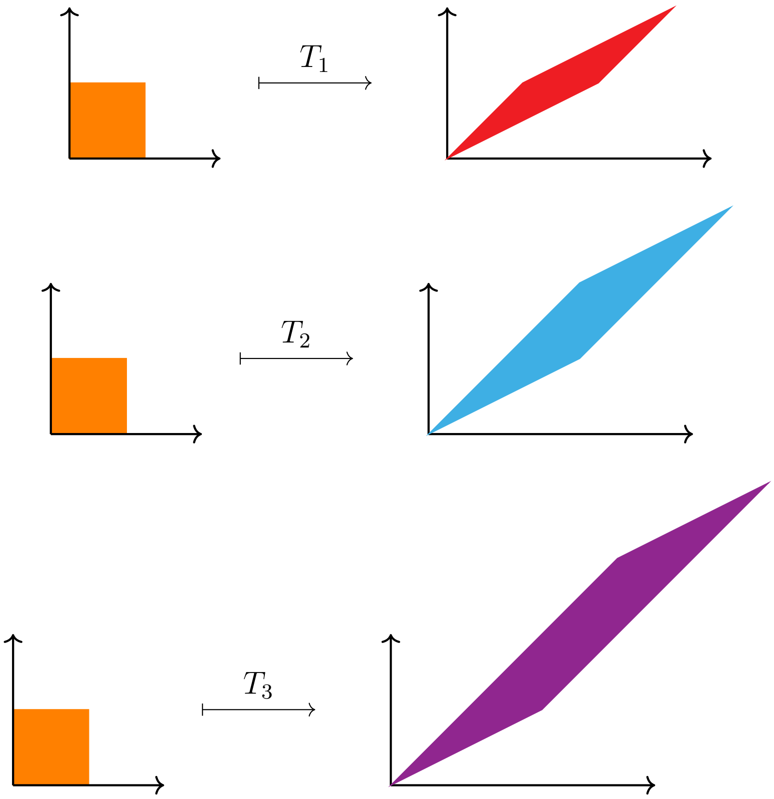
$$\det(—, \lambda v, —) = \lambda \det(—, v, —)$$

   where the dashes simply mean that the other entries of $\det$ don't matter. (If you want, pretend you've fixed all of the other entries and are only letting the i-th entry change.)

Geometrically we should expect this property because it's telling us that if we extend a region along some axis, the areas should add. For example, consider three linear transformations $T_1, T_2, T_3 : \mathbb{R}^2 \to \mathbb{R}^2$ with T_i given by the matrix A_i below.

$$A_1 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad A_2 = \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix} \quad A_3 = \begin{pmatrix} 2 & 3 \\ 1 & 3 \end{pmatrix}$$

Consider how these linear transformation map the unit square to parallelograms:



Notice that the area of the purple parallelogram is the sum of the areas of the red and blue parallelograms.

2. **Alternating**
The determinant is ***alternating*** in each argument. This means that if we swap two arguments, the determinant negates.

$$\det(—, v, —, w, —) = -\det(—, w, —, v, —)$$

This is how the determinant is "aware" of whether a linear transformation is orientation preserving or reversing.

3. **Identity**
The identity map id, whose matrix is the identity matrix I, doesn't do anything to sets, and so doesn't change the size of a set or reverse the set, and so we should expect that $\det(\text{id}) = 1$. Interpreting the determinant as a map from $\underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ times}}$ to $\mathbb{R}$, this means we want

$$\det(e_1, e_2, \cdots, e_n) = 1.$$

These three algebraic properties, that the determinant must be linear in each argument, is alternating, and assigns 1 to the identity, are enough to completely specify the determinant. That is, if you were to come up with another map

$$\varphi : \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ times}} \to \mathbb{R}$$

which was linear, alternating, and assigned 1 to the identity, then you would actually have come up with the same map we are about to define. For the sake of completeness we will prove this fact later in these notes, but you can safely ignore this proof if you want.

Exercise 8.1.
Assuming the determinant is linear in each column as described above, show that for every $n \times n$ matrix A and every scalar $\lambda \in \mathbb{R}$,

$$\det(\lambda A) = \lambda^n \det(A).$$

8.3 Computing the Determinant

So far we have described some properties we want our notion of determinant to have, but haven't said how to actually compute the determinant. We will start off by first saying what the determinant of a 1×1 matrix (i.e., a linear transformation $\mathbb{R}^1 \to \mathbb{R}^1$) $A = \begin{pmatrix} a \end{pmatrix}$ is simply

$$\det(A) = a.$$

To compute the determinant of an $n \times n$ matrix we will combine determinants of some $(n-1) \times (n-1)$ submatrices. We will use the following (non-standard) notation. Suppose our matrix A is $n \times n$ with a_{ij} denoting the entry in the i-th row and j-th column,

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ & & \vdots & & \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix}.$$

We will let A_{ij} denote the submatrix of A obtained by deleting the i-th row and j-th column of A (that is, we remove the row and column containing a_{ij}). For example,

$$A_{32} = \begin{pmatrix} a_{11} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{23} & \cdots & a_{2n} \\ a_{41} & a_{43} & \cdots & a_{4n} \\ & & \vdots & \\ a_{n1} & a_{n3} & \cdots & a_{nn} \end{pmatrix}$$

The determinant of A_{ij} is called the (i, j) ***minor*** of the matrix and is sometimes denoted

$$M_{ij} = \det(A_{ij}).$$

If we multiply the (i, j) minor by $(-1)^{i+j}$ we have the (i, j) ***cofactor*** of the matrix, sometimes denoted

$$C_{ij} = (-1)^{i+j} M_{ij} = (-1)^{i+j} \det(A_{ij}).$$

The determinant of A is then given by calculating the ***cofactor expansion*** of A along any row or column. The cofactor expansion of A along the i-th row is

$$a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in},$$

and the cofactor expansion of A along the j-th row is

$$a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

(The cofactor expansion is sometimes also referred to as the ***Laplace expansion***.)

Somewhat surprisingly, the cofactor expansion along any row or any column always gives you the same value, and this value is the ***determinant*** of the matrix.

It is important to realize that the cofactors, regardless of what row or column you expand along, alternate between positive and negative. To keep this straight we can rewrite the determinant as

$$\det(A) = \underbrace{\sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det(A_{ij})}_{\text{cofactor expansion using the } i\text{-th row}} = \underbrace{\sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det(A_{ij})}_{\text{cofactor expansion using the } j\text{-th column}}$$

We claim that the function $\det$ defined by cofactor expansion like this satisfies the three algebraic properties above. Before verifying this, let's use the cofactor expansion to evaluate some determinants.

Example 8.6.
Consider a 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

If we do cofactor expansion along the first row we have

$$\begin{aligned} \det(A) =& (-1)^{1+1}a \cdot \det(d) + (-1)^{1+2}b \cdot \det(c) \\ =& ad - bc \end{aligned}$$

Example 8.7.
Consider the 3×3 matrix

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

If we perform the cofactor expansion along the third column we have

$$\begin{aligned}
&\det(A)\\
=&(-1)^{1+3}\cdot 0\cdot \det\begin{pmatrix}2 & 3\\ 1 & 1\end{pmatrix}\\
&+(-1)^{2+3}\cdot 2\cdot \det\begin{pmatrix}1 & 2\\ 1 & 1\end{pmatrix}\\
&+(-1)^{3+3}\cdot 0\cdot \det\begin{pmatrix}1 & 2\\ 2 & 3\end{pmatrix}\\
=&0-2(1-2)+0\\
=&2
\end{aligned}$$

Notice that in the previous example we chose to expand along a column that had some zeros in it, and this in turn makes our calculation a little bit simpler: we don't need to bother calculating the determinants that get multiplied by zero!

Exercise 8.2.
Show that the determinant of a general 3×3 matrix

$$A=\begin{pmatrix}a & b & c\\ d & e & f\\ g & h & i\end{pmatrix}$$

is given by the following formula:

$$\det(A)=aei+bfg+cdh-ceg-bdi-afh.$$

Remark.
In general computing determinants using the cofactor expansion is extremely slow. To compute the determinant of an $n\times n$ matrix you need to perform on the order of $n! = n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1$ operations. This grows *extremely* quickly, and so becomes excruci-

atingly slow – even if you want were to perform the calculations on a computer – as we work with larger and larger matrices. After proving some theorems below, however, we will see that there is a quicker way to compute determinants.

Example 8.8.

$$
\begin{aligned}
&\det \begin{pmatrix} 1 & 2 & 4 & 1 \\ 2 & 1 & 1 & -1 \\ 3 & 1 & 2 & 4 \\ 1 & -1 & 2 & 2 \end{pmatrix} \\
=&1 \det \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & 4 \\ -1 & 2 & 2 \end{pmatrix} - 2 \det \begin{pmatrix} 2 & 1 & -1 \\ 3 & 2 & 4 \\ 1 & 2 & 2 \end{pmatrix} \\
&+ 4 \det \begin{pmatrix} 2 & 1 & -1 \\ 3 & 1 & 4 \\ 1 & -1 & 2 \end{pmatrix} - 1 \det \begin{pmatrix} 2 & 1 & 1 \\ 3 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \\
=&1 \cdot \left(1 \cdot \det \begin{pmatrix} 2 & 4 \\ 2 & 2 \end{pmatrix} - 1 \cdot \det \begin{pmatrix} 1 & 4 \\ -1 & 2 \end{pmatrix} + (-1) \cdot \det \begin{pmatrix} 1 & 2 \\ -1 & 2 \end{pmatrix} \right) \\
&- 2 \cdot \left(2 \cdot \det \begin{pmatrix} 2 & 4 \\ 2 & 2 \end{pmatrix} - 1 \cdot \det \begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix} + 1 \cdot \det \begin{pmatrix} 3 & 1 \\ 1 & -1 \end{pmatrix} \right) \\
&+ 4 \cdot \left(2 \cdot \det \begin{pmatrix} 1 & 4 \\ -1 & 2 \end{pmatrix} - 1 \cdot \det \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} + (-1) \cdot \det \begin{pmatrix} 3 & 1 \\ 1 & -1 \end{pmatrix} \right) \\
&- 1 \cdot \left(2 \cdot \det \begin{pmatrix} 1 & 2 \\ -1 & 2 \end{pmatrix} - 1 \cdot \det \begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix} + 1 \cdot \det \begin{pmatrix} 3 & 1 \\ 1 & -1 \end{pmatrix} \right) \\
=&1 \cdot (1 \cdot (2 \cdot 2 - 4 \cdot 2) - 1 \cdot (1 \cdot 2 - 4 \cdot (-1)) + (-1) \cdot (1 \cdot 2 - 2 \cdot (-1))) \\
&- 2 \cdot (2 \cdot (2 \cdot 2 - 4 \cdot 2) - 1 \cdot (1 \cdot 2 - 4 \cdot 1) + (-1) \cdot (3 \cdot 2 - 2 \cdot 1)) \\
&+ 4 \cdot (2 \cdot (1 \cdot 2 - 4 \cdot (-1)) - 1 \cdot (3 \cdot 2 - 4 \cdot 1) + (-1)(3 \cdot (-1) - 1 \cdot 1)) \\
&- 1 \cdot (2 \cdot (1 \cdot 2 - 2 \cdot (-1)) - 1 \cdot (3 \cdot 2 - 2 \cdot 1) + 1 \cdot (3 \cdot (-1) - 1 \cdot 1)) \\
=&1 \cdot (1 \cdot (-4) - 1 \cdot 6 - 1 \cdot 4) \\
&- 2 \cdot (2 \cdot (-4) - 1 \cdot (-2) - 1 \cdot 4) \\
&+ 4 \cdot (2 \cdot 6 - 1 \cdot 2 - 1 \cdot (-4)) \\
&- 1 \cdot (2 \cdot 4 - 1 \cdot 4 + 1 \cdot (-4)) \\
=&1 \cdot (-14) - 2 \cdot (-14) + 4 \cdot (14) - 1 \cdot 0 \\
=&-14 + 28 + 56 - 0 \\
=&70
\end{aligned}
$$

8.4 Existence and Uniqueness

This section is considerably more technical than the other sections and should be ignored on a first reading. The proof below is included only for the sake of completeness and to justify our claim that if we know the determinant is supposed to be linear in each column, alternating, and $\det(I) = 1$, then there is only one possible choice for what the determinant could be. The important thing for right now is to learn the basic properties of the determinant and how to calculate determinants, and so you can safely skip the theorem and proof below.

Theorem 8.1.
If φ and ψ are functions from $\underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ times}}$ to $\mathbb{R}$, which are linear in each argument, alternating, and assigning 1 to $(e_1, e_2, ..., e_n)$, then $\varphi = \psi$.

Proof.
Notice that if φ is linear and $\varphi(e_1, ..., e_n) = 1$, then for any $v = \lambda_1 e_1 + \cdots + \lambda_n e_n$, we have

$$\varphi(e_1, ..., e_{i-1}, v, e_{i+1}, ..., e_n) = \lambda_i.$$

That is, if we place v in the i-th argument of φ and otherwise put e_j in the j-th position, then we just pull of the scalar λ_i appearing the linear combination of v with respect to the standard basis.

If we were to place e_i in one of the positions above instead of e_j, then since φ is alternating we have

$$\varphi(e_1, ..., e_{i-1}, v, e_{i+1}, ..., e_i, ..., e_n) = -\lambda_j$$

by swapping v and e_i (which is in the j-th position), which forces us to change signs, but now v is in the j-th position and so by the above we pick up the scalar λ_j.

Now we claim that φ is uniquely determined, in fact we can use the properties above to get a (somewhat complicated) formula for φ in terms of all possible permutations[a] of the entries to v. We claim in particular that

$$\varphi(v_1, v_2, ..., v_n) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} \lambda_{\sigma(i),i}.$$

Here $\mathfrak{S}_n$ denotes the set of all possible permutations of $\{1, 2, ..., n\}$, and the ***sign*** of a permutation σ, denoted $\operatorname{sgn}(\sigma)$ is either 1 or -1 depending on whether the permutation σ can be written as a product of an even or odd number of transpositions[b]. We are supposing above that

$$v_j = \lambda_{1,j} e_1 + \lambda_{2,j} e_2 + \cdots + \lambda_{n,j} e_n = \sum_{i=1}^{n} \lambda_{i,j} e_i$$

Since φ is linear in each argument we have

$$\begin{aligned}
&\varphi(v_1, v_2, ..., v_n) \\
=&\varphi\left(\sum_{i=1}^{n} \lambda_{i,1} e_i, ..., \sum_{i=1}^{n} \lambda_{i,n} e_i\right) \\
=&\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} \prod_{j=1}^{n} \lambda_{i_k,j} \varphi(e_{i_1}, ..., e_{i_n}) \\
=&\sum_{\sigma \in \mathfrak{S}_n} \prod_{i=1}^{n} \lambda_{\sigma(i),i} \varphi(e_{\sigma(1)}, e_{\sigma(2)}, ..., e_{\sigma(n)}) \\
=&\sum_{\sigma \in \mathfrak{S}_n} \prod_{i=1}^{n} \lambda_{\sigma(i),i} \operatorname{sgn}(\sigma) \varphi(e_1, ..., e_n) \\
=&\sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} \lambda_{\sigma(i),i}
\end{aligned}$$

Notice that this formula for φ uses only the fact that φ is linear in each argument, alternating, and $\varphi(e_1, ..., e_n) = 1$. Hence any other function ψ which is linear in each argument, alternating, and has $\psi(e_1, ..., e_n) = 1$ can also be put into the same formula. Thus $\varphi = \psi$.

The formula above is called the ***Leibniz formula*** and is another way of computing determinants. Since we are summing over all permutations of n to obtain the Leibniz formula, this is provides us with another easy way of seeing that computing determinants in general takes about $n!$ arithmetic operations. □

[a] A ***permutation*** of a set is just a bijection from the set to itself. A set with n elements will have $n!$ possible permutations.

[b] A ***transposition*** is a permutation which simply swaps two elements. It is a basic fact about permutations that every permutation can be written as a product of transpositions.

We have now established that if the we want the determinant to be linear in each column, alternating, and assigns 1 to the identity matrix I, then there is only one possible function the determinant could be. In fact, by Leibniz's formula above we even know what the function is. This means we have established the existence and the uniqueness of the determinant.

8.5 More Properties of Determinants

We had stated above that determinants should satisfy certain geometric and algebraic properties. By the definition of the determinant the algebraic properties are already satisfied, so there's nothing to prove, but we'll recall the algebraic properties in terms of matrices.

Theorem 8.2.
If A is an $n \times n$ matrix, then

1. *If A, A', and A'' are three $n \times n$ matrices which are the same except for the i-th column, and if the i-th column of A is the sum of the i-th columns of A' and A'', then* $\det(A) = \det(A') + \det(A'')$.

2. *If one column of A is multiplied by a scalar λ to produce a new matrix A', then* $\det(A') = \lambda \det(A)$.

3. *If two columns of A are exchanged producing a new matrix A', then* $\det(A') = -\det(A)$.

4. $\det(I) = 1$.

The following theorem will let us convert the algebraic properties for columns of the matrix into algebraic properties for rows.

Theorem 8.3.
If A is an $n \times n$ matrix, then $\det(A^T) = \det(A)$.

Proof.
To simplify notation, let $B = A^T$. Notice $b_{ij} = a_{ji}$ and likewise $B_{ij} = A_{ji}$. Now compute the determinant of B using cofactor expansion along the first column:

$$\begin{aligned}\det(A^T) &= \det(B)\\ &= \sum_{i=1}^{n} (-1)^{1+i} b_{i1} \det(B_{i1})\\ &= \sum_{i=1}^{n} (-1)^{1+i} a_{1i} \det(A_{1i})\end{aligned}$$

Notice that this is precisely the formula for the determinant of A using the cofactor expansion along the first row: thus $\det(A) = \det(A^T)$. □

Combining Theorem 8.2 and Theorem 8.3 we obtain the following.

Corollary 8.4.
If A is an $n \times n$ matrix, then

1. *If A, A', and A'' are three $n \times n$ matrices which are the same except for the i-th row, and if the i-th row of A is the sum of the i-th rows of A' and A'', then* $\det(A) = \det(A') + \det(A'')$.
2. *If one row of A is multiplied by a scalar λ to produce a new matrix A', then* $\det(A') = \lambda \det(A)$.

3. *If two rows of A are exchanged producing a new matrix A', then* $\det(A') = -\det(A)$.

It will also be helpful if we notice that if A is a triangular matrix, then it is very easy to compute the determinant of A.

Theorem 8.5.
If A is a triangular $n \times n$ matrix, then the determinant of A is equal to the product of the diagonal entries of A. That is,

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1,n-1} & a_{1n} \\ 0 & a_{22} & a_{23} & a_{24} & \cdots & a_{2,n-1} & a_{2n} \\ 0 & 0 & a_{33} & a_{34} & \cdots & a_{3,n-1} & a_{3n} \\ & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$
$$= a_{11} \cdot a_{22} \cdot \ldots \cdot a_{nn}$$

Proof.
We will do a proof by induction. First note that the theorem is obviously true if $n = 1$. For $n > 1$, suppose the theorem has been proven for $(n-1) \times (n-1)$ matrices. If we perform cofactor expansion along the first column, then the only non-zero entry in the cofactor expansion is

$$\det(A) = a_{11} \det(A_{11}).$$

But A_{11} is a $(n-1) \times (n-1)$ triangular matrix obtained by deleting the first row and first column of A. Thus

$$\det(A_{11}) = a_{22} \cdot a_{33} \cdot \ldots \cdot a_{nn}.$$

Hence

$$\det(A) = a_{11} \cdot a_{22} \cdot \ldots \cdot a_{nn}.$$

We have proven the theorem for upper triangular matrices, but since the transpose of a lower triangular matrix is an upper triangular matrix, Theorem 8.3 provides the result for lower triangular matrices. □

Example 8.9.

$$\det\begin{pmatrix} 2 & 3 & 4 & 1 & 2 \\ 0 & 1 & 9 & 0 & -1 \\ 0 & 0 & 3 & 4 & 7 \\ 0 & 0 & 0 & -3 & 3 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} = 2 \cdot 1 \cdot 3 \cdot (-3) \cdot 4 = -72$$

Combining Theorem 8.5 and the properties of determinants already discussed we can easily compute the determinant of any elementary matrix.

Theorem 8.6.
If E is an $n \times n$ elementary matrix, then

1. $\det(E) = \lambda$ *if E multiplies a row by λ.*
2. $\det(E) = -1$ *if E swaps two rows.*
3. $\det(E) = 1$ *if E adds a multiple of one row to another.*

Notice the first two properties tell us that if E is an elementary matrix obtained by swapping two rows or multiplying one row by λ, then $\det(EA) = \det(A)$. We can extend this result to all elementary matrices, but first we need one little observation.

Lemma 8.7.
If A is an $n \times n$ matrix and two rows of A are identical, then $\det(A) = 0$.

Proof.
Suppose rows i and j of A are identical. Then if we swap rows i and j, we still have the same matrix. However, swapping two rows results in our negating the determinant, which means

$$\det(A) = -\det(A),$$

and so the only option is $\det(A) = 0$. □

Theorem 8.8.
If A is any $n \times n$ matrix and B is obtained from A by adding λ times the i-th row of A to the j-th row of A,

$$A \xrightarrow{\lambda R_i + R_j \to R_j} B$$

then $\det(A) = \det(B)$.

Proof.
This follows from the fact that the determinant is linear in the rows of the matrix. If we let B' be the matrix which is the same as A except row j is replaced by row i, then linearity in the rows tells us

$$\det(B) = \det(A) + \det(B') = \det(A) + 0$$

□

We now have the following.

Corollary 8.9.

If A is any $n \times n$ matrix and E is an $n \times n$ elementary matrix, then

$$\det(EA) = \det(E)\,\det(A)$$

Exercise 8.3.
Prove Corollary 8.9 by combining Corollary 8.4, Theorem 8.6, and Theorem 8.8.

Now we are able to prove the following useful property of determinants described earlier.

Theorem 8.10.
An $n \times n$ matrix A is invertible if and only if $\det(A) \neq 0$.

Proof.
Suppose that A is invertible. Then $A^{-1}A = I$ and so

$$\det(A^{-1})\det(A) = 1,$$

which implies $\det(A) \neq 0$.

Now suppose $\det(A) \neq 0$. Let E_1, E_2, ..., E_m be the elementary matrices such that

$$E_m E_{m-1} \cdots E_2 E_1 A$$

is in row reduced echelon form. Notice that the determinant of an elementary matrix is never zero, and so

$$\begin{aligned}&\det(E_m E_{m-1} \cdots E_2 E_1 A)\\ =&\det(E_m)\det(E_{m-1})\cdots\det(E_2)\det(E_1)\det(A)\\ \neq&0.\end{aligned}$$

But $E_m E_{m-1} \cdots E_2 E_1 A$, being in RREF, is upper triangular. Since the determinant of an upper triangular matrix is the product of the diagonal entries, this must mean that $E_m E_{m-1} \cdots E_2 E_1 A$ has no zeros on the diagonal. But, being in RREF, that means $E_m E_{m-1} \cdots E_2 E_1 A$ is the identity. This means A is invertible with inverse $A^{-1} = E_m \cdots E_1$. □

We saw above that if E is an elementary matrix, $\det(EA) = \det(E)\det(A)$. We can actually extend this property to all products of matrices if we are able to show one useful property of matrix inverses.

Exercise 8.4.
Suppose that A and B are two $n \times n$ matrices and A *is not* invertible. Show that AB can not be invertible either.

Theorem 8.11.
If A and B are two $n \times n$ matrices, then $\det(AB) = \det(A)\ \det(B)$.

Proof.
Suppose that one of A or B is not invertible – say A is not invertible. Then AB is not invertible either, by Exercise 8.4, and so $\det(AB) = 0$, but $\det(A)\det(B) = 0$ as well since $\det(A) = 0$.

Now suppose that both A and B are invertible. Then as A is the inverse of A^{-1}, we can write A as a product of elementary matrices: say

$$A = E_m E_{m-1} \cdots E_2 E_1.$$

and

$$\det(A) = \det(E_m)\det(E_{m-1}) \cdots \det(E_2)\det(E_1)$$

thus

$$\begin{aligned}\det(AB) &= \det(E_m E_{m-1} \cdots E_2 E_1 B) \\ &= \det(E_m)\det(E_{m-1}) \cdots \det(E_2)\det(E_1)\det(B) \\ &= \det(A)\det(B).\end{aligned}$$

□

8.6 Faster Algorithms

We have seen that the determinant is uniquely characterized by three algebraic properties, and we saw that the Laplace expansion gave an algorithm for computing the determinant that satisfies these three properties. The problem with the Laplace expansion is that it is extraordinarily slow to compute: not just by hand, but even on a computer.

Luckily there are faster ways to compute the determinant. In particular, we can very easily compute the determinant of triangular matrices and we can also compute the determinant of a product of matrices as the product of the determinants. Thus if we have a factorization of a matrix into triangular matrices, such as the LU factorization, then it becomes *much* easier to compute the determinant.

Example 8.10.
Computing the determinant of a 7×7 matrix using the Laplace expansion would take about $7! = 5,040$ operations. If we have an LU factorization, however, then our work becomes much simpler.

For example, suppose

$$A = \begin{pmatrix} 3 & 2 & 7 & 9 & 1 & 1 & 1 \\ 9 & 8 & 27 & 31 & 7 & 13 & 5 \\ 3 & 2 & 5 & 8 & 9 & -1 & -2 \\ 6 & 3 & 17 & 21 & -20 & 11 & 18 \\ 3 & 8 & 39 & 30 & -38 & 56 & 37 \\ 12 & 8 & 34 & 38 & -20 & 16 & 15 \\ 21 & 21 & 88 & 89 & -42 & 91 & 67 \end{pmatrix}$$

This matrix has the LU factorization

$$U = \begin{pmatrix} 3 & 2 & 7 & 9 & 1 & 1 & 1 \\ 0 & 1 & 3 & 2 & 2 & 5 & 1 \\ 0 & 0 & 2 & 1 & -8 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 & 4 & 4 \\ 0 & 0 & 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & -1 & 3 & 2 & 0 & 0 & 0 \\ 1 & 6 & 7 & 2 & 1 & 0 & 0 \\ 4 & 0 & 3 & -1 & 2 & 2 & 0 \\ 7 & 7 & 9 & 3 & 3 & 5 & 1 \end{pmatrix}$$

So instead of doing the $5,040$ operations to compute the determinant using the cofactor expansion, we can instead just do 14: multiplying together the diagonal entries of L and U to obtain

$$\det(A) = \det(LU) = \det(L) \cdot \det(U) = -8 \cdot 12 = -96$$

Computing a determinant this way is of course only helpful if we can compute the LU factorization of a matrix in a reasonable amount of time. There are algorithms to compute LU factorizations (or LUP factorizations if we require row swaps) in $O(n^3)$ time (this means that for an $n \times n$ matrix we have to do on the order of n^3 computations), and so this is a reasonable, effective way to actually compute the determinant.

8.7 Cramer's Rule

Cramer's rule is a method for describing the solution to a system $Ax = b$, provided A is invertible, in terms of determinants. In order to state Cramer's rule we need one bit of notation. Given an $n \times n$ matrix A and an n-dimensional vector b, let $A_i(b)$ denote the matrix obtained by replacing the i-th column of A with b.

Example 8.11.
Let A and b be the following:

$$A = \begin{pmatrix} 6 & 2 & 1 & 1 \\ 3 & 7 & 8 & 2 \\ 4 & 1 & 3 & 3 \\ 2 & 2 & 0 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 4 \\ 2 \\ 3 \\ 1 \end{pmatrix}$$

Then

$$A_3(b) = \begin{pmatrix} 6 & 2 & 4 & 1 \\ 3 & 7 & 2 & 2 \\ 4 & 1 & 3 & 3 \\ 2 & 2 & 1 & 1 \end{pmatrix}$$

Theorem 8.12 (Cramer's Rule)**.**
If A is an invertible $n \times n$ matrix, and b is an n-dimensional vector, then the unique solution to $Ax = b$ has components

$$x_i = \frac{\det(A_i(b))}{\det(A)}$$

Proof.
Notice that if we consider the matrix obtained by replacing the i-th column of the identity matrix I with x, $I_i(x)$, then the determinant of this matrix is simply x_i. To see this, perform cofactor expansion along the i-th row of $I_i(x)$: all entries in the i-th row are zero for the entry coming from the column which was replaced by x which has x_i in this location: the i-th row and i-th column. The corresponding (i, i)-minor is simply the $(n-1) \times (n-1)$ identity matrix which has determinant 1, and so we have that the cofactor (and hence the entire determinant) is

$$(-1)^{i+i} x_i \det(I) = x_i.$$

Note too that

$$A \cdot I_i(x) = \begin{pmatrix} Ae_1 & Ae_2 & \cdots & Ax & \cdots & Ae_n \end{pmatrix}$$

So if x solves the equation $Ax = b$ (since A is invertible there is exactly one vector solving the equation) this becomes

$$A \cdot I_i(x) = \begin{pmatrix} Ae_1 & Ae_2 & \cdots & b & \cdots & Ae_n \end{pmatrix}$$

But notice that Ae_j is the j-th column of A, which we'll denote a_j, and so

$$A \cdot I_i(x) = \begin{pmatrix} a_1 & a_2 & \cdots & b & \cdots & a_n \end{pmatrix} = A_i(b)$$

If we take the determinant of both sides of the equation we have

$$\begin{aligned} & \det(A \cdot I_i(x)) = \det(A_i(b) \\ \implies & \det(A) \cdot \det(I_i(x)) = \det(A_i(b)) \\ \implies & \det(A) \cdot x_i = \det(A_i(b)) \\ \implies & x_i = \frac{\det(A_i(b))}{\det(A)} \end{aligned}$$

□

Cramer's rule is generally not a very efficient way to solve large systems of equations, but can sometimes be helpful for theoretical situations (e.g., in proving some theorem it might be helpful to have a way to express the components of a solution, and Cramer's rule allows us to do precisely that).

Example 8.12.

Use Cramer's rule to solve the following system

$$\begin{pmatrix}1 & 3 & 2\\0 & 1 & 4\\2 & 0 & 1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}0\\0\\2\end{pmatrix}$$

First note that the determinant of our matrix is

$$\det\begin{pmatrix}1 & 3 & 2\\0 & 1 & 4\\2 & 0 & 1\end{pmatrix} = 21$$

By Cramer's rule,

$$\begin{aligned}
x =& \frac{1}{21}\det\begin{pmatrix}0 & 3 & 2\\0 & 1 & 4\\2 & 0 & 1\end{pmatrix} = \frac{20}{21}\\
y =& \frac{1}{21}\det\begin{pmatrix}1 & 0 & 2\\0 & 0 & 4\\2 & 2 & 1\end{pmatrix} = \frac{-8}{21}\\
z =& \frac{1}{21}\det\begin{pmatrix}1 & 3 & 0\\0 & 1 & 0\\2 & 0 & 2\end{pmatrix} = \frac{2}{21}
\end{aligned}$$

Thus our system is solved by

$$\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}{}^{20}\!/_{21}\\-{}^{8}\!/_{21}\\{}^{2}\!/_{21}\end{pmatrix}$$

We can use Cramer's rule to get a formula for the inverse of a matrix, assuming of course the matrix is in fact invertible. Note that for an invertible $n \times n$ matrix A, the determinant of $A_i(e_j)$ – the matrix obtained by replacing the i-th row of A with the standard basis vector e_j – easily seen via cofactor expansion as

$$\det(A_i(e_j)) = (-1)^{i+j}\det(A_{ji}).$$

Recall that this is the (j, i) cofactor of A,

$$\det(A_i(e_j)) = C_{ji}.$$

Now, as $AA^{-1} = I$, the j-th column of A^{-1} – let's call this $\vec{\alpha}_j$ – satisfies the equation $A\vec{\alpha}_j = e_j$. So by Cramer's rule the i-th component of this vector (and hence the entry in the i-th row, j-th column of A^{-1}) is

$$\frac{\det(A_i(e_j))}{\det(A)} = \frac{C_{ji}}{\det(A)}.$$

Factoring out the denominator of $\det(A)$ from each entry we then have

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ & & \vdots & \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{pmatrix}$$

Notice that the matrix on the right is the transpose of the matrix obtained by putting the (i, j)-cofactor of A into the i-th row and j-th column. This matrix is sometimes called the ***adjugate***[1] of A and denoted $\operatorname{adj}(A)$. Hence we have

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A).$$

Example 8.13.
For a 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

the adjugate is

$$\operatorname{adj}(A) = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

and thus the inverse of A is

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}.$$

[1]In older texts this matrix is called the ***adjoint*** of A, but in modern mathematics "adjoint" usually means something different and so the term "adjugate" is preferred.

8.8 Practice Problems

Problem 8.1.
Compute the determinant of each of the matrices below.

(a) $\begin{pmatrix} 3 & 2 \\ -1 & 4 \end{pmatrix}$

(b) $\begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 3 \\ 1 & 0 & 3 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & -2 & 7 \\ 0 & 0 & 0 & 3 \end{pmatrix}$

Problem 8.2.
Each matrix below is given together with its LU, or LUP, factorization. Use this to compute the determinant.

(a)

$$A = \begin{pmatrix} 3 & 8 & 18 & 23 \\ 2 & 4 & 6 & 8 \\ 2 & 4 & 8 & 9 \\ 1 & 2 & 3 & 5 \end{pmatrix} \qquad L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2/3 & 1 & 0 & 0 \\ 2/3 & 1 & 1 & 0 \\ 1/3 & 1/2 & 0 & 1 \end{pmatrix}$$

$$U = \begin{pmatrix} 3 & 8 & 18 & 23 \\ 0 & -4/3 & -6 & -22/3 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

(b)

$$A = \begin{pmatrix} 3 & 4 & 1 & 7 \\ 2 & 2 & 3 & 8 \\ 0 & 1 & 0 & 1 \\ 1 & 2 & 2 & 5 \end{pmatrix} \qquad L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2/3 & -2/3 & 1 & 0 \\ 1/3 & 2/3 & 5/7 & 1 \end{pmatrix}$$

$$U = \begin{pmatrix} 3 & 4 & 1 & 7 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 7/3 & 4 \\ 0 & 0 & 0 & -6/7 \end{pmatrix} \qquad P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Problem 8.3.
Use Cramer's rule to solve the following system of equations:

$$\begin{aligned} 2x + 3y - 4z =& 1 \\ 6x - y + 3z =& 2 \\ 2x + y + z =& 3 \end{aligned}$$

Problem 8.4.
For each of the matrices given below, first compute the determinant to determine whether the matrix is invertible or not. If the matrix is invertible, then use the formula $A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$ to compute the inverse of A.

(a) $\begin{pmatrix} 2 & 4 & 8 \\ 1 & 5 & 7 \\ -1 & 3 & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 7 & 2 & 3 \\ 1 & 4 & 1 \\ 8 & 5 & 2 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 3 & 6 & 4 \\ 2 & 3 & 5 & 1 \\ 3 & 3 & 11 & 5 \\ 0 & 1 & 0 & 2 \end{pmatrix}$

9

Subspaces, Dimension, and Rank

9.1 Subspaces

Suppose that V is a subset of vectors $\mathbb{R}^n$, $V \subseteq \mathbb{R}^n$. We will say that V is a a ***subspace*** of $\mathbb{R}^n$ if the following two properties are satisfied:

1. V is not empty: $V \neq \emptyset$.
2. Given any pair of vectors $\vec{v}_1, \vec{v}_2 \in V$ inside of V, their sum $\vec{v}_1 + \vec{v}_2$ is also in V.
3. Given any vector $\vec{v} \in V$ in V, every scalar multiple $\lambda\vec{v}$ is also in V.

We sometimes say that the set V is ***closed*** under vector addition and scalar multiplication in a case such as this. This means simply what we've stated above: elements of V can be added together and we obtain a new element of V; and scalar multiples of elements of V are also elements of V.

Lemma 9.1.
If V is a subspace of $\mathbb{R}^n$, then $\vec{0} \in V$.

Proof.
Since V is a subspace, by definition V is non-empty. Say $\vec{v} \in V$. Then $\vec{0} = 0 \cdot \vec{v}$ is in V since V is closed under scalar multiplication. □

Example 9.1.
The xy-plane in $\mathbb{R}^3$,

$$V = \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \middle| \, x, y \in \mathbb{R} \right\}$$

is a subspace of $\mathbb{R}^3$.

To see this, notice that the sum of two elements of the xy-plane is also in the xy-plane,

$$\begin{pmatrix} x_1 \\ y_1 \\ 0 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ 0 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ 0 \end{pmatrix},$$

and any scalar multiple of someting in the xy-plane is in the xy-plane,

$$\lambda \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \\ 0 \end{pmatrix}.$$

Example 9.2.

More generally, given any collection of vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, ..., \vec{v}_m\} \in \mathbb{R}^n$, the span of those vectors is a subspace.

Let V be the span of the vectors,

$$\begin{aligned} V &= \text{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_m\} \\ &= \left\{\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \cdots + \lambda_m\vec{v}_m \,\middle|\, \lambda_1, \lambda_2, ..., \lambda_m \in \mathbb{R}\right\}. \end{aligned}$$

Notice again that V is closed under vector addition and scalar multiplication:

$$\begin{aligned} &(\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \cdots + \lambda_m\vec{v}_m) + (\mu_1\vec{v}_1 + \mu_2\vec{v}_2 + \cdots + \mu_m\vec{v}_m) \\ &= (\lambda_1 + \mu_1)\,\vec{v}_1 + (\lambda_2 + \mu_2)\,\vec{v}_2 + \cdots + (\lambda_m + \mu_m)\,\vec{v}_m \end{aligned}$$

$$\mu\,(\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \cdots + \lambda_m\vec{v}_m) = \mu\lambda_1\vec{v}_1 + \mu\lambda_2\vec{v}_2 + \cdots + \mu\lambda_m\vec{v}_m.$$

When V is the span of vectors $\vec{v}_1$, ..., $\vec{v}_m$, we say that V is ***spanned by*** those vectors, or is ***generated*** by the vectors.

Exercise 9.1.

(a) Show that the set containing only the zero vector, $\{\vec{0}\}$, is a subspace of $\mathbb{R}^n$.

(b) Show that every $\mathbb{R}^n$ is a subspace of itself.

If V is spanned by $\vec{v}_1$, ..., $\vec{v}_m$ and if these vectors are linearly independent, we say that $\vec{v}_1$, ..., $\vec{v}_m$ form a ***basis*** for the vector space.

Example 9.3.
A basis for the xy-plane in $\mathbb{R}^3$ is given by

$$\begin{pmatrix}1\\0\\0\end{pmatrix} \qquad \begin{pmatrix}0\\1\\0\end{pmatrix}$$

Example 9.4.
The n-dimensional $\vec{e}_1$, $\vec{e}_2$, ..., $\vec{e}_n$ vectors form a basis for $\mathbb{R}^n$. These are called the ***standard basis vectors*** for $\mathbb{R}^n$.

Theorem 9.2.
Every subspace of $\mathbb{R}^n$ has a basis. That is, given any *subspace $V \subseteq \mathbb{R}^n$, there is some collection of vectors $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_m$ such that these vectors are linearly independent and $V = \operatorname{span}\{\vec{v}_1, ..., \vec{v}_m\}$.*

We will skip the proof of the above theorem since this theorem will follow from a more general theorem about *abstract vector spaces* that we will discuss later.

Notice that the basis of a subspace *is not* unique! With the exception of the zero subspace, every subspace has infinitely-many bases.

Example 9.5.
Another basis for the xy-plane in $\mathbb{R}^3$ is

$$\begin{pmatrix}2\\3\\0\end{pmatrix} \qquad \begin{pmatrix}4\\1\\0\end{pmatrix}$$

Example 9.6.
Another basis for $\mathbb{R}^n$ is

$$\{\vec{e}_1, 2\vec{e}_2, 3\vec{e}_3, ..., m\vec{e}_m\}.$$

When we choose to write a vector as a list of numbers, what we're really doing is saying how that vector is written as a linear combination of the vectors in some basis. For $\mathbb{R}^n$ we usually implicitly assume this basis is the standard basis vector. That is, when we write

$$\vec{v} = \begin{pmatrix}2\\4\\-1\end{pmatrix}$$

what we really mean is

$$\vec{v} = 2\vec{e}_1 + 4\vec{e}_2 - \vec{e}_3.$$

But if we picked another basis, we would write our vector as some other linear combination. For example, say in $\mathbb{R}^3$ we used the basis

$$\mathcal{B} = \left\{ \vec{b}_1 = \begin{pmatrix}1\\1\\0\end{pmatrix}, \vec{b}_2 = \begin{pmatrix}2\\0\\2\end{pmatrix}, \vec{b}_3 = \begin{pmatrix}1\\1\\1\end{pmatrix} \right\}.$$

Then our vector $\vec{v}$ from before can be written as

$$\vec{v} = 3\vec{b}_1 - \vec{b}_2 + \vec{b}_3$$

It would then be reasonable for us to write

$$\vec{v} = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}_{\mathcal{B}}.$$

We will adopt the convention that a subscript $\mathcal{B}$ means we are writing the vector with respect to some basis $\mathcal{B}$. If we don't say what basis we're using, then we will always assume it is the standard basis, $\{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$.

Remark.
Writing a vector with respect to different bases is comparable to writing a real number with respect to a different base. For example, when you see 271.5 this really means the number

$$2 \cdot 10^2 + 7 \cdot 10^1 + 1 \cdot 10^0 + 5 \cdot 10^{-1}$$

but we could also write this number in binary as 100001111.1 which means

$$1 \cdot 2^8 + 0 \cdot 2^7 + 0 \cdot 2^6 + 0 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 + 1 \cdot 2^{-1}.$$

We're writing the same number, just in two different ways: once with base 10, and once with base 2. Writing a vector with respect to two different bases is the same idea: we're expressing the same vector in different ways.

Exercise 9.2.
Write the vector

$$\vec{v} = \begin{pmatrix} 2 \\ 4 \\ -1 \end{pmatrix}$$

above, but with respect to each of the following bases:

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \right\}$$

$$\mathcal{B}_2 = \left\{ \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}$$

9.2 Dimension

We are now ready to properly define the notion of "dimension." Intuitively, you should think that the dimension of a space tells you how many pieces of information you need to specify a point in the space.

Theorem 9.3.
Suppose $\{\vec{v}_1, ..., \vec{v}_m\}$ *and* $\{\vec{w}_1, ..., \vec{w}_p\}$ *are two bases for a subspace* $V \subseteq \mathbb{R}^n$. *(That is,* $V = \text{span}\{\vec{v}_1, ..., \vec{v}_m\} = \text{span}\{\vec{w}_1, ..., \vec{w}_p\}$ *and each set of vectors is linearly independent.) Then* $p = m$. *I.e., any two bases for a vector space have to have the same number of elements.*

Proof.
If $p \neq m$, then one of p or m is larger. Suppose first that $p > m$.

Notice that each $\vec{w}_i$ may be written as a linear combination of the $\vec{v}_j$ since the $\vec{v}_j$'s form a basis:

$$\vec{w}_i = \sum_{j=1}^{m} \alpha_{ij} \vec{v}_j.$$

Now consider the equation

$$\lambda_1 \vec{w}_1 + \lambda_2 \vec{w}_2 + \cdots + \lambda_p \vec{w}_p = \vec{0}.$$

Rewriting each $\vec{w}_i$ in terms of the $\vec{v}_j$ we have

$$\lambda_1 (\alpha_{11}\vec{v}_1 + \cdots + \alpha_{1m}\vec{v}_m) + \cdots + \lambda_p (\alpha_{p1}\vec{v}_1 + \cdots + \alpha_{pm}\vec{v}_m) = \vec{0}$$

Grouping the $\vec{v}_j$'s together this becomes

$$\begin{aligned} &(\lambda_1\alpha_{11} + \lambda_2\alpha_{21} + \cdots + \lambda_p\alpha p1)\,\vec{v}_1 \\ &+ (\lambda_1\alpha_{12} + \lambda_2\alpha_{22} + \cdots + \lambda_p\alpha p2)\,\vec{v}_2 \\ &+ \cdots \\ &+ (\lambda_1\alpha_{1m} + \lambda_2\alpha_{2m} + \cdots + \lambda_p\alpha_{pm})\,\vec{v}_m \\ &= \vec{0} \end{aligned}$$

Since the $\vec{v}_j$ are linearly independent, we must have that each of these scalars is zero. This gives us a system of linear equations,

$$\begin{aligned} \alpha_{11}\lambda_1 + \alpha_{21}\lambda_2 + \cdots + \alpha_{p1}\lambda_p &= 0 \\ \alpha_{12}\lambda_1 + \alpha_{22}\lambda_2 + \cdots + \alpha_{p2}\lambda_p &= 0 \\ &\vdots \\ \alpha_{1m}\lambda_2 + \alpha_{2m}\lambda_2 + \cdots + \alpha_{pm}\lambda_p &= 0. \end{aligned}$$

Notice the coefficient matrix of this system has more columns than rows, hence there must be a non-trivial solution. That is, some of the λ_1, λ_2, ..., λ_p are non-zero. But this implies we have a non-trivial solution to the equation

$$\lambda_1\vec{w}_1 + \cdots + \lambda_p\vec{w}_p = \vec{0}$$

which means the $\vec{w}_i$ are not linearly independent. This contradicts the assumption that the $\vec{w}_i$ form a basis, however. Hence $p \not> m$ – that is, $p \leq m$.

Suppose now that $p < m$. Repeating the argument above, but with the roles of $\vec{v}_j$ and $\vec{w}_i$ exchanged, shows that $p \not< m$, so $p \geq m$.

As $p \geq m$ and $p \leq m$, the only option is that $p = m$. □

So given any subspace, any two bases have to have the same number of elements. This common number of elements to all bases is called the ***dimension*** of the subpsace. The dimension of V is often denoted $\dim(V)$ or $\dim_{\mathbb{R}}(V)$.

Example 9.7.
$\mathbb{R}^n$ is n-dimensional since $\{\vec{e}_1, ..., \vec{e}_n\}$ forms a basis.

Example 9.8.
The xy-plane in $\mathbb{R}^3$ is two-dimensional since

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$$

forms a basis.

9.3 Change of Basis

As we have seen above, when a vector in $\mathbb{R}^n$ is written as

$$v = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}$$

this really means that the scalars v_1 through v_n tell us how to write v as a linear combination of the standard basis vectors:

$$v = \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n.$$

If we chose another basis for $\mathbb{R}^n$, however, like

$$\mathcal{B} = \{b_1, b_2, ..., b_n\},$$

then we would represent v as a linear combination with different scalars:

$$v = \mu_1 b_1 + \mu_2 b_2 + \cdots + \mu_n b_n.$$

It would be reasonable for us to then write the vector v as

$$v = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}_{\mathcal{B}}$$

where we use the subscript $\mathcal{B}$ to tell us that these coordinates are with respect to the $\mathcal{B}$ basis.

Example 9.9.
Consider the following basis for $\mathbb{R}^3$:

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} \right\}$$

The vector

$$v = \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix}$$

in this basis can be written as

$$v = \begin{pmatrix} 5 \\ -3 \\ -1 \end{pmatrix}_{\mathcal{B}}$$

as

$$v = 5\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} - 3\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}.$$

We can go back and forth between two different bases by using matrix multiplication. For example, suppose we know that the coordinates of a vector v with respect to the standard basis are $v = \begin{pmatrix} \lambda_1 & \cdots & \lambda_n \end{pmatrix}^T$. How can we find the coordinates for the basis $\mathcal{B}$? We need to set this up as a system of equations: the coordinates with respect to $\mathcal{B}$, call them μ_1, μ_2, ..., μ_n should have the property that $v = \mu_1 b_1 + \cdots + \mu_n b_n$. If we use the b_i vectors as the columns of a matrix A, and use the μ_i's as components of a vector μ, then we have the equation $v = A\mu$. We want to solve this for μ, however, so we'll multiply both sides by A^{-1} to obtain $\mu = A^{-1}v$ where v. (Notice that A will always be invertible: since the b_i vectors form a basis, the columns of A are linearly independent.)

Example 9.10.
Consider the following basis for $\mathbb{R}^2$:

$$\mathcal{B} = \left\{ \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}.$$

We can convert any vector v from standard coordinates (the coordinates given using the standard basis vectors),

$$v = \begin{pmatrix} x \\ y \end{pmatrix}$$

to coordinates with respect to the $\mathcal{B}$ basis by multiplying by

$$A^{-1} = \begin{pmatrix} 3 & -1 \\ 2 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1/5 & 1/5 \\ -2/5 & 3/5 \end{pmatrix}.$$

Hence in $\mathcal{B}$-coordinates,

$$v = \begin{pmatrix} x/5 + y/5 \\ -2x/5 + 3y/5 \end{pmatrix}_{\mathcal{B}}$$

The matrix A^{-1} above is called the ***change of basis*** matrix from the standard basis to the $\mathcal{B}$ basis. There isn't a single standard notation for this matrix. In the Lay textbook this matrix is denoted $\mathcal{P}_{\mathcal{B}}^{-1}$ (so $\mathcal{P}_{\mathcal{B}}$ is the matrix obtained by using the vectors of $\mathcal{B}$ as the columns of the matrix); another common notation which we will use in class is ${}_{\mathcal{B}}I_{\mathcal{S}}$. Here we will use $\mathcal{S}$ to mean the standard basis $\{e_1, e_2, ..., e_n\}$. Thus

$${}_{\mathcal{B}}I_{\mathcal{S}} = \left(b_1 \,\middle|\, b_2 \,\middle|\, \cdots \,\middle|\, b_n \right)^{-1}$$

If $v = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_n \end{pmatrix}^T$ with respect to the standard basis, then with respect to $\mathcal{B}$ the coordinates of v are

$$[v]_{\mathcal{B}} = {}_{\mathcal{B}}I_{\mathcal{S}} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}$$

where $[v]_{\mathcal{B}}$ means the coordinates of v with respect to the basis $\mathcal{B}$.

We can find the coordinates of v with respect to the standard basis by multiplying by the inverse of ${}_{\mathcal{B}}I_{\mathcal{S}}$ which we denote ${}_{\mathcal{S}}I_{\mathcal{B}} = {}_{\mathcal{B}}I_{\mathcal{S}}^{-1}$.

$$[v] = {}_{\mathcal{S}}I_{\mathcal{B}}[v]_{\mathcal{B}}$$

where $[v]$ denotes the vector of coordinates for v with respect to the standard basis, $\mathcal{S}$.

More generally, if we have two different bases $\mathcal{B}_1$ and $\mathcal{B}_2$ we can go back and forth between $\mathcal{B}_1$-coordinates and $\mathcal{B}_2$ coordinates by multiplying by by a matrix ${}_{\mathcal{B}_2}I_{\mathcal{B}_1}$ where

$${}_{\mathcal{B}_2}I_{\mathcal{B}_1} = {}_{\mathcal{B}_2}I_{\mathcal{S}}\, {}_{\mathcal{S}}I_{\mathcal{B}_1}.$$

That is, we switch from $\mathcal{B}_1$-coordinates to standard coordinates, then from standard coordinates to $\mathcal{B}_2$-coordinates.

$$[v]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}I_{\mathcal{B}_1}[v]_{\mathcal{B}_1}.$$

Example 9.11.
Consider the following two bases of $\mathbb{R}^3$:

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$$

$$\mathcal{B}_2 = \left\{ \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \right\}$$

The change of basis matrices between the $\mathcal{B}_i$ and $\mathcal{S}$ bases are

$$ {}_{\mathcal{B}_1}I_{\mathcal{S}} = \begin{pmatrix} 0 & 1 & -1 \\ 1/3 & -4/3 & 4/3 \\ -1/3 & 1/3 & 2/3 \end{pmatrix} \qquad {}_{\mathcal{S}}I_{\mathcal{B}_1} = \begin{pmatrix} 4 & 3 & 0 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} $$

$$ {}_{\mathcal{B}_2}I_{\mathcal{S}} = \begin{pmatrix} -5/4 & 3/2 & 1/4 \\ 1 & -1 & 0 \\ 1/2 & 0 & -1/2 \end{pmatrix} \qquad {}_{\mathcal{S}}I_{\mathcal{B}_2} = \begin{pmatrix} 2 & 3 & 1 \\ 2 & 2 & 1 \\ 2 & 3 & -1 \end{pmatrix} $$

Hence the $\mathcal{B}_1$-to-$\mathcal{B}_2$ change of basis matrix is

$$ {}_{\mathcal{B}_2}I_{\mathcal{B}_1} = {}_{\mathcal{B}_2}I_{\mathcal{S}}\,{}_{\mathcal{S}}I_{\mathcal{B}1} = \begin{pmatrix} -7/4 & -2 & 7/4 \\ 2 & 2 & -1 \\ 3/2 & 1 & -1/2 \end{pmatrix} $$

We can now convert from $\mathcal{B}_1$ to $\mathcal{B}_2$ coordinates using this matrix. For example, if v is a vector which in $\mathcal{B}_1$ coordinates is given by

$$ v = \begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix}_{\mathcal{B}_1} $$

then

$$ [v]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}I_{\mathcal{B}_1} = \begin{pmatrix} -7/4 & -2 & 7/4 \\ 2 & 2 & -1 \\ 3/2 & 1 & -1/2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} -9/4 \\ 6 \\ 9/2 \end{pmatrix}_{\mathcal{B}_2} $$

9.4 Matrices of Linear Transformations

When we write a linear transformation as a matrix, the matrix we use depends on the choice of basis. Recall that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then T is represented by an $m \times n$ matrix A with the property that $T(v) = Av$. The columns of this matrix are exactly the coordinates (with respect to the standard basis vectors) of $T(e_1)$, $T(e_2)$, ..., $T(e_n)$. If we were using a different basis, though, what should this matrix look like?

There is a very easy way to determine what the matrix should be: we can switch from some basis $\mathcal{B}$ to the standard basis where we know what the matrix is, then switch back to $\mathcal{B}$. That is, if A represents the linear transformation with respect to the basis $\mathcal{B}$, then $A_\mathcal{B}$, the matrix with respect to $\mathcal{B}$, should be

$$A_\mathcal{B} = {}_\mathcal{B}I_\mathcal{S} A \, {}_\mathcal{S}I_\mathcal{B}.$$

So, when we multiply $[v]_\mathcal{B}$ by $A_\mathcal{B}$ we have

$$\begin{aligned} A_\mathcal{B}[v]_\mathcal{B} &= {}_\mathcal{B}I_\mathcal{S} A \, {}_\mathcal{S}I_\mathcal{B}[v]_\mathcal{B} \\ &= {}_\mathcal{B}I_\mathcal{S} A[v]_\mathcal{S} \\ &= {}_\mathcal{B}I_\mathcal{S}[T(v)]_\mathcal{S} \\ &= [T(v)]_\mathcal{B}. \end{aligned}$$

Remark.
Notice that ${}_\mathcal{B}I_\mathcal{S} = {}_\mathcal{S}I_\mathcal{B}^{-1}$. Let's momentarily denote this matrix J. Then the matrix $A_\mathcal{B}$ has the form

$$J^{-1}AJ.$$

This operation of taking A and multiplying it on one side by some J and on the other side by J^{-1} is sometimes called ***conjugation*** and appears in many different areas of mathematics. Conjugation basically means "do something (e.g., apply A) somewhere else (e.g., in a different coordinate system)."

Example 9.12.
Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the linear transformation which in standard coordinates is represented by the matrix

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

Now suppose that $\mathcal{B}$ is the basis of $\mathbb{R}^2$ given by

$$\mathcal{B} = \left\{ \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix} \right\}.$$

Then

$${}_{\mathcal{B}}I_{\mathcal{S}} = \begin{pmatrix} 1 & -3/2 \\ -1 & 2 \end{pmatrix} \qquad {}_{\mathcal{S}}I_{\mathcal{B}} = \begin{pmatrix} 4 & 3 \\ 2 & 2 \end{pmatrix}$$

Then with respect to the basis $\mathcal{B}$, the linear transformation T is given by

$$A_{\mathcal{B}} = {}_{\mathcal{B}}I_{\mathcal{S}} A\, {}_{\mathcal{S}}I_{\mathcal{B}} = \begin{pmatrix} 10 & 17/2 \\ -8 & -7 \end{pmatrix}$$

Example 9.13.
Suppose $T : \mathbb{R}^3 \to \mathbb{R}^3$ is the linear transformation which in standard coordinates is given by the matrix

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

and consider the basis of $\mathcal{B}$ of $\mathbb{R}^3$ given by

$$\mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

The change of basis matrices are

$${}_{\mathcal{B}}I_{\mathcal{S}} = \begin{pmatrix} 3/4 & -1/4 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad {}_{\mathcal{S}}I_{\mathcal{B}} = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Then with respect to the $\mathcal{B}$ basis, T is given by the matrix

$$A_{\mathcal{B}} = \begin{pmatrix} 5/2 & 13/4 & -1/4 \\ -1 & -5/2 & 1/2 \\ 2 & 3 & 1 \end{pmatrix}$$

We can extend the idea above to linear transformations between spaces of different dimensions. That is, suppose we choose a basis $\mathcal{B}$ for $\mathbb{R}^n$ and a basis $\mathcal{C}$ for $\mathbb{R}^m$ and we want to express the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ with respect to these bases. If T is represented by the $m \times n$ matrix A in the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, which we'll denote $\mathcal{S}_n$ and $\mathcal{S}_m$ respectively, then with respect to the $\mathcal{B}$ and $\mathcal{C}$ bases we have the matrix

$${}_{\mathcal{C}}A_{\mathcal{B}} = {}_{\mathcal{C}}I_{\mathcal{S}_m}\, A \; {}_{\mathcal{S}_n}I_{\mathcal{B}}.$$

That is, we again convert from the $\mathcal{B}$ basis of $\mathbb{R}^n$ to the standard basis, apply the matrix A, then convert from the standard basis of $\mathbb{R}^m$ to the $\mathcal{C}$ basis.

Example 9.14.
Suppose $T : \mathbb{R}^3 \to \mathbb{R}^2$ is given with respect to the standard bases by

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 0 & 1 \end{pmatrix}.$$

What matrix represents this linear transformation with respect to the following $\mathcal{B}$ and $\mathcal{C}$ bases for $\mathbb{R}^3$ and $\mathbb{R}^2$ given by

$$\mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

$$\mathcal{C} = \left\{ \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix} \right\}$$

The change of basis matrices are

$$_{\mathcal{S}_3}I_{\mathcal{B}} = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad _{\mathcal{C}}I_{\mathcal{S}_2} = \begin{pmatrix} 1 & -{}^{3}\!/_{2} \\ -1 & 2 \end{pmatrix}$$

The matrix $_{\mathcal{C}}A_{\mathcal{B}}$ is then

$$_{\mathcal{C}}A_{\mathcal{B}} = {}_{\mathcal{C}}I_{\mathcal{S}_2}\, A \, {}_{\mathcal{S}_3}I_{\mathcal{B}} = \begin{pmatrix} 2 & 7 & -{}^{1}\!/_{2} \\ 0 & -6 & 1 \end{pmatrix}$$

9.5 Column Space, Null Space, and Rank

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then ***image*** or T (also called the ***range*** of T) is the collection of all the vectors in $\vec{w} \in \mathbb{R}^m$ such that there exists a $\vec{v} \in \mathbb{R}^n$ with $T(\vec{v}) = \vec{w}$.

Example 9.15.
Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ x-y \\ 2x \end{pmatrix}$$

Then the vector

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

is in the image of T since

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

However, the vector

$$\begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}$$

is not in the image: there is no vector $\vec{v} \in \mathbb{R}^2$ such that $T(\vec{v})$ equals the vector above.

Exercise 9.3.
Convince yourself that

$$\begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}$$

is not in the image of

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ x-y \\ 2x \end{pmatrix}.$$

To really justify this you need to show there is no solution to some system of linear equations. Figure out what that system of equations is, and then show there is no solution.

The image (aka range) of a linear transformation doesn't have a very standard notation, but some relatively common notations are $T(\mathbb{R}^n)$, $\mathrm{range}(T)$ and $\mathrm{im}(T)$.

Theorem 9.4.
Given any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, the image $\mathrm{im}(T)$ is a subspace of $\mathbb{R}^m$.

Proof.
We need to show that $\operatorname{im}(T)$ is closed under vector addition and scalar multiplication. Suppose $\vec{v}_1, \vec{v}_2 \in \operatorname{im}(T)$. Then, by definition of the image, there must exist vectors $\vec{w}_1, \vec{w}_2 \in \mathbb{R}^n$ such that $T(\vec{w}_1) = \vec{v}_1$ and $T(\vec{w}_2) = \vec{v}_2$. Notice

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

so $\operatorname{im}(T)$ is closed under vector addition.

Similarly,

$$\begin{aligned} \lambda\vec{v}_1 &= \lambda T(\vec{w}_1) \\ &= T(\lambda\vec{w}_1) \end{aligned}$$

and so $\operatorname{im}(T)$ is closed under scalar multiplication. □

Example 9.16.
Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ x-y \\ 2x \end{pmatrix}.$$

Since T is given by the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 0 \end{pmatrix}$$

every vector $T(\vec{v})$ is a linear combination of the columns of this matrix. That is,

$$\operatorname{im}(T) = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}.$$

More generally, given any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, the image of T consists of all linear combinations of the columns of the matrix representing T, and so $\mathrm{im}(T)$ is the span of the columns of the matrix representing T. For this reason the image of a linear transformation is sometimes also called the ***column space*** of the corresponding matrix.

The columns of the matrix do not in general form a basis for the column space (aka image of T, aka range of T) because they may not be linearly independent. However some subset of these columns will form a basis: namely the largest linearly independent subset of columns of the matrix forms a basis for the column space.

Theorem 9.5.
The column space of a matrix A (which is synonymous with the image, or range, of the corresponding linear transformation) has a basis given by the pivot columns of A.

Proof.
We simply need to show that the pivot columns of A are a maximal linearly independent set. That is, we need to show the pivot columns are linearly independent, and every column of A can be written as a linear combination of the pivot columns.

Suppose that A is an $m \times n$ matrix and for simplicity suppose the first k columns, $\vec{a}_1$, $\vec{a}_2$, ..., $\vec{a}_k$, are the pivot columns of A. Performing row operations turns $\vec{a}_1$, $\vec{a}_2$, ..., $\vec{a}_k$ into the first k standard basis vectors of $\mathbb{R}^m$, $\vec{e}_1$, $\vec{e}_2$, ..., $\vec{e}_k$. If the columns $\vec{a}_1$, ..., $\vec{a}_k$ were not linearly independent, then there is some non-zero solution to the homogeneous system

$$\begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_k \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} = \vec{0}.$$

Multiplying on the left by the product of the elementary matrices which take $\vec{a}_1$, ..., $\vec{a}_k$ to $\vec{e}_1$, ..., $\vec{e}_k$, we would then have a non-trivial solution to

$$\begin{pmatrix} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_k \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} = E \begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_k \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} = \vec{0}.$$

But this is clearly impossible since the $\vec{e}_i$ are linearly independent. Hence the pivot columns of the matrix must be linearly independent as well.

We now need to show that adding a non-pivot column would give us a linearly dependent set. Suppose $\vec{a}$ is any column of A which is not a pivot column. Then the homogeneous system

$$\begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_k & \vec{a} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \\ \lambda \end{pmatrix} = \vec{0}$$

has a non-trivial solution since not all of the columns are pivot columns. That is,

$$\lambda_1 \vec{a}_1 + \lambda_2 \vec{a}_2 + \cdots + \lambda_k \vec{a}_k + \lambda \vec{a} = \vec{0}$$

Has some solution where not all λ_1, ..., λ_k, λ are zero. Notice that in fact λ must not be zero: if $\lambda = 0$ then the above reduces to $\lambda_1 \vec{a}_1 + \cdots + \lambda_k \vec{a}_k = \vec{0}$ which we know only has the trivial solution. Since $\lambda \neq 0$, we can solve the above for $\vec{a}$, writing $\vec{a}$ as a linear combination of the other vectors:

$$\vec{a} = -\frac{\lambda_1}{\lambda} \vec{a}_1 - \frac{\lambda_2}{\lambda} \vec{a}_2 - \cdots - \frac{\lambda_k}{\lambda} \vec{a}_k.$$

Hence this collection of columns is not linearly independent.

To summarize: we have shown that the pivot columns of A are linearly independent, but every non-pivot column can be written as a linear combination of the pivot columns. Hence the span of all of the columns is equal to the span of the pivot columns, and since these are linearly independent they form a basis for this span. □

Example 9.17.
What is a basis for the column span of the following matrix?

$$A = \begin{pmatrix} 3 & 2 & 6 & 6 & 2 & 13 \\ 2 & 3 & 9 & 4 & 1 & 8 \\ -1 & 1 & 3 & -2 & -1 & -5 \\ 0 & 6 & 18 & 0 & -1 & -2 \end{pmatrix}$$

A basis for the column span is given by the pivot columns. To determine the pivot columns we put A into RREF to obtain

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

Notice that the first, second, and fifth columns are the pivot columns. Hence the column span of our matrix has basis given by the first, second, and fifth columns *of the original matrix*:

$$\left\{ \begin{pmatrix} 3 \\ 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \\ 6 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -1 \\ -1 \end{pmatrix} \right\}$$

The dimension of the column space (or range or image) is called the ***rank*** of the matrix (or linear transformation) and is denoted $\text{rank}(T)$ or $\text{rank}(A)$ for a linear transformation T or matrix A.

Related to the column space of a matrix is the *null space*, also called the *kernel*. Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, the ***kernel*** of T, denoted $\ker(T)$, is the collection of all the vectors $\vec{v} \in \mathbb{R}^n$ such that $T(\vec{v}) = \vec{0}$.

$$\ker(T) = \left\{ \vec{v} \in \mathbb{R}^n \,\middle|\, T(\vec{v}) = \vec{0} \right\}.$$

Theorem 9.6.
$\ker(T)$ *is a subspace of* $\mathbb{R}^n$.

Proof.
Suppose $\vec{v}, \vec{w} \in \ker(T)$. That is, $T(\vec{v}) = T(\vec{w}) = \vec{0}$. Then

$$T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w}) = \vec{0} + \vec{0} = \vec{0}.$$

Similarly, for any $\lambda \in \mathbb{R}$,

$$T(\lambda\vec{v}) = \lambda T(\vec{v}) = \lambda\vec{0} = \vec{0}.$$

□

Notice that if T is represented by the matrix A, the kernel of T is precisely the set of solutions to the homogeneous equation $A\vec{x} = \vec{0}$.

Notice that finding a basis for the kernel of T is essentially the same as parametrizing the solutions to $A\vec{x} = \vec{0}$.

Example 9.18.
Find a basis for the kernel of the linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ given by the following matrix:

$$A = \begin{pmatrix} 3 & 1 & -5 & -1 \\ 0 & 1 & 1 & -1 \\ 2 & 1 & -3 & -1 \end{pmatrix}$$

If we put the matrix into RREF we have

$$\begin{pmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

This $A\vec{x} = \vec{0}$ is equivalent to the system

$$\begin{aligned} x_1 - 2x_4 &= 0 \\ x_2 + x_3 - x_4 &= 0 \end{aligned}$$

Thus

$$x_1 = 2x_4$$
$$x_2 = -x_3 + x_4$$

and x_3 and x_4 are free variables. So, our vectors $\vec{x}$ solving the equation look like

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2x_4 \\ -x_3 + x_4 \\ x_3 \\ x_4 \end{pmatrix}$$

which we can write as

$$\vec{x} = x_3 \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

So the solutions to our equation are precisely the vectors in

$$\text{span}\left(\left\{\begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}\right\}\right).$$

That is, a basis for the kernel is

$$\left\{\begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}\right\}$$

Just as the dimension of the image $\text{im}(T)$ has a special name (the *rank* of T), the dimension of the kernel $\ker(T)$ has a special name: it is called the ***nullity*** of T and we will denote it $\text{null}(T)$:

$$\text{null}(T) = \dim(\ker(T)).$$

9.6 The Rank-Nullity Theorem

One of, if not *the*, most fundamental result in linear algebra is the *rank-nullity theorem* which tells us how the rank and nullity of a linear transformation are related. It is hard to overstate how important the following result is, not just in linear algebra but in many other areas that use linear algebraic ideas.

Theorem 9.7 (The Rank-Nullity Theorem)**.**
If $T : \mathbb{R}^n \to \mathbb{R}^m$ is any linear transformation, then

$$\operatorname{rank}(T) + \operatorname{null}(T) = n.$$

Proof.
For simplicity let's write $N = \operatorname{null}(T)$, so we want to show that

$$\operatorname{rank}(T) + N = n$$

Suppose that $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_N\}$ is a basis for the kernel of T. We can get a basis for all of $\mathbb{R}^n$ by adding more vectors. Let's say we add M new vectors, $\vec{w}_1$, $\vec{w}_2$, ..., $\vec{w}_M$. That is, we have a basis

$$\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_N, \vec{w}_1, \vec{w}_2, ..., \vec{w}_M\}$$

is a basis for $\mathbb{R}^n$. Since $\mathbb{R}^n$ is n-dimensional, this must mean that

$$N + M = n.$$

It's clear that

$$\{T(\vec{v}_1), T(\vec{v}_2), ..., T(\vec{v}_N), T(\vec{w}_1), T(\vec{w}_2), ..., T(\vec{w}_M)\}$$

spans $\operatorname{im}(T)$. But notice that

$$T(\vec{v}_1) = T(\vec{v}_2) = \cdots = T(\vec{v}_N) = \vec{0}.$$

Thus

$$\operatorname{im}(T) = \operatorname{span}\left(\{T(\vec{w}_1), T(\vec{w}_2), ..., T(\vec{w}_M)\}\right).$$

Hence $M \geq \text{rank}(T)$. If we can show that the $T(\vec{w}_i)$ vectors are linearly independent, that will imply that in fact $M = \text{rank}(T)$ since $\{T(\vec{w}_1), ..., T(\vec{w}_M)\}$ would form a basis for $\text{im}(T)$ which we know has dimension $\text{rank}(T)$.

Suppose that we have λ_1, ..., λ_M satisfying the following equation:

$$\lambda_1 T(\vec{w}_1) + \lambda_2 T(\vec{w}_2) + \cdots + \lambda_M T(\vec{w}_M) = \vec{0}$$

Since T is linear we can rewrite this as

$$T\left(\lambda_1 \vec{w}_1 + \lambda_2 \vec{w}_2 + \cdots + \lambda_M \vec{w}_M\right) = \vec{0}.$$

But this means

$$\lambda_1 \vec{w}_1 + \lambda_2 \vec{w}_2 + \cdots + \lambda_M \vec{w}_M \in \ker(T).$$

However the vectors $\vec{w}_1$, ..., $\vec{w}_M$ are linearly independent of the vectors $\vec{v}_1$, ..., $\vec{v}_\rho$ that formed a basis for $\ker(T)$. The only possibility, then, is that $\lambda_1 = \cdots = \lambda_M = 0$. That is, the vectors are linearly independent and form a basis for $\text{im}(T)$, which means $M = \text{rank}(T)$.

We already know $M + N = n$, but $M = \text{rank}(T)$ and $N = \text{null}(T)$, so $\text{rank}(T) + \text{null}(T) = n$. □

With the rank-nullity theorem at our disposal, we are now able to go back and prove some of the theorems and propositions from earlier in the semester that we had to take on faith earlier, such as the following (Proposition 1 from Lecture 5).

Proposition 9.8.
The number of free variables in the solution set of $A\vec{x} = \vec{0}$ is precisely the number of columns of th ematrix that do not contain a pivot.

Proof.
Parametrizing the solution set of $A\vec{x} = 0$ is the same as finding a basis for the nullspace of A, with each basis vector for the nullspace corresponding to a free variable in the parametrization. That is, the

dimension of the nullspace is the number of free variables in the parametrization. By the rank nullity theorem, however, the dimension of the nullspace is exactly n (the number of columns of A) minus $\mathrm{rank}(A)$ (the number of pivot columns of A), which is simply the number of columns without a pivot. $\square$

9.7 Practice Problems

Problem 9.1.
Suppose that $V \subseteq \mathbb{R}^3$ consists of all vectors $\begin{pmatrix} x & y & z \end{pmatrix}^T$ whose components satisfy the equation $3x - 2y + z = 0$. Is V a subspace of $\mathbb{R}^3$?

Problem 9.2.
Suppose that $V \subseteq \mathbb{R}^3$ consists of all vectors $\begin{pmatrix} x & y & z \end{pmatrix}^T$ whose components satisfy the equation $3x - 2y + z = 1$. Is V a subspace of $\mathbb{R}^3$?

Problem 9.3.
Suppose that $V \subseteq \mathbb{R}^2$ consists of all vectors $\begin{pmatrix} x & y \end{pmatrix}^T$ whose components satisfy the equation $x^2 - y = 0$. Is V a subspace of $\mathbb{R}^2$?

Problem 9.4.
Suppose that U and V are two subspaces of $\mathbb{R}^n$. Is the intersection $U \cap V$ necessarily also a subspace of $\mathbb{R}^n$?

Problem 9.5.
Suppose that U and V are two subspaces of $\mathbb{R}^n$. Is the union $U \cup V$ necessarily also a subspace of $\mathbb{R}^n$?

Problem 9.6.
Consider the linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^5$ given by

$$T \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} w - x + 3y \\ y + z \\ x - w \\ 2w - 4z \\ -w + x + y + z \end{pmatrix}$$

Find a basis for the image and kernel of this linear transformation.

Problem 9.7.
In each of the problems below, find the coordinates of the given vector v (given with respect to the standard basis) using the basis $\mathcal{B}$ of $\mathbb{R}^n$.

(a)

$$v = \begin{pmatrix} -2 \\ 3 \end{pmatrix} \qquad \mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 5 \end{pmatrix} \right\}$$

(b)

$$v = \begin{pmatrix} -2 \\ 3 \end{pmatrix} \qquad \mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \end{pmatrix} \right\}$$

(c)

$$v = \begin{pmatrix} 3 \\ 7 \\ 0 \end{pmatrix} \qquad \mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} \right\}$$

Problem 9.8.
In each of the problems below you are given a matrix A which represents some linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ with respect to the standard basis, as well as bases $\mathcal{B}$ of $\mathbb{R}^n$ and $\mathcal{C}$ of $\mathbb{R}^m$. Find the matrix representing T with respect to $\mathcal{B}$ and $\mathcal{C}$.

(a)

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

$$\mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 7 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\}$$

(b)

$$A = \begin{pmatrix} 1 & 6 & 0 \\ 3 & 2 & 1 \end{pmatrix}$$

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right\}$$

(c)

$$A = \begin{pmatrix} 1 & 4 & 3 \\ 1 & 7 & 2 \end{pmatrix}$$

$$\mathcal{B} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

(d)

$$A = \begin{pmatrix} 1 & 4 & 3 \\ 1 & 7 & 2 \end{pmatrix}$$

$$\mathcal{B} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$

10 Eigenvectors & Eigenvalues

10.1 Introduction

Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, we should typically expect that T will move the vectors around in $\mathbb{R}^n$ in a somewhat complicated way. However, it may happen that some vectors only get stretched out (multiplied by a scalar). For example, consider the map $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by the matrix

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 1 & 2 \\ 0 & 1 & 0 \end{pmatrix}.$$

It's not immediately obvious, but it's easy to check that this matrix leaves the vector

$$u = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

alone; that is, $Au = u$:

$$Au = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 1 & 2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} = - \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} = u.$$

Similarly, the vector

$$v = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

is simply negated; that is, $Av = -v$:

$$Av = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 1 & 2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} = 0 \cdot \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = -v$$

And finally, the vector

$$w = \begin{pmatrix} 5 \\ 4 \\ 1 \end{pmatrix}$$

gets multiplied by a factor of 4:

$$
\begin{aligned}
Aw =& \begin{pmatrix} 3 & 1 & 1 \\ 2 & 1 & 2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 5 \\ 4 \\ 1 \end{pmatrix} \\
=& 5 \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} + 4 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \\
=& \begin{pmatrix} 20 \\ 16 \\ 4 \end{pmatrix} \\
=& 4 \begin{pmatrix} 5 \\ 4 \\ 1 \end{pmatrix} \\
=& 4w.
\end{aligned}
$$

At first glance this may not seem like the most useful observation, but notice that these vectors u, v, and w form a basis for $\mathbb{R}^3$. Thus every vector can be written as some linear combination of these vectors,

$$\alpha u + \beta v + \gamma w.$$

It is now extremely easy to describe how our linear transformation acts on these vectors:

$$A(\alpha u + \beta v + \gamma w) = \alpha u - \beta v + 4\gamma w.$$

That is, if we write our vector as

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}_{\mathcal{B}}$$

then

$$T \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}_{\mathcal{B}} = \begin{pmatrix} \alpha \\ -\beta \\ 4\gamma \end{pmatrix}_{\mathcal{B}}.$$

Thus having vectors which are simply stretched out by the linear transformation can make it extremely easy to describe the linear transformation, and this can make studying linear transformations we may have interest in considerably easier.

10.2 Eigenvectors and Eigenvalues

Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, we say that a vector $v \neq 0$ is an ***eigenvector*** of T with ***eigenvalue*** λ if v and λ satisfy the following equation:

$$T(v) = \lambda v.$$

That is, eigenvectors are precisely the vectors that simply get stretched out by T and the eigenvalue tells us how much the vector gets stretched out.

Exercise 10.1.
Notice that eigenvectors are by definition never zero, but eigenvalues *are* allowed to be zero. Show that a linear transformation will have zero as an eigenvalue if and only if the linear transformation is not injective.

Remark.
The word *eigen* is an adjective in German that means something like "owned by." So the eigenvectors and eigenvalues are the vectors and scalars "owned" by the linear transformation.

The vectors u, v, and w from the example in the introduction are thus eigenvectors of the linear transformation with matrix

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

while 1, -1, and 4 are the respective eigenvalues.

Perhaps that first thing to notice about eigenvectors and eigenvalues is that they are somewhat special: not every linear transformation will have eigenvector and eigenvalues.

Example 10.1.
The linear transformation in $\mathbb{R}^2$ with matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

does not have any eigenvectors/eigenvalues. The easiest way to see this is to think geometrically: this matrix acts on the plane $\mathbb{R}^2$ by 90° rotations, and no non-zero vector in the plane is simply stretched out by a 90° rotation.

The other thing to notice about eigenvectors is that they come in families. For example, the vector

$$w = \begin{pmatrix} 5 \\ 4 \\ 1 \end{pmatrix}$$

from before is an eigenvector with eigenvalue 4 of the matrix

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

Notice that any scalar multiple λw is also an eigenvector with eigenvalue 4: As $Aw = 4w$ we have $A\lambda w = \lambda A w = \lambda 4w = 4\lambda w$. The collection of all eigenvectors of a given eigenvalue forms a subspace of $\mathbb{R}^n$ called the ***eigenspace*** of T with the given eigenvalue.

Lemma 10.1.
Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and suppose λ is an eigenvalue of T. The set of all eigenvectors of T with eigenvalue λ is a subspace of $\mathbb{R}^n$, provided this collection of eigenvectors is not empty (i.e., that λ really is an eigenvalue).

Proof.
The set of all eigenvectors with eigenvalue λ is precisely the set of all vectors $v \in \mathbb{R}^n$ satisfying $T(v) = \lambda v$. Let's momentarily denote this set U:

$$U = \left\{ v \in \mathbb{R}^n \,\middle|\, T(v) = \lambda v \right\}.$$

Since we're already assuming U is non-empty, we need to show that U closed under vector addition, and closed under scalar multiplication. Suppose $u, u' \in U$ and μ is any scalar.

Checking that U is closed under vector addition is easy:

$$T(u + u') = T(u) + T(u') = \lambda u + \lambda u' = \lambda(u + u'),$$

as is checking that U is closed under scalar multiplication:

$$T(\mu u) = \mu T(u) = \mu \lambda u = \lambda \cdot (\mu u).$$

Thus U, the eigenspace of vectors in $\mathbb{R}^n$ which are eigenvectors of T with eigenvalue λ, is a subspace of $\mathbb{R}^n$. □

10.3 Computing Eigenvectors and Eigenvalues

The question now is how do we go about finding the eigenvectors and eigenvalues of a matrix. This is a two-step process: first we have to find the eigenvalues λ, and then for each eigenvalue we need to find the associated eigenvectors.

If we want to find eigenvalues of T, then we need to find the scalars λ for which there is a solution v to the equation $T(v) = \lambda v$. For simplicity, let's suppose our linear transformation has domain $\mathbb{R}^n$ and codomain $\mathbb{R}^n$ so that we can represent T by an $n \times n$ matrix A. We then want to find the scalars λ for which there is a solution to

$$Av = \lambda v.$$

Equivalently, we want to find the λ's for which there is a solution to

$$Av - \lambda v = 0.$$

We can rewrite $Av - \lambda v$ as $(A - \lambda I)v$: just distribute the v and notice that λI is the matrix with all zeros except for λ's on the diagonal, thus

$(\lambda I)v = \lambda v$. So we want to find the λ's for which there is a non-zero v solving

$$(A - \lambda I)v = 0.$$

Since $A - \lambda I$ is an $n \times n$ matrix, this equation has a non-zero solution precisely when $A - \lambda I$ is not invertible: i.e., there is a non-zero solution exactly when $\det(A - \lambda I) = 0$. Long story short, we have the following:

Proposition 10.2.
A scalar λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

Example 10.2.
Find the eigenvalues of the matrix

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

We want to find the values of λ for which

$$\det(A - \lambda I) = \det\left(\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) = \det\left(\begin{pmatrix} 2-\lambda & 3 \\ 2 & 1-\lambda \end{pmatrix}\right)$$

is zero.

$$\begin{aligned}
\det\left(\begin{pmatrix} 2-\lambda & 3 \\ 2 & 1-\lambda \end{pmatrix}\right) &= 0 \\
\implies (2-\lambda)(1-\lambda) - 6 &= 0 \\
\implies 2 - 3\lambda + \lambda^2 - 6 &= 0 \\
\implies \lambda^2 - 3\lambda - 4 &= 0 \\
\implies (\lambda - 4)(\lambda + 1) &= 0
\end{aligned}$$

Thus our eigenvalues are $\lambda = 4$ and $\lambda = -1$.

Once the eigenvalues of A are known, we can then search for the eigenvectors.

Example 10.3.
Find the eigenvectors associated with eigenvalue $\lambda = 4$ for the matrix

$$A = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}.$$

We are trying to find the solutions to $Av = 4v$, or equivalently $(A - 4I)v = 0$. That is, we want to find the solutions to the homogeneous system

$$\begin{pmatrix} 2-4 & 3 \\ 2 & 1-4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2 & 3 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Putting the coefficient matrix into RREF we have

$$\begin{pmatrix} 1 & -3/2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Thus the eigenspace of this matrix, associated with the eigenvector 4, is

$$\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \,\middle|\, x - \frac{3}{2}y = 0 \right\}$$

Exercise 10.2.
Find the eigenvectors associated with eigenvalue $\lambda = -1$ for the matrix

$$A = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}.$$

Perhaps unsurprisingly (since eigenvalues are related to determinants), eigenvalues for triangular matrices are very easy to compute.

Theorem 10.3.
If A is a triangular matrix, then the eigenvalues of A are the entries on the diagonal.

Proof.
Suppose that A is a triangular matrix with diagonal entries a_{11}, a_{22}, ..., a_{nn}. Then $A - \lambda I$ is also a triangular matrix, but with diagonal entries $a_{11} - \lambda$, $a_{22} - \lambda$, ..., $a_{nn} - \lambda$. Thus

$$\det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda)\cdots(a_{nn} - \lambda).$$

This will only be zero when one of the factors is zero, that is when λ equals a_{ii} for some diagonal entry a_{ii}. □

Example 10.4.
The eigenvalues of

$$\begin{pmatrix} 3 & 2 & 1 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

are 3, -1, 0, 5, and 3.

Example 10.5.
Find the eigenvalues and associated eigenvectors of the following matrix:

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thanks to the previous theorem we can easily determine that the eigenvalues are 2 and 1. Now we simply need to find the associated eigenvectors.

For the eigenvalue $\lambda = 2$ we need to solve the equation $(A - 2I)v = 0$,

$$\begin{pmatrix} 0 & 1 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Putting the matrix into RREF, this system is equivalent to

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Thus the eigenspace associated to 2 is

$$\left\{ \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} \middle| \, x \in \mathbb{R} \right\}$$

For the eigenvalue $\lambda = 1$ we need to solve the equation $(A - I)v = 0$,

$$\begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

This is equivalent to solving the system

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and so the eigenspace associated to 1 is

$$\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \, x = -2z, y = -z, z \in \mathbb{R} \right\}$$

In our motivating example at the start of the lecture, notice that the eigenvectors for

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

formed a basis for $\mathbb{R}^3$. Even though this doesn't happen in Example 10.4, notice that the eigenvectors associated to different eigenvalues are linearly independent. This is true in general.

Theorem 10.4.
If v_1, v_2, ..., v_m are eigenvectors associated with distinct eigenvalues λ_1, λ_2, ..., λ_m of A, then $\{v_1, v_2, ..., v_m\}$ is a linearly independent set.

Proof.
Suppose instead that $\{v_1, v_2, ..., v_m\}$ is linearly dependent. Rearranging the order of eigenvectors and eigenvalues if necessary, we may assume that $\{v_1, v_2, ..., v_r\}$ is linearly independent (it could be that $r = 1$) and $\{v_1, v_2, ..., v_r, v_{r+1}\}$ is linearly dependent. That is, there exist scalars μ_1 through μ_r such that

$$v_{r+1} = \mu_1 v_1 + \cdots + \mu_r v_r.$$

If we apply A to both sides of the equation we have

$$\begin{aligned} Av_{r+1} &= A(\mu_1 v_1 + \cdots + \mu_r v_r) \\ \Longrightarrow \lambda_{r+1} v_{r+1} &= \mu_1 \lambda_1 v_1 + \cdots + \mu_r \lambda_r v_r \end{aligned}$$

But notice that if we multiply both sides of

$$v_{r+1} = \mu_1 v_1 + \cdots + \mu_r v_r$$

by λ_{r+1} we have

$$\lambda_{r+1} v_{r+1} = \mu_1 \lambda_{r+1} v_1 + \cdots + \mu_r \lambda_{r+1} v_r.$$

Thus

$$\mu_1 \lambda_1 v_1 + \cdots + \mu_r \lambda_r v_r = \mu_1 \lambda_{r+1} v_1 + \cdots + \mu_r \lambda_{r+1} v_r.$$

Subtracting the right-hand side from the left-hand side gives

$$\mu_1(\lambda_1 - \lambda_{r+1})v_1 + \cdots + \mu_r(\lambda_r - \lambda_{r+1})v_r = 0.$$

But this is a contradiction since $\{v_1, ..., v_r\}$ is a linearly independent set. □

10.4 Application: The PageRank Algorithm

Our initial motivation for finding eigenvectors and eigenvalues was to try to find a basis with respect to which a linear transformation becomes diagonal. The first time you learn linear algebra, however, it's probably difficult to appreciate why this is something special that you might care about. In this section we will outline another interesting application of eigenvectors and eigenvalues. The PageRank algorithm is used by Google to determine the "importance" of a webpage, which determines where the page lands on the list Google presents you after searching.

To each webpage containing a certain phrase we will associate a number; the higher the number is, the more important the page is. In order to determine these numbers we will have to solve an equation of the form $Ax = x$. That is, we need to find an eigenvector with eigenvalue 1 for a particular matrix.

The problem is that if A was allowed to be any arbitrary matrix, there's no guarantee that 1 is an eigenvalue. However we can get around this issue if we make sure our matrix A satisfies the hypotheses of the Perron-Frobenius theorem.

The Perron-Frobenius Theorem

The Perron-Frobenius theorem says that if we have a square matrix A with the property that each entry a_{ij} is positive, then we know a good bit of information about the eigenvalues of A.

Theorem 10.5 (The Perron-Frobenius Theorem).
Suppose that A is a matrix with real, positive entries. Then the following properties hold:

(a) A has a real eigenvalue ρ with the property that for every other eigenvalue λ, whether λ is real or complex, $|\lambda| < \rho$. (The absolute value of a complex number is $|x + iy| = \sqrt{x^2 + y^2}$.) We call ρ the **dominant eigenvalue** *of A.*

(b) There exists an eigenvector v_0 associated to eigenvalue ρ that contains only positive entries.

(c) The eigenspace of ρ has dimension 1, so all other eigenvectors with eigenvalue ρ are scalar multiples of the v_0 from part (b).

(d) The eigenvectors of every other eigenvalue of A do not have all positive entries: the only eigenvectors of A with positive entries are scalar multiples of the v_0 from part (b).

The proof of the Perron-Frobenius theorem involves some more advanced mathematics, so we won't give it here, but we will make use of the theorem. One very important consequence of the Perron-Frobenius theorem is the existence of a solution to the equation $Ax = x$ if A has the property that in addition to having only positive entries, the entries in each column of A sum up to 1. Matrices with this property are called ***stochastic matrices*** or sometimes ***Markov matrices***.

Example 10.6.
The following matrices are stochastic:

$$\begin{pmatrix} 1/2 & 1/8 & 1/2 \\ 1/4 & 3/8 & 3/16 \\ 1/4 & 1/2 & 5/16 \end{pmatrix} \qquad \begin{pmatrix} 0.2 & 0.13 \\ 0.8 & 0.87 \end{pmatrix}$$

Remark.
The word *stochastic* is basically a synonym for the word *random*. Matrices with positive (or more generally non-negative) entries that add up to 1 come up in probability theory and statistics when study-

ing certain random phenomena, such as a coin toss or a roll of dice, which is why these matrices are called stochastic.

Theorem 10.6.
If A is a positive, stochastic matrix (i.e., a square matrix with positive entries where the entries in each column sum up to 1), then $\rho = 1$ is the dominant eigenvalue of A guaranteed to exist by the Perron-Frobenius theorem.

Just to reiterate, the Perron-Frobenius theorem and Theorem 10.6 promise us the following: If A is a square matrix where every entry of A is positive and the sums of the entries in each column add up to 1, then 1 is the dominant eigenvalue of the matrix. Thus, by the Perron-Frobenius theorem, the only eigenvectors with all positive entries are scalar multiples of an eigenvector with eigenvalue 1. The existence of such an eigenvector is essential for the PageRank algorithm used by Google to determine the "importance" of webpages.

PageRank

When you search Google for a particular phrase, such as *algebra*, Google returns a list of web pages containing that phrase. As you know from experience, this list of web pages is typically very long. For example, searching for *algebra* brought back 81,900,000 results. You also know from experience that, usually, the most helpful pages show up near the top of the list, and things on later pages are typically not what you're looking for. How is it that Google is able to determine out of all of the web pages on the Internet, which ones are about algebra? And of those concerning algebra, how does Google determine which ones are most important and should go to the top of the list?

To determine which webpages contain *algebra*, or any other phrase, Google uses a program called a ***web crawler*** (or ***spider***) which basically starts from one web page (or collection of webpages) and makes note of the words that appear on that page, and also the links that page has to other pages. The web crawler then follows each link on the web page, and starts the process over again: noting what text is on the webpage, and what other pages the page links to, then follows those links and repeat the process.

In this way Google is able to build a large database of web pages and what content is on them. This is relatively easy to do. (If you like to write code, writing a simple web crawler is a fun afternoon project.)

The more interesting thing is how Google ranks the web pages to present you a list with the most relevant pages at the top. Google does this using an algorithm called ***PageRank***, which is technically intellectual property of Stanford University, where Google founders Larry Page and Sergey Brin were graduate students before they started Google.

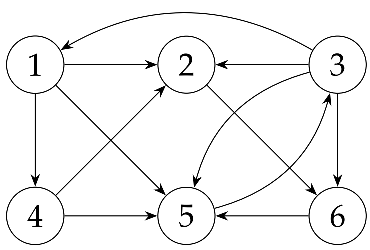
PageRank basically works by solving an eigenvector/eigenvalue problem to determine which web pages are important. The idea is that all of the webpages that contain a given phrase, like *algebra*, will correspond to entries in one enormous vector. The entries in the vector will be real numbers that describe how important the corresponding page is.

For example, just for the sake of simplicity let's suppose there were only six webpages that mentioned *algebra*. Let's call these page 1, page 2, ..., page 6. We will build a vector with entries $(x_1, x_2, ..., x_6)$ where x_i tells us how important page i is.

Now we will build a matrix A according to the following rules: if page j does not have a link to page i, then $a_{ij} = 0$. If page j links to k different pages, then $a_{ij} = {}^1\!/\!_k$ for each page i that page i links to. Multiple links to the same page are ignored; links from a page to itself are also ignored.

The idea is that each webpage has one "vote" that it can distribute to the other pages it links to. So if page 2 only links to page 6, then a_{62} = 1. If page 3 links to pages 1, 2, 5, and 6, then $a_{13} = a_{23} = a_{53} = a_{63} = 1/4$.

We can represent this information describing how pages link to one another with a ***graph***, which is a collection of ***nodes*** (also called ***vertices***) representing our web pages, and ***edges***, arrows from one node to another if there's a link between the corresponding pages. Let's say that our six pages are related to one another as indicated below:



We would then associate the following matrix:

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

Now, we want to consider the vector

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

where x_i tells us how important page i is. This notion of "importance" should follow the following two rules: the more pages that link to you, the more important you are; the more important the pages linking to you are, the more important you are. Figuring out what the x_i's should be then seems a bit tricky because we have to know how important all the other pages are before we can figure out how important a particular page is. So how can we get started figuring out which pages are important?

Let's first make an observation about what we get when we multiply A and x:

$$Ax = \begin{pmatrix} 0 & 0 & 1/4 & 0 & 0 & 0 \\ 1/3 & 0 & 1/4 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/4 & 1/2 & 0 & 1 \\ 0 & 1 & 1/4 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} x_3/4 \\ x_1/3 + x_3/4 + x_4/2 \\ x_5 \\ x_1/3 \\ x_1/3 + x_3/4 + x_4/2 + x_6 \\ x_2 + x_3/4 \end{pmatrix}$$

Notice that the entries of Ax are what we *want* the x_i to be. That is, we want to find the x_i's that solve the equation $Ax = x$, so we have an eigenvector/eigenvalue problem.

A few obvious issues crop up at this point.

1. It may be that $Ax = x$ doesn't have a solution, so how can be sure 1 is an eigenvalue of A?

2. Even if 1 is an eigenvalue of A, what happens if some of the entries of x_i are negative and some are positive? What would it mean if one webpage can have a positive importance and another has a negative importance?

3. Assuming that we can find eigenvectors with only positive entries, what happens if there are multiple, linearly independent such eigenvectors? Each one would correspond to a different notion of which webpages are important. Which one is the "right" one to use?

We can sidestep exactly these issues if we can apply the Perron-Frobenius theorem. The Perron-Frobenius theorem requires that the entries in our matrix are all positive, however, and our matrix is generally going to have lots of zeros in it. If we want to use the Perron-Frobenius theorem, then, we will have to modify our matrix a little bit to get rid of those zeros.

We will modify A in three steps:

1. If there is a column of A that consists entirely of zeros, then replace each entry in that column with $1/n$, if A is an $n \times n$ matrix.

   Having a column of all zeros means that we have a page that doesn't link to any other page. Intuitively we can think of this as meaning that page doesn't think any other pages are important, so it evenly distributes its vote to every page because no page is considered more important than another.

2. We need to get rid of any zeros in a column that isn't all zeros. We will do this by adding a matrix of all positive numbers to our matrix. We'll let this matrix be denoted B, and B will be the $n \times n$ matrix where every entry is $1/n$.

3. At this point if we just replaced A with $A + B$ (after replacing columns of zeros in A as described in the first step), we would have a matrix with all positive entries, but the entries in each column might not add up to 1 any more. To fix this we will actually replace A with $(1 - t)A + tB$ where $0 < t < 1$. This will give us a matrix where we can apply the Perron-Frobenius theorem.

The value of t is slightly arbitrary in the sense that we will get a matrix satisfying the hypotheses of the Perron-Frobenius theorem regardless of what t is (provided $0 < t < 1$). We might get different rankings for our pages if we choose different values of t, however, so we could try lots of different values of t to try to figure out which value of t gave us rankings we liked. Supposedly Google uses the value $t = 0.15$, so that's what we'll use in these notes as well.

Continuing with our example above, our original matrix A will be replaced by the matrix $0.85A + 0.15B$, where B is the 6×6 matrix consisting entirely of entries $1/6$. This gives us the matrix

$$\begin{pmatrix} 1/40 & 1/40 & 19/80 & 1/40 & 1/40 & 1/40 \\ 37/120 & 1/40 & 19/80 & 9/20 & 1/40 & 1/40 \\ 1/40 & 1/40 & 1/40 & 1/40 & 7/8 & 1/40 \\ 37/120 & 1/40 & 1/40 & 1/40 & 1/40 & 1/40 \\ 37/120 & 1/40 & 19/80 & 9/20 & 1/40 & 7/8 \\ 1/40 & 7/8 & 19/80 & 1/40 & 1/40 & 1/40 \end{pmatrix}$$

Now that we have a stochastic matrix, we can calculate the eigenvector associated to the dominant eigenvalue 1. In our particular example this is (approximately) the following:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 0.0820507243340146 \\ 0.125803704283873 \\ 0.268473996865951 \\ 0.0482477052279708 \\ 0.286439996312884 \\ 0.188983872975307 \end{pmatrix}$$

The numbers here are ugly, but we don't actually care about what the particular numbers are: we only care about the ordering they tell us. The larger x_i is, the more important page i is. Here we see that pages, ranked from most important to least important according to the vector above, is

1. Page 5
2. Page 3
3. Page 6
4. Page 2
5. Page 1
6. Page 4

So if these six pages were the webpages mentioning the word *algebra*, then this ranking would be used by Google to determine that Page 5 was the most imporant, while Page 4 was the least important.

10.5 Practice Problems

Problem 10.1.
Compute the eigenvalues and eigenspaces of the following matrices.

(a)
$$A = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$$

(b)
$$B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

(c)
$$C = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 2 \end{pmatrix}$$

(d)
$$D = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}$$

Problem 10.2.
Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} 2x - y \\ 3x + 2y + 7z \\ 6x + y + 8z \end{pmatrix}$$

Find a basis $\mathcal{B}$ of $\mathbb{R}^3$ for which T is represented by a diagonal matrix, and write down the matrix $A_{\mathcal{B}}$.

11

The Characteristic Polynomial

In this lecture we describe the characteristic polynomial which is a polynomial we can associate to a matrix whose roots are precisely the eigenvalues of the matrix.

11.1 Algebraic Background

Given any polynomial, a ***root*** of the polynomial is a value which makes the polynomial equal to zero. For example, the roots of $x^3 + 2x^2 - x - 2$ are $x = -2$, $x = 1$ and $x = -1$. When a polynomial has a root, we can always factor the polynomial into smaller pieces:

Theorem 11.1.
The polynomial $f(x)$ has a root of $x = a$ (i.e., $f(a) = 0$) if and only if $f(x)$ factors as $f(x) = (x - a) \cdot g(x)$ for some polynomial $g(x)$.

Proof.
First suppose that a is a root of f. By polynomial long division, we can always divide $x - a$ into $f(x)$ to write $f(x) = (x - a)g(x) + r(x)$ where $f(x)$ is a polynomial of lesser degree than $x - a$. Since $x - a$ has degree 1 and the degree of $r(x)$ has to have degree less than 1, it must be that $r(x)$ is in fact a constant: say $r(x) = r$. Thus $f(x) = (x - a)g(x) + r$. By assumption, however, $f(a) = 0$, but plugging a into the above we have $f(a) = (a - a)g(a) + r = r$, thus $r = 0$.

If $x - a$ is a factor of $f(x)$, then clearly $x = a$ is a root: $f(a) = (a - a)g(a) = 0$. □

So each root of a polynomial contributes a linear factor. We say that a polynomial ***factors completely*** if it can be written as a product of linear factors.

Given a polynomial with real coefficients, you know that sometimes the polynomial will factor completely and sometimes it won't: sometimes there are ***irreducible factors***. For example, the polynomial

$$x^3 + 2x^2 - x - 2$$

factors as $(x+2)(x-1)(x+1)$, but if we change the signs slightly to obtain

$$x^3 + 2x^2 + x + 2$$

then the best we can do is to factor this as $(x + 2)(x^2 + 1)$. Here $x^2 + 1$ is an irreducible factor: we can't break the polynomial down any further.

It turns out that *we can* always factor polynomials completely if we are okay with having complex roots. For example, $x^2 + 1$ can factor as $(x-i)(x+i)$, and so x^3+2x^2+x+2 factors completely as $(x+2)(x-i)(x+i)$.

The fact that this always happens is a significant result called the ***fundamental theorem of algebra***. The proof of the fundamental theorem of algebra requires more advanced mathematical techniques than what we've seen in class, so we won't give a proof.

Theorem 11.2 (The Fundamental Theorem of Algebra)**.**
Every polynomial with complex coefficients factors completely if complex roots are allowed.

Notice that since each root gives us a linear factor of the polynomial, each root contributes to the degree of the polynomial. Thus a polynomial of degree n can not have any more than n distinct roots.

It may happen that when we factor a polynomial completely some of the roots are "repeated." For example, $x^2 - 6x + 9$ factors as $(x - 3)^2$. If we group all of the repeated factors together so that the polynomial is written as

$$(x - a_1)^{m_1} \cdot (x - a_2)^{m_2} \cdots (x - a_n)^{m_n}$$

where each a_j is distinct, then we call m_j the ***multiplicity*** of the root a_j.

Each root contributes its multiplicity to the degree of the polynomial, so in fact every polynomial of degree n has exactly n roots if we count by a root's multiplicity.

11.2 The Characteristic Polynomial

In the last lecture we saw that the eigenvalues of a matrix A were precisely the values of λ for which $\det(A - \lambda I) = 0$. This expression, $\det(A - \lambda I)$, is a polynomial of degree n in λ.

Theorem 11.3.
If A is an $n \times n$ matrix, then $\det(A - \lambda I)$ is a polynomial of degree n.

(Sketch of proof).
We can perform an LUP factorization of A so that the diagonal entries of $A - \lambda I$ have the form $a_{jj} - \lambda + c_j$ (the c_j's appear when we try to zero out below the diagonal), and L and P are matrices just of numbers (no λ's). Hence the determinant of U will be

$$(a_{11} - \lambda + c_1) \cdot (a_{22} - \lambda + c_2) \cdots (a_{nn} - \lambda + c_n)$$

which is clearly a polynomial in λ of degree n. The determinants of L and P are just numbers, so they don't change the fact that our determinant of $A - \lambda I$ is a polynomial. □

The polynomial $\det(A - \lambda I)$ is called the ***characteristic polynomial*** of A. We will denote the characteristic polynomial by $p_A(\lambda)$, or sometimes $p_A(x)$ if we want to use x as the variable instead of λ – this is equivalent to $\det(A - xI)$.

Notice that the eigenvalues of a matrix A are precisely the roots of the characteristic polynomial $p_A(x)$. The multiplicity of a root is called the ***algebraic multiplicity*** of the eigenvalue. The ***geometric multiplicity*** of an eigenvalue is the dimension of the corresponding eigenspace.

Since $\det(A - \lambda I)$ is a polynomial of degree n, we see that every $n \times n$ square matrix has at most n distinct eigenvalues, and has exactly n eigenvalues (not necessarily distinct) if we count by (algebraic) multiplicity.

Example 11.1.
Find the characteristic polynomial, and the algebraic and geometric multiplicities of each eigenvalue of the matrix

$$\begin{pmatrix} 1 & -2 & -2 \\ 2 & 6 & 3 \\ -2 & -3 & 0 \end{pmatrix}.$$

The characteristic polynomial is

$$\begin{aligned} p_A(x) =& \det(A - xI) \\ =& \det \begin{pmatrix} 1-x & -2 & -2 \\ 2 & 6-x & 3 \\ -2 & -3 & -x \end{pmatrix} \\ =&(1-x) \cdot \det \begin{pmatrix} 6-x & 3 \\ -3 & -x \end{pmatrix} - (-2) \cdot \det \begin{pmatrix} 2 & 3 \\ -2 & -x \end{pmatrix} + (-2) \cdot \det \begin{pmatrix} 2 & 6-x \\ -2 & -3 \end{pmatrix} \\ =&(1-x) \cdot ((6-x)(-x)+9) + 2 \cdot (-2x+6) - 2 \cdot (-6 + 2(6-x)) \\ =&(1-x) \cdot (x^2 - 6x + 9) + 2 \cdot (6-x) - 2 \cdot (6-x) \\ =&(1-x) \cdot (x-3)^2 \end{aligned}$$

Thus the eigenvalues are 1 and 3, where 1 has algebraic multiplicity 1, and 3 has algebraic multiplicity 2.

To get the geometric multiplicities we need to find the dimensions of the corresponding eigenspaces.

For 1, the eigenspace is the set of solutions to $(A - I)v = 0$:

$$\begin{pmatrix} 0 & -2 & -2 \\ 2 & 5 & 3 \\ -2 & -3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The coefficient matrix in RREF is

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

so the eigenspace is

$$\left\{ \begin{pmatrix} z \\ -z \\ z \end{pmatrix} \,\middle|\, z \in \mathbb{R} \right\}.$$

This is a one-dimensional subspace of $\mathbb{R}^3$, so the geometric multiplicity of the eigenvalue 1 is 1.

For the eigenvalue 3, to find the eigenspace we need to find the set of solutions to $(A - 3I)v = 0$:

$$\begin{pmatrix} -2 & -2 & -2 \\ 2 & 3 & 3 \\ -2 & -3 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The RREF of the matrix is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

so the eigenspace is

$$\left\{ \begin{pmatrix} 0 \\ -z \\ z \end{pmatrix} \,\middle|\, z \in \mathbb{R} \right\},$$

and the geometric multiplicity of 3 is 1.

Notice in the previous example that the eigenvalue 3 had a larger algebraic multiplicity than its geometric multiplicity. This is true in general:

Theorem 11.4.
The geometric multiplicity of an eigenvalue is less than or equal to the eigenvalue's algebraic multiplicity.

Proof.
Suppose A is an $n \times n$ matrix and λ is an eigenvalue of A with ge-

ometric multiplicity m. Suppose that $\{v_1, v_2, ..., v_m\}$ form a basis for the eigenspace of λ. Now consider the basis $\mathcal{B}$ for $\mathbb{R}^n$ formed by taking the v_1, ..., v_m vectors, plus some other vectors w_{m+1}, w_{m+2}, ..., w_n. With respect to this basis we can represent A as

$$A_{\mathcal{B}} = \begin{pmatrix} \lambda & 0 & 0 & 0 & \cdots & 0 & 0 & \alpha_{1,m+1} & \cdots & \alpha_{1,n} \\ 0 & \lambda & 0 & 0 & \cdots & 0 & 0 & \alpha_{2,m+1} & \cdots & \alpha_{2,n} \\ 0 & 0 & \lambda & 0 & \cdots & 0 & 0 & \alpha_{3,m+1} & \cdots & \alpha_{3,n} \\ & & & & \ddots & & & & & \\ 0 & 0 & 0 & 0 & \cdots & \lambda & 0 & \alpha_{m-1,m+1} & \cdots & \alpha_{m-1,n} \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda & \alpha_{m,m+1} & \cdots & \alpha_{m,n} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \alpha_{m+1,m+1} & \cdots & \alpha_{m+1,n} \\ & & & & & \vdots & & & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \alpha_{n-1,m+1} & \cdots & \alpha_{n-1,n} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \alpha_{n,m+1} & \cdots & \alpha_{n,n} \end{pmatrix}$$

We can think of this as a upper block triangular matrix,

$$A_{\mathcal{B}} = \begin{pmatrix} \lambda I & C_1 \\ 0 & C_2 \end{pmatrix}$$

The characteristic polynomial of this matrix is then

$$\begin{aligned} \det(A_{\mathcal{B}} - xI) &= \det(\lambda I - xI) \cdot \det(C_2 - xI) \\ &= (\lambda - x)^k \cdot p_{C_2}(x). \end{aligned}$$

The algebraic multiplicity of λ is thus *at least* k, but it may be bigger if $p_{C_2}(x)$ has λ as a root. □

Remark.
You may be wondering how we know that $A_{\mathcal{B}}$ and A in the proof above have the same characteristic polynomial. This will be explained in the next lecture.

Example 11.2.

(a) Find the characteristic polynomial of

$$A = \begin{pmatrix} 4 & 10 & 8 \\ -2 & -4 & -3 \\ 0 & 0 & 3 \end{pmatrix}.$$

The characteristic polynomial $p_A(x)$ is the determinant $\det(A - xI)$, so we need to take the determinant of the matrix

$$A - xI = \begin{pmatrix} 4-x & 10 & 8 \\ -2 & -4-x & -3 \\ 0 & 0 & 3-x \end{pmatrix}$$

We can then take the determinant using cofactor expansion. Since the bottom row is mostly zeros, we will do the expansion along this row.

$$\begin{aligned} p_A(x) =& \det(A - xI) \\ =&(3-x) \cdot \det \begin{pmatrix} 4-x & 10 \\ -2 & -4-x \end{pmatrix} \\ =&(3-x) \cdot ((4-x)(-4-x) + 20) \\ =&(3-x) \cdot \left(-16 - 4x + 4x + x^2 + 20\right) \\ =&(3-x) \cdot \left(x^2 + 4\right) \\ =& - x^3 + 3x^2 - 4x + 12 \end{aligned}$$

(b) Find the algebraic multiplicity of each eigenvalue of A.

The eigenvalues of A are the roots of the characteristic polynomial, $p_A(x)$. From the above we see that our characteristic polynomial,

$$p_A(x) = -x^3 + 3x^2 - 4x + 12,$$

factors as

$$p_A(x) = (x^2 + 4)(3 - x).$$

Notice that $x^2 + 4$ factors as $(x - 2i)(x + 2i)$, thus

$$p_A(x) = (x - 2i)(x + 2i)(3 - x)$$

so the roots of $p_A(x)$ – aka the eigenvalues of A – are $2i$, $-2i$, and 3. Each of these has algebraic multiplicity 1.

To find the geometric multiplicities we need to find the dimension of each eigenspace.

(c) Find the geometric multiplicity of the eigenvalue $2i$.

For the eigenvector $2i$, the eigenspace is the set of solutions to

$$(A - 2iI)v = 0.$$

Thus we need to solve the system

$$\begin{pmatrix} 4-2i & 10 & 8 \\ -2 & -4-2i & -3 \\ 0 & 0 & 3-2i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Even though the entries of this matrix are complex numbers, we still solve this system like we would if the matrix had real entries: we put the matrix into RREF.

Let's first swap the first and second rows:

$$\begin{pmatrix} -2 & -4-2i & -3 \\ 4-2i & 10 & 8 \\ 0 & 0 & 3-2i \end{pmatrix}$$

Now multiply the first row by $-1/2$:

$$\begin{pmatrix} 1 & 2+i & -3/2 \\ 4-2i & 10 & 8 \\ 0 & 0 & 3-2i \end{pmatrix}$$

Now subtract $4 - 2i$ times the first row from the second. Notice that $(4 - 2i) \cdot (2 + i) = 8 + 4i - 4i - 2i^2 = 10$.

$$\begin{pmatrix} 1 & 2+i & -3/2 \\ 0 & 0 & 14-3i \\ 0 & 0 & 3-2i \end{pmatrix}$$

Now divide the second row by $14 - 3i$:

$$\begin{pmatrix} 1 & 2+i & -3/2 \\ 0 & 0 & 1 \\ 0 & 0 & 3-2i \end{pmatrix}$$

And we can then zero out the entries in the third column of the first and third rows:

$$\begin{pmatrix} 1 & 2+i & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

This is our matrix in RREF. So our original system is equivalent to the system

$$\begin{pmatrix} 1 & 2+i & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

which means

$$\begin{aligned} x =& -(2+i)y \\ z =& 0 \end{aligned}$$

so the eigenspace associated with the eigenvalue $2i$ is

$$\left\{ \begin{pmatrix} -(2+i)y \\ y \\ 0 \end{pmatrix} \middle| \, y \in \mathbb{C} \right\}.$$

Notice the free variable y above we allowed to be in $\mathbb{C}$ instead of just $\mathbb{R}$. Here our eigenvectors necessarily have complex values, so we consider our vectors as living inside of $\mathbb{C}^3$ instead of $\mathbb{R}^3$. This means our eigenspace is one-dimensional complex subspace; the geometric multiplicity is 1.

Exercise 11.1.
Find the geometric multiplicities of the eigenvalues $-2i$ and 3 of the matrix in the example above.

Complex conjugates

We say that two complex numbers are ***conjugates*** if they have the same real part, but their imaginary parts are the negatives of one another. For example, $3+2i$ and $3-2i$ are complex conjugates. Given any complex number $z = x+iy$, its complex conjugate is denoted by writing a bar over the number: $\overline{z} = \overline{x+iy} = x-iy$.

Notice that the conjugate of a real number is the same number: if $z = x+0i$, then

$$\overline{z} = \overline{x} = \overline{x+0i} = x - 0i = x = z.$$

Lemma 11.5.
The map $T : \mathbb{C} \to \mathbb{C}$ given by $T(z) = \overline{z}$ is a real *linear map (but it* is not *complex linear).*

Proof.
Let $z = x+iy$ and $z' = x'+iy'$ be any complex numbers, and let λ

be any real number. Then

$$
\begin{aligned}
T(\lambda\, z) =& T(\lambda(x+iy)) \\
=& T(\lambda x + i\lambda y) \\
=& \overline{\lambda x + i\lambda y} \\
=& \lambda x - i\lambda y \\
=& \lambda(x - iy) \\
=& \lambda\overline{x + iy} \\
=& \lambda T(z)
\end{aligned}
$$

$$
\begin{aligned}
T(z+z') =& T(x+iy+x'+iy') \\
=& T((x+x')+i(y+y')) \\
=& \overline{(}(x+x')+i(y+y')) \\
=& (x+x')-i(y+y') \\
=& x - iy + x' - iy' \\
=& \overline{x+iy} + \overline{x'+iy'} \\
=& T(z) + T(z')
\end{aligned}
$$

□

Exercise 11.2.
Convince yourself that complex conjugation is not complex linear by considering any example of $\overline{\lambda z}$ where λ and z are both complex numbers with non-zero imaginary part.

Exercise 11.3.
Show that if z and z' are complex numbers, then $\overline{z} \cdot \overline{z'} = \overline{z \cdot z'}$.

Theorem 11.6.
If $p(x)$ is a polynomial with real coefficients and z is a complex root of $p(x)$, then its conjugate $\overline{z}$ is also a complex root of $p(x)$.

Proof.
Write

$$p(x) = \sum_{k=0}^{n} a_i x^k.$$

Now suppose that $p(z) = 0$. We need to show that $p(\overline{z}) = 0$ as well, but this is simple to check:

$$\begin{aligned}
p(\overline{z}) &= \sum_{k=0}^{n} a_i \overline{z}^k \\
&= \sum_{k=0}^{n} a_i \overline{z^k} \quad \text{(by the exercise above)} \\
&= \sum_{k=0}^{n} \overline{a_i z^k} \quad \text{(by linearity)} \\
&= \overline{\sum_{k=0}^{n} a_k z^k} \quad \text{(by linearity)} \\
&= \overline{p(z)} \\
&= \overline{0} \\
&= 0
\end{aligned}$$

□

So the complex roots of a real polynomial come in pairs of complex conjugates. Applying this to characteristic polynomials we have the following:

Corollary 11.7.
The complex eigenvalues of a matrix with real entries come in complex conjugate pairs.

This *is not* true for matrices with complex entries, however. For example,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -2+3i \end{pmatrix}$$

has eigenvalues 1, i, and $-2+3i$, but *does not* have the complex conjugates $-i$ and $-2-3i$ as eigenvalues.

11.3 Eigenvalues Without Determinants

We end this lecture by mentioning an alternative technique for computing eigenvectors and eigenvalues without using determinants. There is some practical computational reason for wanting to do this: determinants are difficult to calculate. By hand it's practically impossible to calculate the determinant of a relatively small matrix, like say 10×10: in principle we can do it, but in practice it's going to be an unrealistic amount of work. If we can instead turn this question of finding eigenvectors and eigenvalues into a question of solving a system of equations without first calculating any determinants, then the problem becomes a bit more tractable.

In order to do this, however, we need to make a quick detour to talk about polynomials of matrices.

Polynomials of Matrices

Recall that matrices can be mulitiplied by scalars, added together, and can be multiplied – at least if the sizes of the matrices match up. In particular, two $n \times n$ matrices can always be multiplied together. We can thus define what it means to take powers of matrices:

$$A^k = \underbrace{A \cdot A \cdot A \cdots A}_{k \text{ copies of } A}$$

Notice that these three operations: multiplying by a number, raising to a power, and adding, are exactly the operations performed by polynomials. Thus it *almost* makes sense to evaluate a polynomial at a matrix. For example, if A is a square matrix then an expression like

$$A^3 - 3A^2 + A$$

makes sense: this is something we can compute. This is like evaluating the polynomial $p(x) = x^3 - 3x^2 + x$ at the matrix A. The one thing that might be a cause for concern is what should happen if we add a constant at the end of the polynomial: e.g., $q(x) = x^3 - 3x^2 + x + 2$. How should we add 2 to our matrix? The way around this is to think about the constant factor as being multiplied by x^0,

$$q(x) = x^3 - 3x^2 + x^1 + 2x^0.$$

So we should figure out how to interpret A^0. If we want the rule $A^j \cdot A^k = A^{j+k}$ to hold, then this only leaves one option for what A^0 could be: A^0 is the identity matrix I. So when we see a constant added in our polynomial, we should interpret this as the constant times the identity matrix.

Example 11.3.
Let $p(x)$ be the polynomial

$$p(x) = 3x^2 + 6x - 4$$

and let A be the matrix

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

Then $p(A)$ is the matrix

$$
\begin{aligned}
&p(A)\\
=&3A^2+6A-4I\\
=&3\begin{pmatrix}1&-1&3\\0&2&1\\2&2&0\end{pmatrix}^2+6\begin{pmatrix}1&-1&3\\0&2&1\\2&2&0\end{pmatrix}-4\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\\
=&\begin{pmatrix}21&9&6\\6&18&6\\6&6&24\end{pmatrix}+\begin{pmatrix}6&-6&18\\0&12&6\\12&12&0\end{pmatrix}-\begin{pmatrix}4&0&0\\0&4&0\\0&0&4\end{pmatrix}\\
=&\begin{pmatrix}23&3&24\\6&26&12\\18&18&20\end{pmatrix}
\end{aligned}
$$

Eigenvalues without determinants

Now suppose that A is an $n \times n$ matrix and let v be any non-zero vector in $\mathbb{R}^n$. Notice that

$$\{v, Av, A^2v, A^3v, ..., A^nv\}$$

is a set of $n+1$ vectors. Since we are in n-dimensional space, this set must be linearly dependent. That is, there is some choice of scalars c_n, c_{n-1}, ..., c_1, c_0, not all of which are zero, so that

$$c_nA^nv + c_{n-1}A^{n-1}v + \cdots + c_1Av + c_0v = 0.$$

Factoring out the v's this becomes

$$\left(c_nA^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I\right) v = 0.$$

The expression

$$c_nA^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I$$

is just some $n \times n$ matrix: it's the polynomial

$$p(x) = c_nx^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0$$

evaluated at the matrix A. Now factor this polynomial completely – this might require using complex factors – to get something like

$$p(x) = (x - \lambda_1)^{m_1} \cdot (x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}.$$

So our system above,

$$\big(c_nA^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I\big)\, v = 0,$$

can be rewritten as

$$(A-\lambda_1 I)^{m_1} \cdot (A-\lambda_2 I)^{m_2} \cdots (A-\lambda_k I)^{m_k} v = 0$$

Now, since $v \neq 0$ but is getting mapped to zero, some factor $(A - \lambda_j I)$ above must not be injective.

Remark.
The last statement above might be easier to see if we think of $A-\lambda_j I$ as a linear transformation: say $T_j : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T_j(v) = (A-\lambda_j I)v$. Then the above says

$$T_1^{m_1} \circ T_2^{m_2} \circ \cdots \circ T_k^{m_k}(v) = 0.$$

Since $v \neq 0$, at least one of the T_j's must not be injective. It's not hard to show that the composition of injective maps is injective, so if all of the maps above are injective it would have to be the case that $v = 0$, but this is a contradiction.

If $A - \lambda_j I$ is not injective, then λ_j is an eigenvalue of A. (If you look back at our definition of eigenvectors and eigenvalues, this is basically the definition in disguise.)

Notice we're not saying that the v we started with is an eigenvector: v is any arbitrarily chosen, non-zero vector so we could very well have picked one that wasn't an eigenvector! But now we do have the eigenvalues, so we can proceed to find the eigenvectors as before.

Example 11.4.
Find the eigenvalues of the following matrix without calculating any determinants:

$$A = \begin{pmatrix} 1 & -2 & -1 \\ -1/4 & 3/2 & -1/4 \\ 1/2 & 1 & 5/2 \end{pmatrix}.$$

We can choose v to be any non-zero vector, so let's just choose

$$v = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Now we consider the vectors $\{v, Av, A^2v, A^3v\}$:

$$v = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$
$$Av = \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix}$$
$$A^2v = \begin{pmatrix} -8 \\ 1 \\ 10 \end{pmatrix}$$
$$A^3v = \begin{pmatrix} -20 \\ 1 \\ 22 \end{pmatrix}$$

Since we have four 3-dimensional vectors, they can't form a linearly independent set, so there must be some non-zero solution to the equation

$$c_3 \begin{pmatrix} -20 \\ 1 \\ 22 \end{pmatrix} + c_2 \begin{pmatrix} -8 \\ 1 \\ 10 \end{pmatrix} + c_1 \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix} + c_0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

This means we want to solve the system

$$\begin{pmatrix} -20 & -8 & -2 & 1 \\ 1 & 1 & 1 & 1 \\ 22 & 10 & 4 & 1 \end{pmatrix} \begin{pmatrix} c_3 \\ c_2 \\ c_1 \\ c_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Putting the 3×4 matrix above into RREF gives

$$\begin{pmatrix} 1 & 0 & -1/2 & -3/4 \\ 0 & 1 & 3/2 & 7/4 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This means

$$\begin{aligned} c_3 &= 1/2 c_1 + 3/4 c_0 \\ c_2 &= -3/2 c_1 - 7/4 c_0 \end{aligned}$$

If we take $c_1 = c_0 = 4$, then one non-zero solution of the system is

$$\begin{pmatrix} c_3 \\ c_2 \\ c_1 \\ c_0 \end{pmatrix} = \begin{pmatrix} 5 \\ -13 \\ 4 \\ 4 \end{pmatrix}.$$

So our equation above can be written as

$$5 \begin{pmatrix} -20 \\ 1 \\ 22 \end{pmatrix} - 13 \begin{pmatrix} -8 \\ 1 \\ 10 \end{pmatrix} + 4 \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix} + 4 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0$$

Keeping in mind these vectors are A^3v, A^2v, Av and v we have

$$5A^3v - 13A^2v + 4Av + 4v = 0$$

Factoring out the v this is

$$(5A^3 - 13A^2 + 4A + 4I)v = 0.$$

Thus we are considering the polynomial

$$5x^3 - 13x^2 + 4x + 4$$

evaluated at A. This polynomial factors as

$$(5x + 2)(x - 1)(x - 2)$$

So we can rewrite $(5A^3 - 13A^2 + 4A + 4I)v = 0$ as

$$(5A + 2I) \cdot (A - I) \cdot (A - 2I)v = 0$$

If we multiply everything by $1/5$, we can rewrite this equation as

$$\left(A + \frac{2}{5}I\right) \cdot (A - I) \cdot (A - 2I)v = 0$$

just to make the first factor look like the other two.

Now we have three *candidates* for eigenvalues: $-2/5$, 1, and 2. We aren't guaranteed that all of these are in fact eigenvalues. We will only have an eigenvalue if $(A - \lambda I)v$ has a non-zero solution. But now this is something we can check by solving the systems

$$(A + 2/5 I)v = 0 \qquad (A - I)v = 0 \qquad (A - 2I)v = 0.$$

Putting each of the matrices into RREF we have

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{pmatrix}$$

for $A + 2/5 I$, $A - I$, and $A - 2I$, respectively. From these RREF's we see that $(A + 2/5 I)v = 0$ has no non-zero solutions, while $(A - I)v = 0$ and $(A - 2I)v = 0$ both *do* have non-zero solutions. Hence the eigenvalues of our matrix are 1 and 2.

The above example seems like a lot of work, but the point is that for large matrices this procedure is actually *less* work than calculating determinants.

11.4 Practice Problems

Problem 11.1.
Compute the characteristic polynomial, as well as the algebraic and geometric multiplicities of each eigenvalue for the matrices below:

(a)

$$A = \begin{pmatrix} 4 & 1 & 1 \\ -8 & -2 & -4 \\ 2 & 1 & 3 \end{pmatrix}$$

(b)

$$B = \begin{pmatrix} 5 & 1 & 0 \\ -11 & -2 & -1 \\ 3 & 1 & 2 \end{pmatrix}$$

(c)

$$C = \begin{pmatrix} 5 & 0 & -1 \\ -3 & 4 & 3 \\ 1 & 0 & 3 \end{pmatrix}$$

(d)

$$D = \begin{pmatrix} 9 & 2 & 1 \\ -13 & -1 & -2 \\ 3 & 1 & 4 \end{pmatrix}$$

12

Similarity and Diagonalization

12.1 Introduction

We had seen before that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ can be represented as a matrix, but we can only do this *after* choosing bases for V and W and making a different choice of basis will give us a different matrix.

In this lecture we describe a relationship between different matrices that actually represent the same linear transformation, just with respect to different bases.

We had seen before in Lecture 15 that every change-of-basis matrix is invertible. We can in fact interpret *every* invertible matrix as some change of basis matrix: given any invertible $n \times n$ matrix A with columns a_1, a_2, ..., a_n, A is the change of basis matrix from the standard basis e_1, ..., e_n to the basis given by a_1, ..., a_n.

12.2 Similarity

We say that two matrices A and B are ***similar*** if there exists some invertible matrix P such that

$$B = PAP^{-1}.$$

Notice that this exactly means that A and B represent the same linear transformation, but with respect to different bases. If we think of P as the change of basis matrix from the standard basis to the basis given by the columns of P, then the formula $B = PAP^{-1}$ says that we get B by first applying P^{-1} (switching from the basis given by the columns of P to the standard basis), applying A, then switching back to the other basis.

One crucial property of similarity is that A and $B = PAP^{-1}$ have the same characteristic polynomial. This implies that A and B have the same eigenvalues with the same algebraic multiplicities.

Theorem 12.1.
If A and B are similar matrices, then they have the same characteristic polynomial.

Proof.
Suppose $B = PAP^{-1}$. Then the characteristic polynomial of B is

$$\begin{aligned}
p_B(x) &= \det(B - xI) \\
&= \det(PAP^{-1} - xI) \\
&= \det(PAP^{-1} - xPP^{-1}) \\
&= \det((PA - xP)P^{-1}) \\
&= \det((PA - P(xI))P^{-1}) \\
&= \det(P(A - xI)P^{-1}) \\
&= \det(P)\ \det(A - xI)\ \det(P^{-1}) \\
&= \det(P)\ \det(P^{-1})\ \det(A - xI) \\
&= \det(P)\ \frac{1}{\det(P)}\ \det(A - xI) \\
&= \det(A - xI) \\
&= p_A(x)
\end{aligned}$$

□

Since the characteristic polynomial of a matrix is preserved by similarity (i.e., changing basis), this means we can actually associate characteristic polynomials to linear transformations: pick any basis you want, determine the matrix representation of $T : V \to V$, and calculate the characteristic polynomial of that matrix. You will calculate the same polynomial regardless of what basis you choose, since changing bases just replaces your matrix with a similar matrix.

12.3 Diagonalization

We say that a matrix A is ***diagonalizable*** if it is similar to a diagonal matrix. That is, if we can find a basis for which the linear transformation $T(v) = Av$ can be represented by a diagonal matrix D.

Example 12.1.

The matrix

$$A = \begin{pmatrix} -2 & -5 & 4 \\ 2 & 5 & -2 \\ -1 & -1 & 3 \end{pmatrix}$$

is diagonalizable. If we take

$$P = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ 1 & 2 & -1 \end{pmatrix} \qquad P^{-1} = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

then it is easy to calculate that

$$PAP^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Example 12.2.
The matrix

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

is not diagonalizable. No matter what you take P to be, the matrix PAP^{-1} will never be diagonal.

After seeing Example and Exam , the natural question to ask is if we're given a matrix A how can we determine if it is or is not diagonalizable? If we figure out that the matrix *is* diagonalizable, how can we find the matrix P so that PAP^{-1} is diagonal? These questions are answered by the following theorem:

Theorem 12.2.
An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

Proof.
Suppose that A is diagonalizable. That is, there is some square matrix P so that $P^{-1}AP$ is a diagonal matrix D. Notice that we can write

$$D = P^{-1}AP \implies PD = AP.$$

If we suppose the columns of P are the vectors v_1, v_2, ..., v_n, then the product AP may be written as

$$AP = \begin{pmatrix} Av_1 & Av_2 & \cdots & Av_n \end{pmatrix}$$

While PD may be written as

$$PD = \begin{pmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \end{pmatrix}$$

where λ_1, ..., λ_n are the diagonal entries of D. Since $PD = AP$ this means these two matrices are the same, hence each column is the same, so

$$Av_1 = \lambda_1 v_1 \qquad Av_2 = \lambda_2 v_2 \qquad \cdots Av_n = \lambda_n v_n.$$

Now, since the v_i are the columns of the invertible matrix P, the v_i are linearly independent (and hence are all non-zero), so these equations say that the columns of the matrix P are n linearly independent eigenvectors of A, and also that the entries of the diagonal matrix D are exactly the eigenvalues of A.

For the converse, suppose now that A is a matrix with n linearly independent eigenvectors, v_1 through v_n. Let P be the matrix whose columns are these vectors. Notice that P must be invertible since is a square matrix with linearly independent columns. Let D be the diagonal matrix whose entries are the eigenvalues of A, given in the order (from upper left-hand corner to lower right-hand corner) λ_1, λ_2, ..., λ_n where λ_i is the eigenvalue of the eigenvector v_i in the i-th column of A. Just as above,

$$PD = \begin{pmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \end{pmatrix}$$

and

$$AP = \begin{pmatrix} Av_1 & Av_2 & \cdots & Av_n \end{pmatrix}.$$

Since $Av_i = \lambda_i v_i$, these matrices are equal: $PD = AP$. Multiplying both sides on the left by P^{-1} we have $D = P^{-1}AP$, and so A is diagonalizable. □

The proof of Theorem 12.2 tells us that when a matrix is diagonalizable, the corresponding diagonal matrix has the eigenvalues on the diagonal. (Although we could have figured this out already since similar matrices have the same characteristic polynomial.)

Example 12.3.
Is the matrix

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

diagonalizable? If so, find the diagonal matrix D and invertible matrix P so that $D = P^{-1}AP$.

We need to find the eigenvectors and eigenvalues of A, so we want to find the roots of the characteristic polynomial,

$$\begin{aligned}
p_A(x) =& \det(A - xI) \\
=& \det \begin{pmatrix} -x & -1 & -2 \\ 0 & 1-x & 0 \\ 1 & 1 & 3-x \end{pmatrix} \\
=& (-x) \cdot \det \begin{pmatrix} 1-x & 0 \\ 1 & 3-x \end{pmatrix} + 1 \cdot \det \begin{pmatrix} -1 & -2 \\ 1-x & 0 \end{pmatrix} \\
=& (-x) \cdot (1-x)(3-x) + 2(1-x) \\
=& (-x) \cdot (3 - 4x + x^2) + 2 - 2x \\
=& -x^3 + 4x^2 - 3x + 2 - 2x \\
=& -x^3 + 4x^2 - 5x + 2
\end{aligned}$$

This is our characteristic polynomial, but to find the roots (aka the eigenvalues of A) we need to factor this polynomial. To do this, let's notice that if we factor a polynomial as

$$(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_k),$$

then when we expand this expression by multiplying everything out, the constant term at the end will be the product of $(-\lambda_1)(-\lambda_2) \cdots (-\lambda_k)$. So our candidates for what the λ_k might be can be determined by the factors of the constant term.

Since our constant term here is 2, our candidates are 1 and 2.

Notice what happens when we plug in $x = 1$ into our polynomial above: we get $-1 + 4 - 5 + 2 = 0$, so $x - 1$ is in fact a factor of our polynomial. We *could* now use polynomial long division to figure out what the other factor is, but let's do something different. Since $x - 1$ is a factor of $-x^3 + 4x^2 - 5x + 2$, let's try to rewrite our polynomial by breaking up the existing terms into pairs we can factor an $x - 1$ out of:

$$\begin{aligned}
p_A(x) =& -x^3 + 4x^2 - 5x + 2 \\
=& -x^3 + x^2 + 3x^2 - 3x - 2x + 2 \\
=& -x^2(x-1) + 3x(x-1) - 2(x-1)
\end{aligned}$$

Now we can factor out these $x-1$ factors

$$\begin{aligned} p_A(x) =& -x^2(x-1) + 3x(x-1) - 2(x-1) \\ =& (-x^2+3x-2)(x-1) \\ =& -(x^2-3x+2)(x-1) \end{aligned}$$

Now we can factor x^2-3x+2 a little more easily: $x^2-3x+2 = (x-2)(x-1)$. Putting this altogether our characteristic polynomial factors as

$$p_A(x) = -(x-2)(x-1)^2.$$

Hence our eigenvalues are $x=2$ (with algebraic multiplicity 1) and $x=1$ (with algebraic multiplicity 2).

Now we need to find eigenvectors for each of these eigenvalues.

For the eigenvalue 2 we need to solve the system $(A-2I)v=0$:

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

The RREF of the coefficient matrix is

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

so the eigenspace for 2 is

$$\left\{ \begin{pmatrix} -z \\ 0 \\ z \end{pmatrix} \,\middle|\, z \in \mathbb{R} \right\}.$$

Notice this is one-dimensional, so the geometric multiplicity of 2 is 1.

For the eigenvalue 1 we solve the system $(A-I)v=0$:

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

The RREF of the coefficient matrix is

$$\begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The solutions to this system are the triples $\begin{pmatrix} x & y & z \end{pmatrix}^T$ where $x = -y - 2z$, so our eigenspace is

$$\left\{ \begin{pmatrix} -y \\ y \\ 0 \end{pmatrix} + \begin{pmatrix} -2z \\ 0 \\ z \end{pmatrix} \,\middle|\, y, z \in \mathbb{R} \right\}$$

So the geometric multiplicity of 1 is 2.

Now if we pick three linearly independent eigenvectors, one from the eigenspace of 2 and two from the eigenspace of 1, we can build our P matrix. Let's consider the eigenvectors

$$\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad and \quad \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}.$$

We use these as the columns of our P matrix:

$$P = \begin{pmatrix} -1 & -1 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

We can then row-reduce $\begin{pmatrix} P \mid I \end{pmatrix}$ to $\begin{pmatrix} I \mid P^{-1} \end{pmatrix}$ to find the inverse,

$$P^{-1} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \end{pmatrix}.$$

Now we simply compute

$$P^{-1}AP = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 0 & -1 & -2 \\ 0 & 1 & 0 \\ 1 & 1 & 3 \end{pmatrix} \begin{pmatrix} -1 & -1 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Example 12.4.
Is the matrix

$$A = \begin{pmatrix} 1 & -1 & 1 \\ -2 & 3 & 1 \\ -4 & -2 & 5 \end{pmatrix}$$

diagonalizable?

An $n \times n$ matrix is diagonalizable if and only if it has n linearly independent eigenvectors, so we need to find the eigenspaces of A.

The characteristic polynomial of A is

$$p_A(x) = \det(A - xI) = x^3 - 9x^2 + 27x - 27.$$

We can factor this polynomial by grouping by noting that the factors of 27 are 1, 3, and 9. Of these, only 3 will make the polynomial equal to zero, and so the polynomial factors as $p_A(x) = (x-3)^3$.

Now we find the eigenspace of 3 by solving the system $(A - 3I)v = 0$:

$$\begin{pmatrix} -2 & -1 & 1 \\ -2 & 0 & 1 \\ -4 & -2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The RREF of the matrix is

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus the eigenspace of 3 is

$$\left\{ \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} \middle| \, x \in \mathbb{R} \right\}.$$

Since this has dimension one, we can't find three linearly independent eigenvectors, and so the matrix is not diagonalizable.

Notice that Theorem 12.2 tells us that an $n \times n$ matrix will be diagonalizable if and only if the sum of the geometric multiplicities of the matrix's eigenvalues add up to n. This is guaranteed to happen if, for example, the matrix had n distinct eigenvalues.

Example 12.5.
The matrix

$$\begin{pmatrix} 1 & 7 & 3 & 2 & 1 \\ 0 & -1 & 2 & 2 & 2 \\ 0 & 0 & 3 & 4 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

is diagonalizable. Since this is a triangular matrix, we know the eigenvalues appear on the diagonal of the matrix. This 5×5 matrix has five distinct eigenvalues, $-1, 0, 1, 2, 3$, and so is diagonalizable.

12.4 Jordan Normal Form

Even if a matrix is not diagonalizable, it turns out that we can still find a similar matrix which is *almost* diagonal. We won't take the time to prove that such a matrix exists, though the proof isn't particularly difficult, but we will note that such a matrix can always be found. In order to describe this matrix, we will need to make one preliminary definition.

We say that a square $m \times m$ matrix is a ***Jordan block*** with eigenvalue λ if the matrix has the form

$$\begin{pmatrix} \lambda & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \lambda & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \lambda & 1 & \cdots & 0 \\ & & & & \ddots & & \\ 0 & 0 & \cdots & 0 & \lambda & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \lambda & 1 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \lambda \end{pmatrix}$$

That is, we have a matrix that has λ's on the diagonal, 1's immediately above the diagonal, and 0's everywhere else. Let's denote this $m \times m$ matrix $J_m(\lambda)$. Notice that $J_1(\lambda)$ is just a single λ.

We say a matrix is in ***Jordan normal form*** if the matrix is block diagonal which each block on the diagonal being a Jordan block:

$$\begin{pmatrix} J_{m_1}(\lambda_1) & & & & \\ & J_{m_2}(\lambda_2) & & & \\ & & J_{m_3}(\lambda_3) & & \\ & & & \ddots & \\ & & & & J_{m_d}(\lambda_d) \end{pmatrix}$$

here we *do not* assume that the λ_j's are distinct.

Example 12.6.

The following matrices are in Jordan normal form:

$$\begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} J_2(2) & & \\ & J_1(2) & \\ & & J_1(3) \end{pmatrix}$$

$$\begin{pmatrix} 3 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} J_3(3) & \\ & J_2(3) \end{pmatrix}$$

A matrix in Jordan normal form is the next best thing to a diagonal matrix because it is very easy to read off information about the eigenvalues of the matrix based on the Jordan blocks.

Theorem 12.3.
If A is a matrix in Jordan normal form, then the eigenvalues of A occur on the diagonal of the matrix. Furthermore, the number of times an eigenvalue appears on the diagonal is precisely the algebraic multiplicity of the eigenvalue, and the number of Jordan blocks with a given eigenvalue is the geometric multiplicity of the eigenvalue.

Example 12.7.
Consider the following matrix in Jordan normal form:

$$\begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

This matrix has two eigenvalues: -1 and 2. The eigenvalue -1 has algebraic multiplicity 3 but geometric multiplicity 2; the eigenvalue 2 had algebraic multiplicity 4, but geometric multiplicity 1. From the eigenvalues and the algebraic multiplicities we know that the characteristic polynomial of this matrix is

$$p_A(x) = (x+1)^3\,(x-2)^4$$

Exercise 12.1.

(a) Show that every diagonal matrix is in Jordan normal form.

(b) Show that if A is a matrix which is in Jordan normal form but is not diagonal, then A is not diagonalizable.

The main theorem about Jordan normal form is that *every* square matrix has a Jordan normal form.

Theorem 12.4.
Every square matrix is similar to a matrix in Jordan normal form.

Example 12.8.
Find a Jordan normal form for the matrix

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

Notice that the eigenvalues of this matrix are 1 and 2, and the algebraic multiplicities are 1 and 2, respectively. The eigenspace associated to eigenvalue 1 is one-dimensional since the algebraic multiplicity is 1. The eigenspace associated to eigenvector 2 is

$$\left\{ \begin{pmatrix} 2y \\ y \\ 0 \end{pmatrix} \middle| y \in \mathbb{R} \right\}.$$

This is one-dimensional, so the geometric multiplicity is 1. We now have enough information to construct a Jordan normal form,

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

12.5 Practice Problems

Problem 12.1.
Determine if each of the matrices below is diagonalizable or not. If the matrix is diagonalizable, write down the diagonalization. If the matrix is not diagonalizable, write down a Jordan normal form for the matrix.

(a)

$$A = \begin{pmatrix} 3 & 2 & 4 & 1 \\ 0 & -1 & 3 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 6 \end{pmatrix}$$

(b)

$$B = \begin{pmatrix} -9 & -7 & -9 \\ -6 & -2 & -5 \\ 18 & 12 & 17 \end{pmatrix}$$

(c)

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

(d)

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

(e)

$$E = \begin{pmatrix} 2 & 1 & 2 \\ -1 & 2 & 0 \\ 1/2 & 0 & 2 \end{pmatrix}$$

Part III

Inner Products & Norms

13 The Inner Product in $\mathbb{R}^n$

13.1 The Inner Product

At the start of the semester we initially defined vectors as quantities that had both a magnitude and a direction, and we represented these quantities in two and three dimensions as arrows. We then extended our definition of vector so that we could talk about vectors in any dimension. When we did this, we lost a certain amount of geometric intuition. In the plane and three-space we could visualize things, measure angles, calculate lengths, etc. We will now say how to do these geometric operations in higher dimensions using a tool called an *inner product*. The inner product will allow us to define and compute geometric quantities in any (finite) dimension, even though we might not be able to visualize those quantities.

The ***(standard, or Euclidean) inner product*** in $\mathbb{R}^n$ is a way of pairing two vectors together to get a real number. We denote the number associated to the pair of vectors u and v as $\langle u, v\rangle$, and we calculate this number by the following formula:

$$\langle u, v\rangle = u^T v$$

where u^T is the transpose of u (this really just means writing the components of u as a row instead of a column), and then doing matrix multiplication with v. For example, if our vectors were in the following elements of $\mathbb{R}^4$,

$$u = \begin{pmatrix} 2 \\ 7 \\ 0 \\ 1 \end{pmatrix} \qquad v = \begin{pmatrix} 3 \\ -2 \\ -1 \\ 4 \end{pmatrix}$$

then we have

$$\langle u, v\rangle = \begin{pmatrix} 2\,7\,0\,1 \end{pmatrix} \begin{pmatrix} 3 \\ -2 \\ -1 \\ 4 \end{pmatrix} = -4.$$

In terms of components, if our vectors in $\mathbb{R}^n$ are

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

then this number $\langle u, v \rangle$ is also given by

$$\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i.$$

There are five crucial properties of the inner product we have defined:

Theorem 13.1.
For all $u, v, w \in \mathbb{R}^n$ *and all* $\lambda \in \mathbb{R}$*:*

1. $\langle u, v \rangle = \langle v, u \rangle$
2. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
3. $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$
4. $\langle u, u \rangle \geq 0$
5. $\langle u, u \rangle = 0$ *if and only if* $u = 0$.

Proving these properties hold is really just a matter of computation: compute the left- and right-hand sides of each of the properties claimed above, and verify that they are the same. Since this is entirely computation, we will leave this as an exercise.

Exercise 13.1.
Prove Theorem 13.1.

There are other types of inner products besides the "standard" one we have defined, and we will discuss these later, but for right now we will stick to the standard inner product.

Remark.
The standard inner product is usually called the *dot product* in a multivariable calculus course.

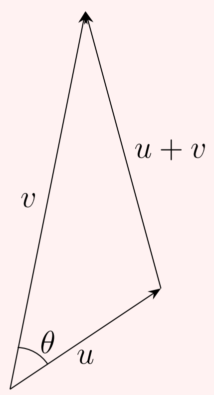
Exercise 13.2.
Show that for every vector $v \in \mathbb{R}^n$,

$$\langle 0, v \rangle = 0.$$

13.2 Orthogonality

In $\mathbb{R}^2$ and $\mathbb{R}^3$, we say two vectors u and v are ***orthogonal*** if they meet at a $90°$ angle. By the law of cosines, this means exactly that $u^T v = 0$.

Remark.
Suppose $u = \begin{pmatrix} u_1 & u_2 \end{pmatrix}^T$ and $v = \begin{pmatrix} v_1 & v_2 \end{pmatrix}^T$ are two vectors in $\mathbb{R}^2$ and place the vectors together at their tails. To measure the angle θ between the vectors we consider the triangle formed by u, v, and $u + v$, and then use the law of cosines.



Letting $\|u\|$ denote the length of u, and likewise for v and $u + v$, the law of cosines then says that

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

Notice however that

$$\|u\|^2 = \left(\sqrt{u_1^2 + u_2^2}\right)^2 = u_1^2 + u_2^2$$

which we can write as $\langle u, u\rangle$. The above expression can then be rewritten as

$$\langle u+v, u+v\rangle = \langle u, u\rangle + \langle v, v\rangle + 2\|u\|\|v\| \cos\theta.$$

As inner products are linear in each argument, the left-hand side of this equation may be rewritten as

$$\begin{aligned} \langle u+v, u+v\rangle &= \langle u, u+v\rangle + \langle v, u+v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \langle u, u\rangle + 2\langle u, v\rangle + \langle v, v\rangle \end{aligned}$$

Hence

$$\begin{aligned} \langle u, v\rangle &= \|u\|\|v\| \cos\theta \\ \implies \cos\theta &= \frac{\langle u, v\rangle}{\|u\|\,\|v\|} \end{aligned}$$

So when $\theta = 90^\circ$ (i.e., the vectors are orthogonal), we have $\cos\theta = 0$, so

$$\begin{aligned} \frac{\langle u, v\rangle}{\|u\|\,\|v\|} &= 0 \\ \implies \langle u, v\rangle &= 0. \end{aligned}$$

Since the standard inner product is just a generalization of the dot product in two and three dimensions, this tells us how we should define orthogonality in general: we say two vectors u and v are ***orthogonal*** if $\langle u, v\rangle = 0$.

Given a subspace U of an inner product space V, the set of vectors orthogonal to every vector in U, denoted $U^\perp$ and sometimes pronounced

"U-perp", is called the ***orthogonal complement*** of U:

$$U^\perp = \left\{ v \in \mathbb{R}^n \;\middle|\; \langle u, v \rangle = 0 \text{ for all } u \in U \right\}$$

Theorem 13.2.
If U is a subspace of $\mathbb{R}^n$, then the orthogonal complement, $U^\perp$, is also a subspace of $\mathbb{R}^n$.

Proof.
We need to show that $U^\perp$ is not empty, and is closed under both vector addition and scalar multiplication. Notice that $U^\perp$ is not empty as $0 \in U^\perp$ by Exercise 13.2 above.

Now suppose that $v, w \in U^\perp$. Then for any $u \in U$,

$$\langle v + w, u \rangle = \langle v, u \rangle + \langle w, u \rangle = 0 + 0 = 0,$$

so $v + w \in U^\perp$.

Let λ be any scalar and let $v \in U^\perp$. Then for any $U \in U$,

$$\langle \lambda v, w \rangle = \lambda \langle v, w \rangle = \lambda \cdot 0 = 0$$

so $U^\perp$ is also closed under scalar multiplication. □

Lemma 13.3.
$\dim(U^\perp) = n - \dim(U)$.

Proof.
Let $\{u_1, u_2, \ldots, u_m\}$ be a basis for U and consider the map $T : \mathbb{R}^n \to U$

given by

$$T(v) = \langle v, u_1 \rangle u_1 + \langle v, u_2 \rangle u_2 + \cdots + \langle v, u_m \rangle u_m.$$

We can suppose that in fact elements of $\{u_1, u_2, ..., u_m\}$ are pairwise orthogonal. (This is proven later when we show every inner product space has an orthogonal basis.)

First notice this is a linear transformation:

$$\begin{aligned}
T(v+w) &= \langle v+w, u_1 \rangle u_1 + \langle v+w, u_2 \rangle u_2 + \cdots + \langle v+w, u_m \rangle u_m \\
&= (\langle v, u_1 \rangle + \langle w, u_1 \rangle)u_1 + + \cdots (\langle v, u_m \rangle + \langle w, u_m \rangle)u_m \\
&= \langle v, u_1 \rangle u_1 + \langle w, u_1 \rangle u_1 + \cdots \langle v, u_m \rangle u_m + \langle w, u_m \rangle u_m \\
&= \langle v, u_1 \rangle u_1 + \cdots + \langle v, u_m \rangle u_m + \langle w, u_1 \rangle u_1 + \cdots + \langle w, u_m \rangle u_m \\
&= T(v) + T(w)
\end{aligned}$$

$$\begin{aligned}
T(\lambda v) &= \langle \lambda v, u_1 \rangle u_1 + \langle \lambda v, u_2 \rangle u_2 + \cdots + \langle \lambda u_m, u_m \rangle u_m \\
&= \lambda \langle v, u_1 \rangle u_1 + \lambda \langle v, u_2 \rangle u_2 + \cdots + \lambda \langle v, u_m \rangle \\
&= \lambda(\langle v, u_1 \rangle u_1 + \langle v, u_2 \rangle u_2 + \cdots + \langle v, u_m \rangle u_m) \\
&= \lambda T(v).
\end{aligned}$$

Notice this linear transformation is surjective: in particular, $T(u_i)$ is a scalar multiple of u_i as

$$T(u_i) = \langle u_i, u_1 \rangle u_1 + \langle u_i, u_2 \rangle u_2 + \cdots + \langle u_i, u_m \rangle u_m = \langle u_i, u_i \rangle u_i$$

since $\langle u_i, u_j \rangle = 0$ for $i \neq j$. So, $\mathrm{im}(T) = U$.

Notice too the kernel of this map is exactly $U^\perp$: if $T(v) = 0$ then

$$\langle v, u_1 \rangle u_1 + \langle v, u_2 \rangle u_2 + \cdots + \langle v, u_m \rangle u_m = 0.$$

Since $\{u_1, ..., u_m\}$ is linearly independent, however, the only solution to this equation is to have $\langle v, u_i \rangle = 0$, which means v is orthogonal to everything in U, so $v \in U^\perp$.

By the rank-nullity theorem,

$$\dim(\ker(T)) + \dim(\mathrm{im}(T)) = n,$$

but this becomes

$$\dim(U^\perp) + \dim(U) = n,$$

and moving $\dim(U)$ to the other side gives the result. □

Exercise 13.3.
Show that for any subspace U, $(U^\perp)^\perp = U$.

Given a collection of vectors $\{v_1, v_2, ..., v_n\}$ in $\mathbb{R}^m$, we say that $\{v_1, ..., v_n\}$ is an ***orthogonal set*** if each pair of vectors v_i and v_j (with $i \neq j$) are orthogonal.

Example 13.1.
In $\mathbb{R}^3$, the set

$$\left\{ v_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 6 \\ -3 \\ 2 \end{pmatrix}, v_3 = \begin{pmatrix} 4 \\ -2 \\ -15 \end{pmatrix} \right\}$$

is an orthogonal set. To see this we simply check that each vector is orthogonal to each other vector:

$$\begin{aligned} \langle v_1, v_2 \rangle &= 1 \cdot 6 + 2 \cdot (-3) + 0 \cdot 2 = 0 \\ \langle v_1, v_3 \rangle &= 1 \cdot 4 + 2 \cdot (-2) + 0 \cdot (-15) = 0 \\ \langle v_2, v_3 \rangle &= 6 \cdot 4 + (-2) \cdot (-3) + 2 \cdot (-15) = 0 \end{aligned}$$

Example 13.2.

In $\mathbb{R}^3$, the set

$$\left\{ v_1 = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 11 \\ 3 \end{pmatrix}, v_3 = \begin{pmatrix} 3 \\ 0 \\ -2 \end{pmatrix} \right\}$$

is not an orthogonal set. To see this, notice that even though $\langle v_1, v_2 \rangle = 0$ and $\langle v_1, v_3 \rangle = 0$, we have

$$\langle v_2, v_3 \rangle = 1 \cdot 3 + 11 \cdot 0 + 3 \cdot (-2) = -3 \neq 0.$$

Exercise 13.4.
Show that no collection of three or more non-zero vectors in $\mathbb{R}^2$ can form an orthogonal set.

Theorem 13.4.
If S is an orthogonal set of non-zero vectors in $\mathbb{R}^n$, then S is a linearly independent set.

Proof.
We must show that any finite collection of vectors v_1, ..., v_n in S is linearly independent. So let $v_1, ..., v_n \in S$ and consider the equation

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0.$$

Now consider taking the inner product of the zero vector with some v_i in this collection of vectors. Then, by Exercise 13.2, we know $\langle 0, v_i \rangle = 0$. However, we can write the zero vector as $\lambda_1 v_1 + \cdots + \lambda_n v_n$. Thus,

$$\langle \lambda_1 v_1 + \cdots + \lambda_n v_n, v_i \rangle = 0.$$

If we use the linearity of the inner product, we can rewrite this as

$$\lambda_1 \langle v_1, v_i \rangle + \lambda_2 \langle v_2, v_i \rangle + \cdots + \lambda_n \langle v_n, v_i \rangle = 0.$$

By assumption S is an orthogonal set, however, so each $\langle v_j, v_i \rangle = 0$ *except* for the term containing $\langle v_i, v_i \rangle$. We thus have

$$\lambda_i \langle v_i, v_i \rangle = 0.$$

But we're assuming $0 \notin S$, thus $v_i \neq 0$, and so $\langle v_i, v_i \rangle = 0$. Hence we must have $\lambda_i = 0$. We can repeat this argument to determine that each term of

$$\lambda_1 v_1 + \cdots + \lambda_n v_n$$

is multiplied by zero. Thus these vectors are linearly independent, and since these were arbitrarily chose from S, is a linearly independent set. □

As a corollary of the above theorem, notice that every orthogonal set of non-zero vectors is a basis for some subspace of $\mathbb{R}^n$: namely the subspace spanned by those vectors. Having a basis consisting of orthogonal vectors is often very convenient because it makes it easy to determine the scalars the basis vectors are multiplied by as we will see in the theorem below. We say a basis is an ***orthogonal basis*** if the basis forms an orthogonal set.

Theorem 13.5.
Suppose that $\{b_1, b_2, ..., b_n\}$ *is an orthogonal basis for* $\mathbb{R}^n$. *Then for any* $v \in V$, v *may be written as a linear combination of the basis vectors*

$$v = \lambda_1 b_1 + \cdots + \lambda_n b_n$$

where the scalars λ_i *may be calculated as*

$$\lambda_i = \frac{\langle v, b_i \rangle}{\langle b_i, b_i \rangle}.$$

Proof.
Consider the inner product $\langle v, b_i \rangle$ which we may write as

$$\begin{aligned}\langle v, b_i \rangle &= \langle \lambda_1 b_1 + \cdots + \lambda_n b_n, b_i \rangle \\ &= \lambda_1 \langle b_1, b_i \rangle + \cdots + \lambda_n \langle b_n, b_i \rangle\end{aligned}$$

Since $\{b_1, b_2, ..., b_n\}$ is an orthogonal basis, each $\langle b_j, b_i \rangle = 0$ except for $\langle b_i, b_i \rangle$, and so

$$\langle v, b_i \rangle = \lambda_i \langle b_i, b_i \rangle$$

thus we can solve for λ_i,

$$\lambda_i = \frac{\langle v, b_i \rangle}{\langle b_i, b_i \rangle}.$$

□

Theorem 13.6.
Every subspace of $\mathbb{R}^n$ has an orthogonal basis $\mathcal{B}$.

We will actually prove a stronger version of Theorem 13.6 later, so right now we simply remark that such a basis always exists.

13.3 Orthogonal Projection

There are some applications where it is helpful to split vectors up as a linear combination of orthogonal components. In physics, for example, it might be necessary to compute the work done as a particle moves from one position to another. If we recall that work is force times distance, it's tempting to use this formula to compute the work done. However, "work is force times distance" is only true if the direction of the force is the same as the direction of the displacement. Thus we might want to see "how much" of the force points in the same direction as the displacement. If $\vec{F}$ is the force and $\vec{d}$ the displacement, we thus want to write the force as

$\vec{F} = \vec{F}_d + \vec{F}_o$ where $\vec{F}$ is in the direction of the displacement and $\vec{F}_o$ is orthogonal. We can compute $\vec{F}_d$ using *orthogonal projection*.

In general, if u and v are vectors in an inner product space, then we can always express v as a sum of two vectors: one parallel to u (in our abstract setting this means a vector which is a scalar multiple of u) and another vector orthogonal to u. The vector parallel to u is called the ***orthogonal projection*** of v onto u. The orthogonal projection of v onto u is denoted $\text{proj}_u(v)$ and is defined as

$$\text{proj}_u(v) = \frac{\langle v, u\rangle}{\langle u, u\rangle} u.$$

By definition, $\text{proj}_u(v)$ is a scalar multiple of u. (The scalar, $\frac{\langle v,u\rangle}{\langle u,u\rangle}$, is sometimes called the ***scalar projection*** of v onto u.)

Example 13.3.
Suppose $u = \begin{pmatrix}3 & 2 & 1\end{pmatrix}^T$ and $v = \begin{pmatrix}0 & 1 & 5\end{pmatrix}^T$. Then

$$\text{proj}_u(v) = \frac{0 \cdot 3 + 1 \cdot 2 + 5 \cdot 1}{3^2 + 2^2 + 1^2} \begin{pmatrix}3 \\ 2 \\ 1\end{pmatrix} = \begin{pmatrix}3/2 \\ 1 \\ 1/2\end{pmatrix}$$

Exercise 13.5.
Verify that $\text{proj}_u(v)$ is orthogonal to $v - \text{proj}_u(v)$. (Hint: First show that it suffices to show that $v - \text{proj}_u(v)$ is orthogonal to u.)

Exercise 13.5 shows that we can write v as a sum of two vectors, one of which is parallel to u ($\text{proj}_u(v)$) and one is orthogonal to u ($v - \text{proj}_u(v)$).

More generally, if U is any subspace of $\mathbb{R}^n$, then we can always write any given vector $v \in V$ as $v = u + w$ where $u \in U$ and $w \in U^{\perp}$. Here we will call the term u the ***orthogonal projection of v onto the subspace U*** and denote it $\text{proj}_U(v)$. To define $\text{proj}_U(v)$ we first have to find an orthogonal basis for U, say $\mathcal{B} = \{b_1, b_2, ..., b_n\}$ is such a basis. Now we define our orthogonal projection as

$$\mathrm{proj}_U(v) = \frac{\langle v, b_1 \rangle}{\langle b_1, b_1 \rangle} b_1 + \frac{\langle v, b_2 \rangle}{\langle b_2, b_2 \rangle} b_2 + \cdots + \frac{\langle v, b_n \rangle}{\langle b_n, b_n \rangle} b_n$$

Exercise 13.6.
Verify that $v - \mathrm{proj}_U(v) \in U^{\perp}$.

For reasons we won't describe just yet, the orthogonal projection $\mathrm{proj}_U(v)$ can be interpreted as the element of U which is "closest" to v. (The thing we haven't really described is how to measure how "close" one vector is to another in a general vector space, but we will do that soon.)

13.4 Practice Problems

Problem 13.1.
Let U be the subspace of $\mathbb{R}^4$ spanned by

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ -1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Compute the orthogonal complement, $U^{\perp}$.

Problem 13.2.
Let U be the subspace of $\mathbb{R}^4$ from Problem (1), and find the orthogonal projection of

$$\begin{pmatrix} 3 \\ 2 \\ 0 \\ 4 \end{pmatrix}$$

onto U.

Problem 13.3.
Show that for a subspace U of $\mathbb{R}^n$, a vector $v \in \mathbb{R}^n$ satisfies $\mathrm{proj}_U(v) = v$ if and only if $v \in U$.

14

Norms

14.1 Definition and Basic Properties

To each vector $v \in \mathbb{R}^n$ we can associate a non-negative number called the ***norm*** of v and denoted $\|v\|$,

$$\|v\| = \sqrt{\langle v, v \rangle}$$

Notice this is just generalizing how to measure the length of a vector in $\mathbb{R}^2$ and $\mathbb{R}^3$: The length of a vector $v = \begin{pmatrix} x & y & z \end{pmatrix}^T$ in $\mathbb{R}^3$ is $\sqrt{x^2 + y^2 + z^2} = \sqrt{\langle v, v \rangle}$. So norms give us a way to measure the length of a vector.

There is an abstract inner product space version of the Pythagorean theorem stated in terms of norms.

Theorem 14.1 (The Pythagorean theorem)**.**
Two vectors u and v in $\mathbb{R}^n$ are orthogonal if and only if

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof.
Notice that

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u + v \rangle + \langle v, u + v \rangle \\ &= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\ &= \|u\|^2 + 2 \langle u, v \rangle + \|v\|^2. \end{aligned}$$

If u and v are orthogonal, then $\langle u, v \rangle = 0$ and so $\|u + v\|^2 = \|u\|^2 + \|v\|^2$. If $\|u + v\|^2 = \|u\|^2 + \|v\|^2$, then the above calculation shows we must have $2 \langle u, v \rangle = 0$ which implies $\langle u, v \rangle = 0$, so u and v are orthogonal. $\square$

Exercise 14.1.
Convince yourself that in the familiar setting of $\mathbb{R}^2$, the theorem above is the "usual" Pythagorean theorem applied to the triangle with sides u, v, and $u + v$.

The following theorem tells us how the inner product of two different vectors relates to the length of the vectors.

Theorem 14.2 (The Cauchy-Schwartz Inequality).
For all $u, v \in \mathbb{R}^n$ we have the following inequality:

$$\langle u, v \rangle \leq \|u\| \, \|v\|.$$

Proof.
Notice that we may write v in terms of u by using orthogonal projection:

$$v = \operatorname{proj}_u(v) + (v - \operatorname{proj}_u(v)) \,.$$

Since $\operatorname{proj}_u(v)$ and $v - \operatorname{proj}_u(v)$ are orthogonal, the Pythagorean theorem tells us

$$\|v\|^2 = \| \operatorname{proj}_u(v) + (v - \operatorname{proj}_u(v)) \|^2 = \| \operatorname{proj}_u(v)\|^2 + \|v - \operatorname{proj}_u(v)\|^2$$

Now notice

$$\begin{aligned}
\|\operatorname{proj}_u(v)\|^2 &= \left\| \frac{\langle v, u\rangle}{\langle u, u\rangle} u \right\|^2 \\
&= \left\langle \frac{\langle v, u\rangle}{\langle u, u\rangle} u, \frac{\langle v, u\rangle}{\langle u, u\rangle} u \right\rangle \\
&= \left(\frac{\langle v, u\rangle}{\langle u, u\rangle} \right)^2 \langle u, u\rangle \\
&= \frac{\langle v, u\rangle^2}{\langle u, u\rangle} \\
&= \frac{\langle v, u\rangle^2}{\|u\|^2}
\end{aligned}$$

Combining this with the above we have

$$\|v\|^2 = \frac{\langle v, u\rangle^2}{\|u\|^2} + \|v - \operatorname{proj}_u(v)\|^2.$$

Since $\|v - \operatorname{proj}_u(v)\|^2 \geq 0$ we thus have

$$\begin{aligned}
\|v\|^2 &\geq \frac{\langle v, u\rangle^2}{\|u\|^2} \\
\implies \|u\|^2 \|v\|^2 &\geq \langle u, v\rangle^2 .
\end{aligned}$$

Now taking square-roots of both sides gives

$$\langle u, v\rangle \leq \|u\| \, \|v\|.$$

□

The Cauchy-Schwartz inequality is a fundamental tool in linear algebra (and other areas of mathematics which use linear algebra), although it is often lurking in the background in the sense that while the Cauchy-Schwartz theorem is necessary for the proofs of many other theorems, we won't typically deal with it on a day-to-day basis. (This is similar to the mean value theorem in calculus. The mean value theorem is an essential ingredient for many of the proofs of commonly used theorems in calculus, even though you may not deal with the mean value theorem yourself very often in solving problems.)

One example of where we need the Cauchy-Schwartz theorem is in the proof of property (4) in Theorem 14.3 below.

Theorem 14.3.
The norm $\|\cdot\|$ satisfies the following four properties for all vectors $u, v \in \mathbb{R}^n$ and lal scalars λ:

1. $\|v\| \geq 0$.
2. $\|v\| = 0$ *if and only if* $v = 0$
3. $\|\lambda v\| = |\lambda|\|v\|$
4. $\|u + v\| \leq \|u\| + \|v\|$

Proof.

1. Recall that the properties the inner product promise us that $\langle v, v\rangle \geq 0$, so $\|v\| = \sqrt{\langle v, v\rangle} \geq 0$ as well.
2. If $v = 0$, then $\langle v, v\rangle = \langle 0, 0\rangle = 0$, so $\|v\| = \sqrt{0} = 0$. If $\|v\| = 0$, then $\sqrt{\langle v, v\rangle} = 0$ which means $\langle v, v\rangle = 0$, and so $v = 0$ by the axioms of an inner product space.
3. $$\begin{aligned} \|\lambda v\| &= \sqrt{\langle \lambda v, \lambda v\rangle} \\ &= \sqrt{\lambda \langle v, \lambda v\rangle} \\ &= \sqrt{\lambda^2 \langle v, v\rangle} \\ &= \sqrt{\lambda^2}\sqrt{\langle v, v\rangle} \\ &= |\lambda|\,\|v\| \end{aligned}$$
4. Let's notice for the fourth property that it suffices to show $\|u+v\|^2 \leq (\|u\| + \|v\|)^2$, and then we can take the square root of

both sides.

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u, u+v\rangle + \langle v, u+v\rangle \\ &= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\ &= \|u\|^2 + 2\langle u,v\rangle + \|v\|^2\end{aligned}$$

By the Cauchy-Schwartz theorem, however, we know that $\langle u, v\rangle$ is no larger than $\|u\|\,\|v\|$, and so

$$\begin{aligned}\|u+v\|^2 &\le \|u\|^2 + 2\|u\|\,\|v\| + \|v\|^2 \\ &= (\|u\| + \|v\|)^2\end{aligned}$$

Taking square roots of both sides gives us

$$\|u+v\| \le \|u\| + \|v\|.$$

□

Property (4) above, that $\|u+v\| \le \|u\| + \|v\|$ is often called the ***triangle inequality*** because, in $\mathbb{R}^2$, it tells us that in the triangle with sides u, v, and $u+v$, the length of the $u+v$ side is no longer than the sum of the lenghts of the u and v side.

Example 14.1.
Compute the length of

$$\begin{pmatrix} 4 \\ 2 \\ -2 \\ 1 \end{pmatrix}$$

in $\mathbb{R}^4$ using the inner product $\langle u, v\rangle = u^T v$.

$$\left\| \begin{pmatrix} 4 \\ 2 \\ -2 \\ 1 \end{pmatrix} \right\| = \sqrt{\begin{pmatrix} 4 & 2 & -2 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ -2 \\ 1 \end{pmatrix}}$$
$$= \sqrt{4^2 + 2^2 + (-2)^2 + 1^2}$$
$$= \sqrt{16 + 4 + 4 + 1}$$
$$= \sqrt{25}$$
$$= 5.$$

14.2 Distance Between Vectors

Not only does a norm give us a way to measure how long a vector is, it gives us a way to measure distance between two vectors. We define the distance between u and v, denoted $\operatorname{dist}(u, v)$, as the norm of the difference of u and v:

$$\operatorname{dist}(u, v) = \|u - v\|.$$

Example 14.2.
Consider $\mathbb{R}^2$ with the usual inner product, $\langle u, v \rangle = u^T v$. If

$$u = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \qquad v = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

then the distance between u and v is

$$\operatorname{dist}(u, v) = \|u - v\|$$
$$= \sqrt{\langle u - v, u - v \rangle}$$
$$= \sqrt{(u - v)^T (u - v)}$$
$$= \sqrt{\begin{pmatrix} x_1 - x_2 & y_1 - y_2 \end{pmatrix} \begin{pmatrix} x_1 - x_2 \\ y_1 - y_2 \end{pmatrix}}$$
$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Notice this is the typical distance formula in $\mathbb{R}^2$.

Having a way of measuring how close two vectors are is extremely useful in applications. One interesting application is facial recognition. Some facial recognition algorithms work by converting a picture of a face into a vector by using the colors of the pixel (converted into numbers) as the components of the vector. Given two different pictures, we can compare how close the corresponding vectors are and consider the two pictures to be pictures of the same person if the vectors are "close enough" together.

We can also use this notion of distance to justify the earlier remark that $\text{proj}_U(v)$ is the vector in U which is closest to v.

Theorem 14.4.
Let U be a subspace of $\mathbb{R}^n$. Then for any $u \in U$, $u \neq \text{proj}_U(v)$, we have

$$\|v - \text{proj}_U(v)\| < \|v - u\|$$

and so $\text{proj}_U(v)$ is the closer to v than any other vector in U.

Proof.
Notice that since $\text{proj}_U(v)$ and u are elements of U, $\text{proj}_U(v) - u$ is in U as well. Since $v - \text{proj}_U(v)$ is in $U^\perp$, we thus have that $v - \text{proj}_U(v)$ and $\text{proj}_U(v) - u$ are orthogonal. Thus by the Pythagorean theorem,

$$\begin{aligned} \|v - u\|^2 &= \|(v - \text{proj}_U(v)) + (\text{proj}_U(v) - u)\|^2 \\ &= \|v - \text{proj}_U(v)\|^2 + \|\text{proj}_U(v) - u\|^2 \end{aligned}$$

If $u \neq \text{proj}_U(v)$, then $\|\text{proj}_U(v) - u\|^2 > 0$ and so

$$\|v - u\|^2 = \|v - \text{proj}_U(v)\|^2 + \text{ some positive number}$$

and so $\|v - \text{proj}_U(v)\|^2 < \|v - u\|^2$, and taking square roots of each side gives the desired inequality. $\square$

Corollary 14.5.
Of all the vectors in a subspace U*,* $\operatorname{proj}_U(v)$ *is the one which is closest to* v*.*

14.3 Unit Vectors and Orthonormal Bases

When a vector v has unit length, i.e. when $\|v\| = 1$, we call v a ***unit vector***. Notice that v is a unit vector if and only if $\langle v, v\rangle = 0$. Thus the factors $\langle b_i, b_i\rangle$ that appear above when writing a vector as a linear combination of vectors in an orthogonal basis disappear if each vector is a unit vector. In general, an orthogonal basis where each vector is a unit vector is called an ***orthonormal basis.***

Example 14.3.
The vector

$$\begin{pmatrix} 6/7 \\ -2/7 \\ 3/7 \end{pmatrix}$$

is a unit vector in $\mathbb{R}^3$.

$$\begin{aligned} \left\| \begin{pmatrix} 6/7 \\ -2/7 \\ 3/7 \end{pmatrix} \right\| &= \frac{36}{49} + \frac{4}{49} + \frac{9}{49} \\ &= 1 \end{aligned}$$

Theorem 14.6.
If $\mathcal{B} = \{b_1, b_2, ..., b_n\}$ *is an orthonormal basis for* $\mathbb{R}^n$*, then each* $v \in \mathbb{R}^n$ *can be written as*

$$v = \langle v, b_1\rangle\, b_1 + \langle v, b_2\rangle\, b_2 + \cdots + \langle v, b_n\rangle\, b_n.$$

Proof.
This is just Theorem 13.5 but where each $\langle b_i, b_i \rangle = 1$. □

Notice that every non-zero vector v can be scaled to a unit vector by replacing v with

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

Exercise 14.2.
Show that $\frac{1}{\|v\|}v$ is a unit vector.

14.4 The Gram-Schmidt Algorithm

We now show how to to find an orthonormal (and hence orthogonal) basis for any subspace of $\mathbb{R}^n$ through an algorithm called the ***Gram-Schmidt process***. The algorithm works as follows. Let $\mathcal{B} = \{b_1, b_2, ..., b_n\}$ be a basis for a subspace V of $\mathbb{R}^m$. We will construct a new basis $\mathcal{U} = \{u_1, u_2, ..., u_n\}$ which will be orthonormal: i.e., each $\|u_i\| = 1$ and $\langle u_i, u_j \rangle = 0$ for $i \neq j$. We will build the $\mathcal{U}$ basis iteratively, expressing u_2 in terms of u_1, u_3 in terms of u_1 and u_2, and so on.

To begin, let $u_1 = \frac{1}{\|b_1\|}b_1$. So u_1 is a unit vector pointing in the same direction as b_1. Now consider the orthogonal projection of b_2 onto u_1,

$$\text{proj}_{u_1}(b_2) = \langle b_2, u_1 \rangle u_1.$$

Then the vector $b_2 - proj_{u_1}(b_1)$ is orthogonal to u_1. Scaling to make this vector have length one, we set

$$u_2 = \frac{1}{\|b_2 - \text{proj}_{u_1}(b_1)\|}\left(b_2 - \text{proj}_{u_1}(b_1)\right).$$

Now let W_2 denote the subspace spanned by u_1 and u_2 and consider the orthogonal projection of b_3 onto W_2:

$$\text{proj}_{W_2}(b_3) = \langle b_3, u_1 \rangle u_1 + \langle b_3, u_2 \rangle u_2.$$

Then $b_3 - proj_{W_2}(b_3)$ is an element of $W_2^{\perp}$. Scaling so that this vector has unit length, we set u_3 to be

$$u_3 = \frac{1}{\|b_3 - \mathrm{proj}_{W_2}(b_3)\|}\left(b_3 - \mathrm{proj}_{W_2}(b_3)\right).$$

We continue in this way, building onto our list of orthonormal basis vectors one at a time.

The orthogonality of the basis vectors is the real important part of the process: we could first build an orthogonal basis and then scale all our vectors to unit vectors at the very end. Doing so, we can express the Gram-Schmidt process as follows:

1. Let $v_1 = b_1$.
2. Once v_k is found, set $W_k = \mathrm{span}(v_1, v_2, ..., v_k)$ and define $v_{k+1} = b_{k+1} - \mathrm{proj}_{W_k}(b_{k+1})$. Writing out what the projection gives

$$v_{k+1} = b_{k+1} - \frac{\langle b_{k+1}, v_1\rangle}{\langle v_1, v_1\rangle}v_1 - \frac{\langle b_{k+1}, v_2\rangle}{\langle v_2, v_2\rangle}v_2 - \cdots - \frac{\langle b_{k+1}, v_k\rangle}{\langle v_k, v_k\rangle}v_k$$

   Continue this procedure until v_n is found.
3. Set $u_k = \frac{1}{\|v_k\|}v_k$.

At the end of this procedure, $\{u_1, u_2, ..., u_n\}$ will be an orthonormal basis.

Example 14.4.
Apply the Gram-Schmidt process to orthonormalize the following basis for of $\mathbb{R}^4$:

$$\left\{ b_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 2 \\ 1 \\ 2 \\ 2 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \quad b_4 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

We will first build an orthogonal basis $\{v_1, v_2, v_3, v_4\}$, and then scale each vector to have unit length. We begin by setting

$$v_1 = b_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ -1 \end{pmatrix}.$$

To find v_2 we take $W_1 = \mathrm{span}(v_1)$, and then find a vector in $W_1^\perp$ by setting

$$\begin{aligned} v_2 &= b_2 - \mathrm{proj}_{W_1}(v_1) \\ &= b_2 - \frac{\langle b_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1 \\ &= \begin{pmatrix} 2 \\ 1 \\ 2 \\ 2 \end{pmatrix} - \frac{2+1+0-2}{1^2+1^2+0^1+(-1)^2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 1 \\ 2 \\ 2 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 0 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} 5/3 \\ 2/3 \\ 2 \\ 7/3 \end{pmatrix} \end{aligned}$$

Now we set $W_2 = \mathrm{span}(v_1, v_2)$ and find $v_3 \in W_2^\perp$ by taking

$$\begin{aligned} v_3 &= b_3 - \mathrm{proj}_{W_2}(b_3) \\ &= b_3 - \frac{\langle b_3, v_1\rangle}{\langle v_1, v_1\rangle} v_1 - \frac{\langle b_3, v_2\rangle}{\langle v_2, v_2\rangle} v_2 \\ &= \begin{pmatrix} 5/19 \\ 2/19 \\ -13/19 \\ 7/19 \end{pmatrix} \end{aligned}$$

Next we take W_3 to be $\mathrm{span}(v_1, v_2, v_3)$ and find $v_4 \in W_3^\perp$ by taking

$$\begin{aligned} v_4 &= b_4 - \mathrm{proj}_{W_3}(b_4) \\ &= b_4 - \frac{\langle b_4, v_1\rangle}{\langle v_1, v_1\rangle} v_1 - \frac{\langle b_4, v_2\rangle}{\langle v_2, v_2\rangle} v_2 - \frac{\langle b_4, v_3\rangle}{\langle v_3, v_3\rangle} v_3 \\ &= \begin{pmatrix} -6/13 \\ 8/13 \\ 0 \\ 2/13 \end{pmatrix} \end{aligned}$$

Now, $\{v_1, v_2, v_3, v_4\}$ is an orthogonal set of non-zero vectors, which implies that it is a linearly independent set, and since this is a subset of $\mathbb{R}^4$, it is also a basis. So we have an orthogonal basis. However, our vectors do not have unit length. To get an orthonormal basis we need to scale each vector so that it has unit length, which we do by multiplying the vector by one over its norm.

First we calculate the norm of each of our vectors,

$$\begin{aligned}
\|v_1\| &= \sqrt{\langle v_1, v_1\rangle} = \sqrt{1^2 + 1^2 + 0^2 + (-1)^2} = \sqrt{3} \\
\|v_2\| &= \sqrt{{}^{38}\!/_{3}} \\
\|v_3\| &= \sqrt{{}^{13}\!/_{19}} \\
\|v_4\| &= 2\sqrt{{}^{2}\!/_{13}}
\end{aligned}$$

Dividing each vector by its norm gives us a unit vector,

$$u_1 = \frac{1}{\|v_1\|} v_1 = \begin{pmatrix} {}^{1}\!/_{3\sqrt{3}} \\ {}^{1}\!/_{3\sqrt{3}} \\ 0 \\ -{}^{1}\!/_{3\sqrt{3}} \end{pmatrix}$$

$$u_2 = \frac{1}{\|v_2\|} v_2 = \begin{pmatrix} {}^{5}\!/_{3\sqrt{38/3}} \\ {}^{2}\!/_{3\sqrt{38/3}} \\ {}^{2}\!/_{\sqrt{38/3}} \\ {}^{7}\!/_{3\sqrt{38/3}} \end{pmatrix}$$

$$u_3 = \frac{1}{\|v_3\|} v_3 = \begin{pmatrix} {}^{5}\!/_{19\sqrt{13/19}} \\ {}^{2}\!/_{19\sqrt{13/19}} \\ -{}^{13}\!/_{19\sqrt{13/19}} \\ {}^{7}\!/_{19\sqrt{13/19}} \end{pmatrix}$$

$$u_4 = \frac{1}{\|v_4\|} v_4 = \begin{pmatrix} -{}^{6}\!/_{13\sqrt{8/13}} \\ {}^{8}\!/_{13\sqrt{8/13}} \\ 0{}^{2}\!/_{13\sqrt{8/13}} \end{pmatrix}$$

Now $\{u_1, u_2, u_3, u_4\}$ is an orthonormal basis for $\mathbb{R}^4$.

14.5 Practice Problems

Problem 14.1.
Show that for any three vectors u, v, and w in $\mathbb{R}^n$,

$$\operatorname{dist}(u, w) \leq \operatorname{dist}(u, v) + \operatorname{dist}(v, w).$$

Problem 14.2.
Prove the following ***parallelogram law***: for all $u, v \in \mathbb{R}^n$,

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|y\|^2.$$

A Solutions to Exercises

A.1 Chapter 1

1.1 Example 1.1(e) is linear because it can be written as

$$0 \cdot x_1 + 0 \cdot x_2 + \cdots + 0 \cdot x_n,$$

but Example 1.2(e) is not linear because it does not have the form

$$(x_1, ..., x_n) \mapsto a_1 \cdot x_1 + a_2 \cdot x_2 + \cdots + a_n \cdot x_n.$$

If it did have this form, then we would have

$$(2x_1, ..., 2x_n) \mapsto 2a_1 \cdot x_1 + 2a_2 \cdot x_2 + \cdots + 2a_n \cdot x_n.$$

and as

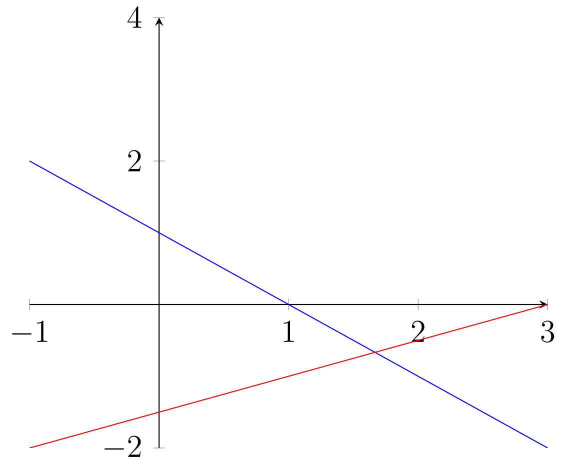
$$a_1 \cdot x_1 + a_2 \cdot x_2 + \cdots + a_n \cdot x_n = 1$$

we would have

$$2a_1 \cdot x_1 + 2a_2 \cdot x_2 + \cdots + 2a_n \cdot x_n = 2$$

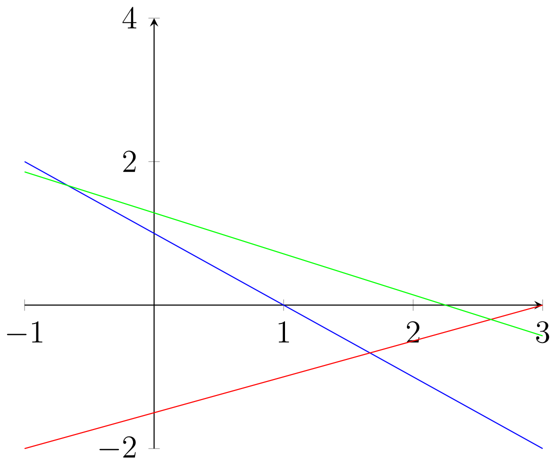
but our function is supposed to send *everything* to 1, yet here is something that gets sent to 2.

1.2 (a) We'll plot $x + y = 1$ in blue and $x - 2y = 3$ in red:



Since the lines intersect, there is a solution to the system. Solving the system gives $(x, y) = (5/3, -2/3)$.

(b) We'll add $4x + 7y = 9$ in green.



Notice that there is no point that lies on each of the lines, and hence there are no solutions to the system.

1.3 Suppose (t_1, t_2) is a solution to the system. I.e.,

$$\begin{aligned}(a_1 + c\alpha_1)t_1 + (a_2 + c\alpha_2)t_2 &= b + c\beta \\ \alpha_1 t_2 + \alpha_2 t_2 &= \beta.\end{aligned}$$

Now subtract c times the second equation from the first equation. On the left-hand side this gives

$$\begin{aligned}&(a_1 + c\alpha_1)t_1 + (a_2 + c\alpha_2)t_2 - c(\alpha_1 t_1 + \alpha_2 t_2) \\ =&a_1 t_1 + a_2 t_2\end{aligned}$$

On the right-hand side we have $b + c\beta - c\beta = b$. Thus (t_1, t_2) satisfies

$$a_1 t_1 + a_2 t_2 = b.$$

Hence (t_1, t_2) is a solution to the system

$$\begin{aligned}a_1 x + a_2 y &= b \\ \alpha_1 x + \alpha_2 y &= \beta.\end{aligned}$$

1.4 Keeping the same notation as in the proof of the first part of Theorem 1.2, we must show that $T \subseteq S$. Let $(t_1, ..., t_n) \in T$ be a solution to the modified system in which the equation

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

has been replaced by

$$(a_1 + c\alpha_1)x_1 + \cdots + (a_n + c\alpha_n)x_n = b + c\beta.$$

We simply need to show that $(t_1, ..., t_n)$ also satisfies our original first equation,

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

Note that $(t_1, ..., t_n)$ satisfies the following:

$$(a_1 + c\alpha_1)t_1 + \cdots + (a_n + c\alpha_n)t_n = b + c\beta.\alpha_1t_1 + \cdots + \alpha_nt_n = \beta$$

If we subtract c times the second equation from the the first, we have on the left-hand side

$$\begin{aligned} &(a_1 + c\alpha_1)t_1 + \cdots + (a_n + c\alpha_n)t_n - c\,(\alpha_1t_1 + \cdots + \alpha_nt_n) \\ =&a_1t_1 + \cdots + a_nt_n \end{aligned}$$

and on the right-hand side, $b + c\beta - c\beta = b$. Hence $(t_1, ..., t_n)$ is a solution to the original system of equations; $(t_1, ..., t_n) \in S$ and so $S \subseteq T$.

1.5 Suppose that we are given a system of equations

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

and we modify the system by replacing one of the equations with some non-zero multiple c of itself. Without loss of generality, we may assume this is the first equation. (If it were a different equation, we could swap two equations to make it the first equation.)

Let S be the set of solutions to the original system, and T the set of solutions to the modified system. Obviously if $(s_1, ..., s_n) \in S$ satisfies

$$a_{11}s_1 + \cdots + a_{1n}s_n = b_1,$$

then in the modified system we simply have

$$ca_{11}s_1 + \cdots + ca_{1n}s_n = c(a_{11}s_1 + \cdots + a_{1n}s_n) = cb$$

and so $S \subseteq T$ as we have a solution to the modified system.

Likewise, if $(t_1, ..., t_n)$ satisfies the first equation in the modified system,

$$ca_{11}t_1 + \cdots + ca_{1n}t_n = cb_1,$$

then

$$a_{11}t_1 + \cdots + a_{1n}t_n == \frac{1}{c}(ca_{11}t_1 + \cdots + ca_{1n}t_n) = \frac{1}{c} \cdot cb_1 = b_1$$

and thus $T \subseteq S$.

A.2 Chapter 2

2.1 Putting this matrix in RREF gives

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus there are three pivot columns.

2.2 In RREF this matrix becomes

$$\left(\begin{array}{cccc|c} 1 & 0 & -1 & -1 & 1 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Since the system is consistent and there are two non-pivot columns, any parametrization of the set of solutions to the system must contain two free variables.

A.3 Chapter 3

3.1 The proofs of these properties are basically identical: we just write out the vectors in components and verify the left-hand side of each equation matches the right-hand side.

Since the proofs are identical we won't give each one, but we will prove the first one which can easily be modified for the other properties:

$$\begin{aligned}
&(\vec{u}+\vec{v})+w \\
=&\left(\begin{pmatrix}u_1\\u_2\\\vdots\\u_n\end{pmatrix}+\begin{pmatrix}v_1\\v_2\\\vdots\\v_n\end{pmatrix}\right)+\begin{pmatrix}w_1\\w_2\\\vdots\\w_n\end{pmatrix} \\
=&\begin{pmatrix}u_1+v_1\\u_2+v_2\\\vdots\\u_n+v_n\end{pmatrix}+\begin{pmatrix}w_1\\w_2\\\vdots\\w_n\end{pmatrix} \\
=&\begin{pmatrix}u_1+v_1+w_1\\u_2+v_2+w_2\\\vdots\\u_n+v_n+w_n\end{pmatrix} \\
=&\begin{pmatrix}u_1\\u_2\\\vdots\\u_n\end{pmatrix}+\begin{pmatrix}v_1+w_1\\v_2+w_2\\\vdots\\v_n+w_n\end{pmatrix} \\
=&\begin{pmatrix}u_1\\u_2\\\vdots\\u_n\end{pmatrix}+\left(\begin{pmatrix}v_1\\v_2\\\vdots\\v_n\end{pmatrix}+\begin{pmatrix}w_1\\w_2\\\vdots\\w_n\end{pmatrix}\right) \\
=&\vec{u}+(\vec{v}+\vec{w})
\end{aligned}$$

A.4 Chapter 4

4.1 Suppose first that the system $A\vec{x}=\vec{b}$ has a solution. Then, by the definition of the product $A\vec{x}$, $\vec{b}$ is a linear combination of the columns of A. In particular, if the components of $\vec{x}$ are x_1, x_2, ..., x_n, then we have

$$\vec{b}=x_1\vec{a}_1+x_2\vec{a}_2+\cdots+x_n\vec{a}_n.$$

Now suppose that $\vec{b}$ is in the span of the columns of A, again, by the definition of the product $A\vec{x}$, this precisely means the system $A\vec{x}=\vec{b}$ has a solution. In particular, if

$$\vec{b} = \lambda_1\vec{a}_1 + \lambda_2\vec{a}_2 + \cdots + \lambda_n\vec{a}_n,$$

then a solution to $A\vec{x} = \vec{b}$ is given by

$$\vec{x} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}$$

A.5 Chapter 5

5.1 For any scalar λ, of which there are infinitely-many choices, the vector $\vec{y} = \lambda\vec{x}$ is another solution.

5.2 1.

$$\left\{ \begin{pmatrix} 1 + 2x_3 \\ -1 - x_3 + x_4 \\ x_3 \\ x_4 \end{pmatrix} \middle| x_3, x_4 \in \mathbb{R} \right\}$$

2.

$$\left\{ \begin{pmatrix} 22 - 7x_4 \\ 2 - x_4 \\ -10 + 3x_4x_4 \end{pmatrix} \middle| x_4 \in \mathbb{R} \right\}$$

3.

$$\left\{ \begin{pmatrix} 18 + x_4 - 16x_5 \\ -45 - 9x_4 + 54x_5 \\ -15 + x_4 + 19x_5 \end{pmatrix} \middle| x_4, x_5 \in \mathbb{R} \right\}$$

4.

$$\left\{ \begin{pmatrix} -3x_2 - 2x_5 \\ 6 - x_3 - 3x_4 - x_5 \\ \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \middle| x_2, x_3, x_4, x_5 \in \mathbb{R} \right\}$$

A.6 Chapter 6

6.1 (a) Consider the vector equation,

$$\lambda_1\vec{e}_1 + \lambda_2\vec{e}_2 + \cdots + \lambda_n\vec{e}_n = \vec{0}.$$

In components this is

$$\begin{pmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_n\end{pmatrix} = \begin{pmatrix}0\\ 0\\ \vdots\\ 0\end{pmatrix}$$

The only solution to this is clearly $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$.

(b) An n-dimensional vector $\vec{x}$ with components x_1, x_2, ..., x_n may be written as

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n.$$

6.2 (a) Consider the vectors

$$\vec{u} = \begin{pmatrix}u_1\\ u_2\end{pmatrix} \quad \vec{v} = \begin{pmatrix}v_1\\ v_2\end{pmatrix}.$$

Notice

$$\begin{aligned} S_m(\vec{u}+\vec{v}) &= S_m\begin{pmatrix}u_1+v_1\\ u_2+v_2\end{pmatrix} \\ &= \begin{pmatrix}u_1+v_1+m(u_2+v_2)\\ u_2+v_2\end{pmatrix} \\ &= \begin{pmatrix}u_1+mu_2\\ u_2\end{pmatrix} + \begin{pmatrix}v_1+mv_2\\ v_2\end{pmatrix} \\ &= S_m(\vec{u}+\vec{v}) \end{aligned}$$

$$\begin{aligned} S_m(\lambda\vec{u}) &= S_m\begin{pmatrix}\lambda u_1\\ \lambda u_2\end{pmatrix} \\ &= \lambda u_1 + m\lambda u_2 \\ \lambda u_2 & \\ &= \lambda\begin{pmatrix}u_1+mu_2\\ u_2\end{pmatrix} \\ &= \lambda S_m(\vec{u}). \end{aligned}$$

(b)

$$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$$

6.3 Notice that

$$R\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}$$

and this implies R is linear.

6.4 The proofs of each of the above properties are basically identical: simply write out the left-hand and right-hand sides of each property and verify that everything is equal. This is easy, but quite tedious, so we will just give the proof of the first property.

Suppose that A and B are $m \times n$ matrices where the element in the i-th row and j-th column of each matrix is a_{ij} and b_{ij}, respectively. Then the element in the i-th row and j-th column of the sum $A + B$ is $a_{ij} + b_{ij}$. Since a_{ij} and b_{ij} are just numbers (real or complex, it doesn't matter), we know $a_{ij} + b_{ij} = b_{ij} + a_{ij}$, but this is the entry in the i-th row and j-th column of $B + A$. Thus $A + B$ and $B + A$ have the same entries are so are the same matrix: $A + B = B + A$.

6.5 This is another theorem that is easy, but tedious, to verify just by writing out what the matrices look like in components. We will give the details for the second property, however.

Suppose that A is $m \times n$, B is $n \times p$, and C is $n \times p$, so that the sums $B + C$ and $AB + AC$ and the products $A(B + C)$, AB and AC are all defined. Let a_{ij} denote the entry in the i-th row and j-th column of A, and likewise the entries of B and C are b_{ij} and c_{ij}.

Notice that the entry in the i-th row and j-th column of AB is

$$\sum_{k=1}^{n} a_{ik}b_{kj}$$

and similarly, the entry in the i-th row and j-th column of AC is

$$\sum_{k=1}^{n} a_{ik}c_{kj}$$

Hence the corresponding entry in $AB + AC$ is

$$\sum_{k=1}^{n} a_{ik}b_{kj} + \sum_{k=1}^{n} a_{ik}c_{kj} = \sum_{k=1}^{n} (a_{ik}b_{kj} + a_{ik}c_{kj}) = \sum_{k=1}^{n} a_{ik}\left(b_{kj} + c_{kj}\right)$$

As the entry in the i-th row and j-th column of $B+C$ is $b_{ij}+c_{ij}$, the entry in the i-th row and j-th column of $A(B+C)$ is thus

$$\sum_{k=1}^{n} a_{ik}\left(b_{kj}+c_{kj}\right).$$

Thus $A(B+C)=AB+AC$ since these matrices have the same entries.

6.6 Notice that the i-th row, j-th column of λA is λa_{ij}, and so the entry in the i-th row, j-th column of $(\lambda A)^T$ is λa_{ji}, but this is exactly the entry in the i-th row, j-th column of $\lambda(A^T)$, and so $(\lambda A)^T = \lambda(A^T)$.

Suppose A is $m\times n$ and B is $n\times p$. The i-th row, j-th column of $(AB)^T$ is the same as the j-th row, i-th column of AB which is

$$\sum_{k=1}^{n} a_{jk}b_{ki}.$$

The i-th row, j-th column of B^TA^T is

$$\sum_{k=1}^{n} b_{ki}a_{jk}.$$

Since $a_{jk}b_{ki}=b_{ki}a_{jk}$, the matrices are the same.

6.7 For $\ell\neq i$, notice that row ℓ of E is all zeros except for a 1 in the ℓ-th position, thus in row ℓ column j of the matrix EA we have

$$\sum_{k=1}^{n} e_{\ell k}a_{kj}=a_{\ell j}$$

since $e_{\ell k}=0$ for all k except $e_{\ell\ell}=1$.

Similarly, in row i the i-th row, j-th column of EA is

$$\sum_{k=1}^{n} e_{ik}a_{kj}=ca_{ij}$$

since $e_{ik}=0$ for all k except $e_{ii}=c$.

6.8 We are assuming $AB=I$. Multiplying both sides on the right by B^{-1} gives $ABB^{-1}=IB^{-1}$ but since $BB^{-1}=I$ this simplifies to $A=B^{-1}$. Similarly if we were to multiply both sides on the left by A^{-1} we have $A^{-1}AB=A^{-1}I$ which simplifies to $B=A^{-1}$.

A.7 Chapter 7

7.1 Suppose the entries of U_1 are u_{ij}, and the entries of U_2 are v_{ij}. Since these matrices are upper triangular, u_{ij} and v_{ij} are both zero if $i > j$ (if $i > j$, then u_{ij} and v_{ij} are below the diagonal).

The entry in the i-th row, j-th column of the product U_1U_2 is

$$\sum_{k=1}^{n} u_{ik}v_{kj}.$$

Suppose that $i > j$ and rewrite the sum above as

$$\sum_{k=1}^{n} u_{ik}v_{kj} = \sum_{k=1}^{j} u_{ik}v_{kj} + \sum_{k=j+1}^{n} u_{ik}v_{kj}.$$

In the first sum on the right-hand side, as $k \leq j$ and $i > j$, we have $i > k$ and so each u_{ik} is zero. In the second sum, $k > j$ and so $v_{kj} = 0$. Hence each term in the sum is zero.

7.2 Every lower triangular matrix without zeros on the diagonal can easily be written as a product of elementary matrices.

A.8 Chapter 8

8.1 Since λA multiplies each column of A by λ, and since the determinant is linear in *each* column, we can pull a λ out of each column. For an $n \times n$ matrix that means we're pulling λ's out of n columns, so we pick up a factor of λ^n.

8.2 Consider cofactor expansion along the first row:

$$\begin{aligned}
&\det(A)\\
=&a \det\begin{pmatrix} e & f \\ h & i \end{pmatrix} - b\det\begin{pmatrix} d & f \\ g & i \end{pmatrix} + c\det\begin{pmatrix} d & e \\ g & h \end{pmatrix}\\
=&a(ei - fh) - b(di - fg) + c(dh - eg)\\
=&aei - afh - bdi + bfg + cdh - ceg\\
=&aei + bfg + cdh - ceg - bdi - afh.
\end{aligned}$$

8.3 Consider the three types of elementary matrices separately.

1. If E is an elementary matrix obtained by multiplying one row of the identity by λ, then Theorem 8.6 tells us that $\det(E) = \lambda$. The second part of Corollary 8.4 tells us that $\det(EA) = \lambda \det(A)$ since EA is A with a row multiplied by λ. Hence
$$\det(E)\det(A) = \lambda \det(A) = \det(EA).$$
2. If E is an elementary matrix obtained by adding λ times the i-th row of the identity to the the j-th row of the identity, then Theorem 8.6 tells us $\det(E) = 1$. The first part of Corollary 8.4 tells us that $\det(EA)$ is $\det(A) + \det(A')$ where A' is the matrix A but with the j-th row obtained by λ times the i-th row. Since the rows of this matrix are linearly dependent (the j-th row of A' is a multiple of the i-th row), we have $\det(A') = 0$. Hence
$$\det(EA) = 1 \cdot \det(A) = \det(E) \cdot \det(A).$$
3. If E is an elementary matrix obtained by swapping two rows, then $\det(E) = -1$ by Theorem 8.6, and $\det(EA) = -\det(A)$ by the third part of Corollary 8.4, and so
$$\det(EA) = -1 \cdot \det(A) = \det(E) \cdot \det(A).$$

8.4 Suppose, for the sake of contradiction, AB was invertible; say the inverse was C so $ABC = I$. This would mean $A \cdot (BC) = I$ and so $BC = A^{-1}$, but this is a contradiction since A was assumed to be non-invertible.

A.9 Chapter 9

9.1 (a) This set is obviously non-empty; it is closed under addition as $\vec{0} + \vec{0} = \vec{0}$; and it is closed under scalar multiplication as $\lambda \cdot \vec{0} = \vec{0}$ for every $\lambda \in \mathbb{R}$.

(b) $\mathbb{R}^n$ is obviously non-zero (e.g., $\vec{0} \in \mathbb{R}^n$); it is clearly closed under vector addition since the sum of two n-dimensional vectors is an n-dimensional vector; and similarly is closed under scalar multiplication as for any $\lambda \in \mathbb{R}$ and any $v \in \mathbb{R}^n$, $\lambda \cdot v \in \mathbb{R}^n$.

9.2 We want to find the scalars α, β, γ such that
$$\begin{pmatrix} 2 \\ 4 \\ -1 \end{pmatrix} = \alpha \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix} + \beta \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} + \gamma \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$$

That is, we're solving the system of equations

$$\begin{pmatrix} 4 & 2 & 2 \\ 1 & 2 & -1 \\ 3 & 2 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ -1 \end{pmatrix}.$$

Since the vectors giving the columns of this matrix are linearly independent (because they form a basis), this matrix is invertible and so we compute

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 1 & 2 & -1 \\ 3 & 2 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 4 \\ -1 \end{pmatrix} = \begin{pmatrix} -1/3 & -2/3 & 1 \\ 1/2 & 1 & -1 \\ 2/3 & 1/3 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ -1 \end{pmatrix} = \begin{pmatrix} -13/3 \\ 6 \\ 11/3 \end{pmatrix}$$

and so with respect to $\mathcal{B}_1$ we have

$$v = \begin{pmatrix} -13/3 \\ 6 \\ 11/3 \end{pmatrix}_{\mathcal{B}}.$$

Similarly, for the basis $\mathcal{B}_2$ we compute

$$\begin{pmatrix} 0 & 0 & 1 \\ 4 & 1 & 2 \\ 0 & 1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 4 \\ -1 \end{pmatrix} = \begin{pmatrix} 7/4 \\ -7 \\ 2 \end{pmatrix}$$

and so

$$v = \begin{pmatrix} 7/4 \\ -7 \\ 2 \end{pmatrix}_{\mathcal{B}}.$$

9.3 If the vector were in the image, then there would be a choice of

$$v = \begin{pmatrix} x \\ y \end{pmatrix}$$

such that

$$\begin{pmatrix} x + y \\ x - y \\ 2x \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}.$$

Writing out the matrix representing this linear transformation we have the following system of equations:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}.$$

Attempting to row-reduce the augmented coefficient matrix of this system we would have

$$\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & -1 & -1 \\ 2 & 0 & 2 \end{array}\right) \xrightarrow{RREF} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)$$

meaning our system of equations is equivalent to one which has the equation $0x + 0y = 1$, which clearly has no solution.

A.10 Chapter 10

10.1 Notice that non-zero $v \in \ker(T)$ is an eigenvector with eigenvalue 0 as $T(v) = 0 = 0 \cdot v$. Thus if T is not injective, and so there are non-zero elements of the kernel, then T has 0 as an eigenvalue. If T has eigenvalue 0, then by definition this means there exists a non-zero vector v such that $T(v) = 0 \cdot v = 0$, so $\ker(T)$ contains non-zero elements, and T is not injective.

10.2 We want to find the vectors v solving $Av = -v$ which we can rewrite as $(A + I)v = 0$. This means we are trying to solve the system

$$(A + I)v = \begin{pmatrix} 3 & 3 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Putting the matrix in RREF gives

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

and so the system is solve when $x = -y$. Hence the space of eigenvectors for this matrix with eigenvalue -1 is

$$\left\{ \begin{pmatrix} y \\ -y \end{pmatrix} \mid y \in \mathbb{R} \right\}$$

A.11 Chapter 11

11.1 To find the geometric multiplicities we can use the rank-nullity theorem.

For the eigenvalue $\lambda = -2i$ our matrix is

$$A = \begin{pmatrix} 4-2i & 10 & 8 \\ -2 & -4-2i & -3 \\ 0 & 0 & 3-2i \end{pmatrix}.$$

which in RREF becomes

$$\begin{pmatrix} 1 & 2+i & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Since this matrix has one non-pivot column, its kernel (which is the eigenspace associated with eigenvalue $\lambda = 2i$) has dimension 1, so the geometric multiplicity of $\lambda = 2i$ is 1.

Similarly, for eigenvalue $\lambda = 3$ we have the matrix

$$A = \begin{pmatrix} 1 & 10 & 8 \\ -2 & -7 & -3 \\ 0 & 0 & 0 \end{pmatrix}.$$

which in RREF becomes

$$\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

which has one non-pivot column, and so the eigenvalue $\lambda = 3$ has geometric multiplicity 1.

11.2 Notice $\overline{i \cdot i} = \overline{-1} = -1$, but if conjugation were complex linear we would require

$$\overline{i \cdot i} = i \cdot \overline{i} = i \cdot -i = -i^2 = 1.$$

11.3 Suppose $z = x + iy$ and $z' = x' + iy'$. We then simply compute

$$\begin{aligned} \overline{z} \cdot \overline{z'} &= (x - iy) \cdot (x' - iy') \\ &= xx' - ixy' - ix'y - yy' \\ &= (xx' - yy') - i(xy' + x'y) \end{aligned}$$

$$\begin{aligned} \overline{z \cdot z'} &= \overline{(x + iy)(x' + iy')} \\ &= \overline{xx' + ixy' + ix'y - yy'} \\ &= \overline{(xx' - yy') + i(xy' + x'y)} \\ &= (xx' - yy') - i(xy' + x'y) \end{aligned}$$

A.12 Chapter 12

12.1 (a) If a matrix is diagonal, then it consists of 1×1 Jordan blocks on the diagonal and zeros everywhere else, so it is in Jordan normal form.

(b) If a matrix is in Jordan normal form but not diagonal, then there must exist a Jordan block of size at least 2×2. This means the corresponding eigenvalue does not have full geometric multiplicity (i.e., the geometric multiplicity is at least one less than the algebraic multiplicity), and hence there are "not enough" eigenvectors for that eigenvalue for the matrix to be diagonalizable.

A.13 Chapter 13

13.1 1.

$$\begin{aligned}\langle u, v\rangle &= u_1v_1 + u_2v_2 + u_3v_3 + \cdots + u_nv_n \\ &= v_1u_1 + v_2u_2 + v_3u_3 + \cdots + v_nu_n \\ &= \langle v, u\rangle\end{aligned}$$

2.

$$\begin{aligned}\langle u+v, w\rangle &= (u_1+v_1)w_1 + (u_2+v_2)w_2 + (u_3+v_3)w_3 + \cdots (u_n+v_n)w_n \\ &= u_1w_1 + v_1w_1 + u_2w_2 + v_2w_2 + u_3w_3 + v_3w_3 + \cdots + u_nw_n + v_nw_n \\ &= u_1w_1 + u_2w_2 + u_3w_3 + \cdots + u_nw_n + v_1w_1 + v_2w_2 + v_3w_3 + \cdots + v_nw_n \\ &= \langle u, w\rangle + \langle v, w\rangle\end{aligned}$$

3.

$$\begin{aligned}\langle \lambda u, v\rangle &= \lambda u_1v_1 + \lambda u_2v_2 + \lambda u_3v_3 + \cdots + \lambda u_nv_n \\ &= \lambda(u_1v_1 + u_2v_2 + u_3v_3 + \cdots + u_nv_n) \\ &= \lambda \langle u, v\rangle\end{aligned}$$

4.

$$\begin{aligned}\langle u, u\rangle &= u_1^2 + u_2^2 + u_3^2 + \cdots u_n^2 \\ &\geq 0\end{aligned}$$

5. If $\langle u, u\rangle = 0$, then $u_1^2 + u_2^2 + u_3^2 + \cdots + u_n^2 = 0$ which means each of u_1, u_2, u_3, ..., u_n must be zero, so $u = 0$. If $u = 0$, clearly $\langle u, u\rangle = \langle 0, 0\rangle = 0$.

13.2

$$\langle 0, v\rangle = \langle 0 \cdot v, v\rangle = 0 \cdot \langle v, v\rangle = 0.$$

13.3 Let $u \in U$. To show that $u \in \left(U^\perp\right)^\perp$ we must show $\langle u, v\rangle = 0$ for each $v \in U^\perp$. But if $v \in U^\perp$, then by definition $\langle v, u\rangle = \langle u, v\rangle = 0$ for all $u \in U$. Thus $u \in \left(U^\perp\right)^\perp$, this establishes that $U \subseteq \left(U^\perp\right)^\perp$. Thus U is a subspace of $\left(U^\perp\right)^\perp$, but notice the dimension of $\left(U^\perp\right)^\perp$ is

$$\dim\left(\left(U^\perp\right)^\perp\right) = \dim(V) - \dim(U^\perp) = \dim(V) - (\dim(V) - \dim(U)) = \dim(U).$$

So U is a subspace of $\left(U^\perp\right)^\perp$ of the same dimension, and hence $U = \left(U^\perp\right)^\perp$.

13.4 No collection of three or more vectors in $\mathbb{R}^2$ can be linearly independent since $\mathbb{R}^2$ is two-dimensional.

13.5 Notice that $\text{proj}_u(v)$ is a scalar multiple of u, so if $v - \text{proj}_u(v)$ is orthogonal to u, then $v - \text{proj}_u(v)$ is orthogonal to $\text{proj}_u(v)$. So it suffices to show $v - \text{proj}_u(v)$ is orthogonal to u.

$$\begin{aligned}
&\langle v - \text{proj}_u(v), u\rangle \\
=&\langle v, u\rangle - \langle \text{proj}_u(v), u\rangle \\
=&\langle v, u\rangle - \left\langle \frac{\langle v, u\rangle}{\langle u, u\rangle} u, u\right\rangle \\
=&\langle v, u\rangle - \frac{\langle v, u\rangle}{\langle u, u\rangle}\langle u, u\rangle \\
=&\langle v, u\rangle - \langle v, u\rangle \\
=&0.
\end{aligned}$$

13.6 Notice that

$$\text{proj}_U(v) = \text{proj}_{b_1}(v)b_1 + \text{proj}_{b_2}(v)b_2 + \cdots + \text{proj}_{b_n}(v)b_n.$$

Consider

$$\langle v - \text{proj}_U(v), b_i\rangle = \left\langle v - \text{proj}_{b_1}(v) - \text{proj}_{b_2}(v) - \cdots - \text{proj}_{b_n}(v), b_i\right\rangle = \langle v, b_i\rangle - \left\langle \text{proj}_{b_i}(v), b_i\right\rangle$$

Notice that each of the $\left\langle \operatorname{proj}_{b_j}(v), b_i \right\rangle$ equals zero as $\operatorname{proj}_{b_j}(v)$ is a scalar multiple of b_j which is orthogonal to b_i. The quantity above can be rewritten as

$$\langle v - \operatorname{proj}_{b_i}(v), b_i \rangle$$

which we know is zero by the previous exercise.

A.14 Chapter 14

14.2

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

B

Solutions to Practice Problems

B.1 Chapter 1

1.1 There is one unique solution: $(2/15, 1/10)$.

1.2 This is one unique solution: $(7, 4)$.

1.3 There are infinitely-many solutions. Each equation represents the same line – solve each equation for y and they will all give you

$$y = -4x + 5$$

Thus every point on this line is a solution to the system. One possible parametrization to the set of solutions is

$$(x, -4x + 5).$$

If we had instead solved for x to get $x = \frac{1}{4}(-y + 5)$ we would obtain the parametrization

$$\left(\frac{-y + 5}{4}, y\right).$$

1.4 This system has infinitely-many solutions. Geometrically, each equation represents a plane in three-dimensional spaces, and these two planes intersect along a line. Every point on that line is a solution to the system.

Algebraically, we could subtract twice the first equation from the second to turn the system into

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

Solving the second equation for y tells us

$$y = \frac{3z - 2}{5}.$$

Plugging this into the first equation and solving for x gives

$$x = \frac{13 - 2z}{5}$$

Thus the solution set is parametrized by

$$\left(\frac{13 - 2z}{5}, \frac{3z - 2}{5}, z\right).$$

1.5 Notice that the first two equations of this system are the same as the first two equations of the last system. Geometrically, we are intersecting the line of solutions from the last system with another plane to pick out a single point on the line.

The unique solution to this system is

$$\left(\frac{-5}{3}, 6, \frac{32}{3}\right).$$

1.6 If we subtract the first equation from the third we obtain the equivalent system

$$\begin{aligned} 2x + 3y &= 4 \\ y - 4z &= 3 \\ y - 4z &= -4 \end{aligned}$$

At this point it's clear the system has no solutions: we can't make the $y - 4z$ equal to both 3 and -4 at the same time since $3 \neq -4$. Since this equivalent system has no solutions, the original system must not have any solutions either.

B.2 Chapter 2

2.1 Since there are many different possible echelon forms for a given matrix, it's easier to give the RREF of the matrices. To check your answer, convert your non-RREF echelon form matrices into RREF and see if you get the RREF of these matrices.

(a)

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(b)

$$\begin{pmatrix} 1 & 0 & 0 & 0 & -75 \\ 0 & 1 & 0 & 0 & 34 \\ 0 & 0 & 1 & 0 & 45 \\ 0 & 0 & 0 & 1 & -3 \end{pmatrix}$$

2.2 (a)

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

(b)

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

(c)

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

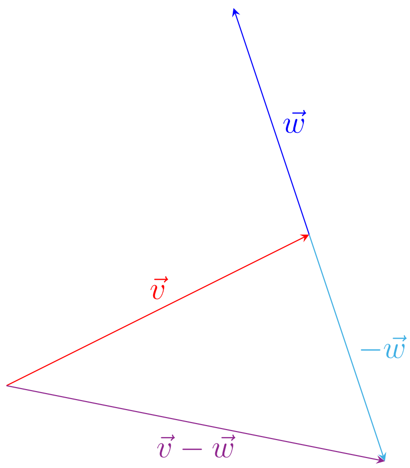
2.3 (a) $(x, y, z) = (2, 3, {}^1\!/\!_2)$.

(b) There are infinitely-many solutions. Taking z to be the free variable, the solutions are parametrized by $(3 - 2z, -z, z)$.

(c) There are no solutions.

B.3 Chapter 3

3.1 Since $\vec{v} - \vec{w}$ is really $\vec{v} + (-\vec{w})$, we simply perform the same triangle law but with $-\vec{w}$.



3.2 We can rewrite this as a system of linear equations,

$$\begin{aligned} 3x + 5y &= 2 \\ -2x + 0y &= -3 \\ 8x + 9y &= 8 \end{aligned}$$

Writing the augmented coefficient matrix of this system and then putting it in to RREF, we see that there are no solutions.

3.3 Given any vector $\begin{pmatrix} a \\ b \end{pmatrix}$ we can write

$$\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

3.4 Given a vector $\begin{pmatrix} a \\ b \end{pmatrix}$, we want to find values of x and y such that

$$x \begin{pmatrix} -1 \\ 3 \end{pmatrix} + y \begin{pmatrix} 4 \\ -2 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

We can turn this into a linear system,

$$\begin{aligned} -x + 4y &= a \\ 3x - 2y &= b \end{aligned}$$

We can always solve such a system: putting the augmented coefficient matrix into RREF we get

$$\left(\begin{array}{cc|c} 1 & 0 & \frac{a+2b}{5} \\ 0 & 1 & \frac{3a+b}{10} \end{array}\right)$$

which means we can write

$$\frac{a+2b}{5} \begin{pmatrix} -1 \\ 3 \end{pmatrix} + \frac{3a+b}{10} \begin{pmatrix} 4 \\ -2 \end{pmatrix}.$$

For example, if $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \end{pmatrix}$, then

$$\frac{16}{5} \begin{pmatrix} -1 \\ 3 \end{pmatrix} + \frac{13}{10} \begin{pmatrix} 4 \\ -2 \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \end{pmatrix}.$$

3.5 In each situation we want to use the theorem saying that a set of vectors is linearly independent if and only if we can only write

$$\lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2 + \cdots + \lambda_n \vec{v}_n = \vec{0}$$

by taking $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$. That is, all zeros is the only solution to a certain system of equations. In the case of the vectors in part (a), for instance, the system is

$$\begin{aligned} \lambda_1 + \lambda_3 &= 0 \\ -2\lambda_1 + \lambda_2 - 6\lambda_3 &= 0 \\ 2\lambda_2 + 8\lambda_3 &= 0 \end{aligned}$$

Of course, $(\lambda_1, \lambda_2, \lambda_3) = (0, 0, 0)$ is a solution to this system, but our theorem tells us that this will be the only solution if and only if the vectors are linearly independent. Thus in each case we want to see if the only solution to the system of linear equations is all zeros.

(a) Not linearly independent.

(b) Not linearly independent.

(c) Not linearly independent. (You don't actually need to do any calculations here. Since the solution to (a) and (b) are not linearly independent, and this is just all of those vectors together, it has no hope of being linearly independent.)

3.6 $\vec{u}$ will be in the span of the other vectors if we can write $\vec{u}$ as a linear combination of the other vectors. To see if we can do this or not, we turn this into a question of systems of linear equations. In particular, we will be able to write $\vec{u}$ as a linear combination of th eother vectors only if the system whose augmented coefficient matrix has these vectors as its columns,

$$\left(\begin{array}{cccc|c} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m & \vec{u} \end{array}\right),$$

is consistent. That is, when we put this matrix into RREF, we *don't* get a row of the form

$$\left(\begin{array}{cccc|c} 0 & 0 & \cdots & 0 & x \neq 0 \end{array}\right)$$

3.7 Again, we turn this into a question of systems of linear equations. We want to see if we can find λ_1, λ_2, λ_3 such that

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 4 \end{pmatrix} + \lambda_3 \begin{pmatrix} 3 \\ 2 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \\ 2 \\ -3 \end{pmatrix}$$

This is the same as seeing if there is a solution to

$$\begin{aligned} \lambda_1 + 3\lambda_3 &= 4 \\ \lambda_2 + 2\lambda_3 &= 7 \\ -2\lambda_1 + \lambda_2 + \lambda_3 &= 2 \\ \lambda_1 + 4\lambda_2 - 2\lambda_3 &= -3 \end{aligned}$$

Writing the augmented coefficient matrix of this system and putting it into RREF we get

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

and so there is no solution – $\vec{u}$ *is not* in the span of the given vectors.

B.4 Chapter 4

4.1 (a)

$$\vec{x} = \begin{pmatrix} -11 \\ 3 \\ 0 \end{pmatrix}$$

(b)

$$\vec{x} = \begin{pmatrix} -11/2 \\ 3/2 \\ 0 \end{pmatrix}$$

(c)

$$\vec{x} = \begin{pmatrix} -4 \\ 4 \\ 3 \end{pmatrix}$$

(d)

$$\vec{x} = \begin{pmatrix} 5/2 \\ 3/2 \\ x_3 \end{pmatrix}$$

where x_3 is free.

(e) No solution.

4.5 No. Put the matrix into RREF and notice there is a row without pivots. By Proposition 4.1 from lecture, the columns can not span all of $\mathbb{R}^4$.

4.3 No. Put the matrix into RREF and notice there is a row without pivots. By Proposition 4.1 from lecture, the columns can not span all of $\mathbb{R}^4$.

4.4 Yes. Put the matrix whose columns are given by these vectors in RREF and notice there is a pivot in every row.

B.5 Chapter 6

6.1 (a) $AB + F$

This is defined and

$$AB + F = \begin{pmatrix} 8 & 7 & 7 & 1 \\ 8 & 6 & 8 & 3 \\ 15 & 3 & 3 & 0 \end{pmatrix}$$

(b) $2D - 3E$

$$2D - 3E = \begin{pmatrix} -17 & -2 \\ 5 & -10 \\ -9 & 11 \end{pmatrix}$$

(c) AC

This product is not defined because A is 3×3 and C is 4×3.

(d) $4AD + E$

$$4AD + E = \begin{pmatrix} 33 & 18 \\ 9 & 20 \\ 15 & -3 \end{pmatrix}$$

(e) FBC

This product is not defined as F is 3×4 and B is 3×4.

(f) BCF

$$BCF = \begin{pmatrix} 56 & 2 & 22 & 33 \\ 101 & 17 & 53 & 36 \\ -9 & -9 & -13 & 9 \end{pmatrix}$$

(g) $AF - DE$

This is undefined. A is 3×3 and F is 3×4, so the product AF *is* defined and is 3×4 matrix. D is 3×2 and E is 3×2, so the product DE is *not* defined.

6.2

$$A^T = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 0 & -1 & -3 \end{pmatrix} \qquad B^T = \begin{pmatrix} 0 & 2 & 4 \\ 1 & 3 & 0 \\ 1 & 3 & 0 \\ 2 & -1 & 1 \end{pmatrix}$$

$$C^T = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & 2 & 4 & 2 \\ -1 & -3 & 5 & 3 \end{pmatrix} \qquad D^T = \begin{pmatrix} -1 & 4 & 0 \\ 2 & 1 & 1 \end{pmatrix}$$

$$E^T = \begin{pmatrix} 5 & 1 & 3 \\ 2 & 4 & -3 \end{pmatrix} \qquad F^T = \begin{pmatrix} 4 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 3 & -1 \\ 1 & 1 & 2 \end{pmatrix}$$

6.3 From problem 5a we know that each column of AB is a linear combination of the columns of A. In particular, if

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then the first column of AB is

$$a \begin{pmatrix} 1 \\ -3 \end{pmatrix} + c \begin{pmatrix} -3 \\ 5 \end{pmatrix},$$

and the second column of AB is

$$b \begin{pmatrix} 1 \\ -3 \end{pmatrix} + d \begin{pmatrix} -3 \\ 5 \end{pmatrix}$$

Hence we have two linear systems we need to solve:

$$\begin{pmatrix} 1 & -3 \\ -3 & 5 \end{pmatrix} \begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & -3 \\ -3 & 5 \end{pmatrix} \begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} -1 \\ 17 \end{pmatrix}$$

In general, if we take the matrix

$$\left(\begin{array}{cc|c} 1 & -3 & x \\ -3 & 5 & y \end{array}\right)$$

and put it into RREF, then we have

$$\left(\begin{array}{cc|c} 1 & 0 & -{}^{5x}/_{4} - {}^{3y}/_{4} \\ 0 & 1 & -{}^{3x}/_{4} - {}^{y}/_{4} \end{array}\right)$$

Taking x and y to be -3 and 1, for the first column of AB, we see that $a = -{}^{5}/_{4}(-3) - {}^{3}/_{4}(1) = 3$, and $c = 2$.

For the second column, take x and y to be -1 and 17 respectively, and the above tells us that $b = -{}^{23}/_{2}$ and $d = -{}^{7}/_{2}$.

Hence our matrix B is

$$B = \begin{pmatrix} 3 & -{}^{23}/_{2} \\ 2 & -{}^{7}/_{2} \end{pmatrix}$$

6.4 Notice that in the product $\vec{r}A$, each element of the first row of A gets multiplied by r_1; each element of the second row of A gets multiplied by r_2; and so on. The elements in the first column are then added together to get the first entry in $\vec{r}A$; the elements in the second column are then added to get the second entry in $\vec{r}A$ and so on.

$$\begin{pmatrix} r_1 & r_2 & \cdots & r_m \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} r_1a_{11} + r_2a_{21} + \cdots + r_ma_{m1} & r_1a_{12} + r_2a_{22} + \cdots + r_ma_{m2} & \cdots & r_1a_{1n} + r_2a_{2n} + \cdots + r_ma_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^m r_ia_{1i} & \sum_{i=1}^m r_ia_{2i} & \cdots & \sum_{i=1}^m r_ia_{ni} \end{pmatrix}$$

Notice this can be written as

$$r_1 \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix} + r_2 \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix} + \cdots + r_m \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

That is, $\vec{r}A$ is a linear combination of the rows of A.

6.5 (a) The entry in the i-th row and j-th column of AB has the form

$$\sum_{k=1}^{n} a_{ik}b_{kj}.$$

So the j-th column of A has the form

$$\begin{pmatrix} \sum_{k=1}^{n} a_{1k}b_{kj} \\ \sum_{k=1}^{n} a_{2k}b_{kj} \\ \vdots \\ \sum_{k=1}^{n} a_{mk}b_{kj} \end{pmatrix}$$

Writing this out we have

$$\begin{pmatrix} a_{11}b_{1j} + a_{12}b_{2j} + \cdots + a_{1n}b_{nj} \\ a_{21}b_{1j} + a_{22}b_{2j} + \cdots + a_{2n}b_{nj} \\ \vdots \\ a_{m1}b_{1j} + a_{m2}b_{2j} + \cdots + a_{mn}b_{nj} \end{pmatrix}$$

Let's separate this into a sum of vectors, where we isolate the components multiplied by each b_{kj} which we may then write as

$$\begin{pmatrix} a_{11}b_{1j} \\ a_{21}b_{1j} \\ \vdots \\ a_{m1}b_{1j} \end{pmatrix} + \begin{pmatrix} a_{12}b_{2j} \\ a_{22}b_{2j} \\ \vdots \\ a_{m2}b_{2j} \end{pmatrix} + \cdots + \begin{pmatrix} a_{1n}b_{nj} \\ a_{2n}b_{nj} \\ \vdots \\ a_{mn}b_{nj} \end{pmatrix}$$

Factoring the b_{kj} out of each vector we have

$$b_{1j}\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + b_{2j}\begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + b_{nj}\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

This is a linear combination of the columns of A, and we get linear combinations of the columns of A by multiplying A with a column vector. In particular, if $\vec{b}_j$ denotes the j-th column of B, then we have show that the j-th column of AB is $A\vec{b}_j$.

(b) Proceeding as in part (a), we know that the i-th row of AB has the form

$$\begin{pmatrix} \sum_{k=1}^{n} a_{ik}b_{k1} & \sum_{k=1}^{n} a_{ik}b_{k2} & & \sum_{k=1}^{n} a_{ik}b_{kp} & \end{pmatrix}$$

Each of these entries has a term with a factor of a_{i1}, and one with a factor of a_{i2}, and so on. We separate the row, breaking it up into a sum of rows where everything is multiplied by a_{i1}, plus a row where everything is multiplied by a_{i2}, and so forth:

$$\begin{aligned} &\begin{pmatrix} a_{i1}b_{11} & a_{i1}b_{12} & \cdots & a_{i1}b_{1p} \end{pmatrix} \\ +&\begin{pmatrix} a_{i2}b_{21} & a_{i2}b_{22} & \cdots & a_{i2}b_{2p} \end{pmatrix} \\ +&\vdots \\ +&\begin{pmatrix} a_{in}b_{n1} & a_{in}b_{n2} & \cdots & a_{in}b_{np} \end{pmatrix} \end{aligned}$$

Now factor out the a_{ik}'s,

$$\begin{aligned} a_{i1}&\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \end{pmatrix} \\ +a_{i2}&\begin{pmatrix} b_{21} & b_{22} & \cdots & b_{2p} \end{pmatrix} \\ +\cdots& \\ +a_{in}&\begin{pmatrix} b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix} \end{aligned}$$

Notice that this is exactly the i-th row of A times the matrix B.

6.6 Notice that the third column is equal to the first column plus the second column. That is, if $\vec{a}_1$, $\vec{a}_2$, and $\vec{a}_3$ denote the first, second, and third columns of A, then

$$\vec{a}_3 = \vec{a}_1 + \vec{a}_2,$$

which means

$$\vec{a}_1 + \vec{a}_2 - \vec{a}_3 = \vec{0}.$$

Thus

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 9 \\ 6 & 7 & 13 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \vec{0}.$$

Since the columns of AB are linear combinations of the columns of A and we now have a linear combination which gives us the zero vector, we know

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 9 \\ 6 & 7 & 13 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ -1 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

6.7 (a) This matrix is invertible and its inverse is

$$\frac{1}{5-6}\begin{pmatrix} 1 & -3 \\ -2 & 5 \end{pmatrix} = \begin{pmatrix} -1 & 3 \\ 2 & -5 \end{pmatrix}$$

(b) This matrix is invertible and its inverse is

$$\frac{1}{4-(-9)}\begin{pmatrix} 4 & -3 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 4/13 & -3/13 \\ 3/13 & 1/13 \end{pmatrix}.$$

(c) This matrix is not invertible because $6 \cdot 4 - 8 \cdot 3 = 0$.

(d) This matrix is invertible and its inverse is

$$\begin{pmatrix} 2/7 & 4/7 & -5/21 \\ 1/7 & 2/7 & 1/21 \\ -4/21 & -1/21 & -4/63 \end{pmatrix}$$

(e) This matrix is invertible and its inverse is

$$\begin{pmatrix} 1/3 & -4/3 & 3/2 & -16/3 \\ 0 & 1 & -2 & 10 \\ 0 & 0 & 1/2 & -4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

6.8 No, if a matrix has two identical columns it can not be invertible. Having two identical columns implies the matrix's columns are not linearly independent, and so the corresponding linear transformation can not be injective.

6.9 No, such a matrix is not invertible for the same reason as the previous problem. The columns are not linearly independent, and so the corresponding linear transformation is not injective.

6.10 If a matrix is not invertible, then the corresponding linear transformation is not bijective. This means the linear transformation is either not surjective or not injective; in general a non-bijective map could be surjective but not injective, or injective but not surjective, but for *linear* maps between two spaces of the same dimension, you're surjective if and only if you're injective.

To see this notice that our non-invertible matrix has either a row or a column which doesn't have a pivot, depending on whether the map is not surjective or not injective. But since the matrix is square, if you're missing a pivot in a row, then you're also missing a pivot in a column.

Hence our map is not injective, so its columns can not be linearly independent, so some column is a linear combination of the others.

6.11 Notice that the matrix corresponding to this linear transformation is

$$\begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix}$$

The inverse of this matrix, which is the matrix of T^{-1}, is

$$\frac{1}{6-(-1)} \begin{pmatrix} 3 & 1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 3/7 & 1/7 \\ -1/7 & 2/7 \end{pmatrix}$$

This tells us that

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (3x+y)/7 \\ (-x+2y)/7 \end{pmatrix}$$

B.6 Chapter 7

7.1 If we multiply out the block matrices we have

$$\left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline A & I & 0 \\ \hline B & D & I \end{array}\right) \left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline P & I & 0 \\ \hline Q & R & I \end{array}\right) = \left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline A+P & I & 0 \\ \hline B+PD+Q & D+R & I \end{array}\right)$$

But we're assuming these matrices are inverses, so this has to be identity,

$$\left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline A+P & I & 0 \\ \hline B+PD+Q & D+R & I \end{array}\right) = \left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline 0 & I & 0 \\ \hline 0 & 0 & I \end{array}\right).$$

But this implies

$$A+P = B+PD+Q = D+R = 0.$$

Thus $P=-A$ and $R=-D$. Then $B+PD+Q = B-AD+Q = 0$, so $Q = AD-B$.

7.2 If the matrix is invertible, then its RREF is the identity, but this implies the RREF of each matrix on the diagonal is also the identity. So if the "large" matrix is invertible, then so are the "small" submatrices on the diagonal.

If the diagonal submatrices are invertible, then when the matrix is put into RREF we also put the submatrices into RREF but these will be identity matrices. Since everything below the diagonal is zero, this means the RREF of the entire matrix is the identity, and so the matrix is invertible.

7.3 The inverse is

$$\begin{pmatrix} 3/2 & -2 & 0 & 0 & 0 & 0 \\ -1/2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 5 & -2 \\ 0 & 0 & 0 & 0 & -2 & 1 \end{pmatrix}$$

7.4 1.

$$\begin{pmatrix} 2 & 6 \\ 4 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 6 \\ 0 & -5 \end{pmatrix}$$

2.

$$\begin{pmatrix} 2 & -4 & 2 \\ -4 & 5 & 2 \\ 6 & -9 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -4 & 2 \\ 0 & -3 & 6 \\ 0 & 0 & 1 \end{pmatrix}$$

3.

$$\begin{pmatrix} 1 & -2 & -2 & -3 \\ 3 & -9 & 0 & -9 \\ -1 & 2 & 4 & 7 \\ -3 & -6 & 26 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -3 & 4 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & -2 & -3 \\ 0 & -3 & 6 & 0 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

7.5 By our LU factorization we have

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -3 & 4 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & -2 & -3 \\ 0 & -3 & 6 & 0 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \\ 0 \\ 3 \end{pmatrix}$$

We first solve the system

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -3 & 4 & -2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \\ 0 \\ 3 \end{pmatrix}$$

Using back substitution we have

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 1 \\ 4 \end{pmatrix}$$

Now we solve

$$\begin{pmatrix} 1 & -2 & -2 & -3 \\ 0 & -3 & 6 & 0 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 1 \\ 4 \end{pmatrix}$$

Using back substitution again this becomes

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -34 \\ -16 \\ -{}^{15}\!/_{2} \\ 8 \end{pmatrix}$$

B.7 Chapter 8

8.1 (a) The determinant is 14.

(b) The determinant is 5.

(c) The determinant is -6.

8.2 (a) The determinant is -8.

(b) The determinant is 6.

8.3

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} {}^{3}\!/_{40} \\ {}^{7}\!/_{4} \\ {}^{11}\!/_{10} \end{pmatrix}$$

8.4 (a) This matrix is not invertible.

(b) The inverse is

$$\frac{1}{-48} \begin{pmatrix} 3 & 11 & -10 \\ 6 & -10 & -4 \\ -27 & -19 & 26 \end{pmatrix}$$

(c) The inverse is

$$\frac{1}{-42}\begin{pmatrix} 50 & -16 & -20 & -42 \\ -14 & -14 & 14 & 0 \\ -13 & 5 & 1 & 21 \\ 7 & 7 & -7 & -21 \end{pmatrix}$$

B.8 Chapter 9

9.1 V is a subspace.

9.2 V is not a subspace.

9.3 V is not a subspace.

9.4 $U \cap V$ is a subspace.

9.5 $U \cup V$ is not a subspace in general, but it will be a subspace if $U \subseteq V$ or $V \subseteq U$.

9.6 A basis for the image is given by

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

A basis for the kernel is given by

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\}$$

9.7 (a) The change of basis matrix is

$$_{\mathcal{B}}I_{\mathcal{S}} = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/5 \end{pmatrix}$$

Thus

$$[v]_{\mathcal{B}} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/5 \end{pmatrix} \begin{pmatrix} -2 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 3/5 \end{pmatrix}$$

(b) The change of basis matrix is

$$ {}_{\mathcal{B}}I_{\mathcal{S}} = \begin{pmatrix} 2 & 3 \\ 1 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 5/7 & -3/7 \\ -1/7 & 2/7 \end{pmatrix} $$

Thus

$$ [v]_{\mathcal{B}} = \begin{pmatrix} 5/7 & -3/7 \\ -1/7 & 2/7 \end{pmatrix} \begin{pmatrix} -2 \\ 3 \end{pmatrix} = \begin{pmatrix} -19/7 \\ 8/7 \end{pmatrix} $$

(c) The change of basis matrix is

$$ {}_{\mathcal{B}}I_{\mathcal{S}} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 4 \\ 1 & -2 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 4/3 & -5/3 & 4/3 \\ 1/3 & -1/6 & -1/6 \\ -1/3 & 2/3 & -1/3 \end{pmatrix} $$

Thus

$$ [v]_{\mathcal{B}} = \begin{pmatrix} 4/3 & -5/3 & 4/3 \\ 1/3 & -1/6 & -1/6 \\ -1/3 & 2/3 & -1/3 \end{pmatrix} \begin{pmatrix} 3 \\ 7 \\ 0 \end{pmatrix} = \begin{pmatrix} -23/3 \\ -1/6 \\ 11/3 \end{pmatrix} $$

9.8 (a)

$$ \mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 7 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\} $$

$$ \begin{pmatrix} 0 & 9 & -27 \\ 4 & 3 & 19 \end{pmatrix} $$

(b)

$$ \mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right\} $$

$$ \begin{pmatrix} 13 & 1 & 20 \\ 10/3 & 1 & 14/3 \end{pmatrix} $$

(c)

$$ \mathcal{B} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} $$

$$ \begin{pmatrix} 3 & 1 & 4 \\ 2 & 1 & 7 \end{pmatrix} $$

(d)

$$\mathcal{B} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$

$$\begin{pmatrix} 2 & 1 & 7 \\ 3 & 1 & 4 \end{pmatrix}$$

B.9 Chapter 10

10.1 (a) The eigenvalues are 5 and 2. The eigenspace of 5 is

$$\left\{ \begin{pmatrix} x \\ x \end{pmatrix} \middle| \, x \in \mathbb{R} \right\}.$$

The eigenspace of 2 is

$$\left\{ \begin{pmatrix} x \\ -x/2 \end{pmatrix} \middle| \, x \in \mathbb{R} \right\}.$$

(b) The eigenvalues are $\frac{1\pm\sqrt{5}}{2}$. The eigenspace of $\frac{1+\sqrt{5}}{2}$ is

$$\left\{ \begin{pmatrix} x \\ x(1+\sqrt{5})/2 \end{pmatrix} \middle| \, x \in \mathbb{R} \right\}.$$

The eigenspace of $\frac{1-\sqrt{5}}{2}$ is

$$\left\{ \begin{pmatrix} x \\ x(1-\sqrt{5})/2 \end{pmatrix} \middle| \, x \in \mathbb{R} \right\}.$$

(c) The eigenvalues are 6, $-\sqrt{2}$ and $\sqrt{2}$. The eigenspace of 6 is

$$\left\{ \begin{pmatrix} x \\ x \\ x \end{pmatrix} \middle| \, x \in \mathbb{R} \right\}$$

The eigenspace of $-\sqrt{2}$ is

$$\left\{ \begin{pmatrix} z(-13-9\sqrt{2})/7 \\ z(5+11\sqrt{2})/7 \\ z \end{pmatrix} \middle| \, z \in \mathbb{R} \right\}$$

The eigenspace of $\sqrt{2}$ is

$$\left\{ \begin{pmatrix} z(-13+9\sqrt{2})/7 \\ z(5-11\sqrt{2})/7 \\ z \end{pmatrix} \middle| \, z \in \mathbb{R} \right\}$$

(d) The eigenvalues are 1 and 2. The eigenspace of 1 is

$$\left\{ \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} \middle| x \in \mathbb{R} \right\}$$

The eigenspace of 2 is

$$\left\{ \begin{pmatrix} 2y \\ y \\ 0 \end{pmatrix} \middle| y \in \mathbb{R} \right\}$$

10.2 Notice that the matrix representing this linear transformation, with respect to the standard basis, is

$$\begin{pmatrix} 2 & -1 & 0 \\ 3 & 2 & 7 \\ 6 & 1 & 8 \end{pmatrix}$$

The eigenvalues of this matrix are 8, 4 and 0. If we take a basis given by an eigenvector of each eigenvalue, such as

$$\left\{ \begin{pmatrix} 7 \\ -42 \\ -39 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \right\}$$

then with respect to this basis, T is represented by the matrix

$$\begin{pmatrix} 8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

B.10 Chapter 11

11.1 (a) The characteristic polynomial is

$$x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2$$

Thus the eigenvalues are 1 and 2. Eigenvalue 1 has algebraic multiplicity 1 and geometric multiplicity 1. Eigenvalue 2 has algebraic multiplicity 2 and geometric multiplicity 2.

(b) The characteristic polynomial is

$$x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2$$

Thus the eigenvalues are 1 and 2. Eigenvalue 1 has algebraic multiplicity 1 and geometric multiplicity 1. Eigenvalue 2 has algebraic multiplicity 2 and geometric multiplicity 1.

(c) This matrix has characteristic polynomial $(x-4)^3$, and 4 is its only eigenvalue. The algebraic multiplicity is 3, but the geometric multiplicity is 2.

(d) This matrix has characteristic polynomial $(x-4)^3$, and 4 is its only eigenvalue. The algebraic multiplicity is 3, but the geometric multiplicity is 1.

B.11 Chapter 12

12.1 (a) This matrix diagonalizes to

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix}$$

(b) This matrix diagonalizes to

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

(c) This matrix diagonalizes to

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

(d) This matrix does not diagonalize. It has eigenvalues 1 and 3, but eigenvalue 1 has geometric multiplicity 1. Hence a Jordan normal form for the matrix is

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

(e) This matrix does not diagonalize. It has one eigenvalue, 2, which has geometric multiplicity 1. Hence a Jordan normal form for the matrix is

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

B.12 Chapter 13

13.1 We need to find the vectors v that solve the equation $\langle v, u \rangle = 0$ for each $u \in U$. Since we have a basis for U, however, each element of U may be written as

$$u = \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3$$

where u_1, u_2, and u_3 are the vectors above. By linearity of the inner product we thus need to find the vectors v which satisfy each of the following:

$$\begin{aligned} \langle v, u_1 \rangle &= 0 \\ \langle v, u_2 \rangle &= 0 \\ \langle v, u_3 \rangle &= 0 \end{aligned}$$

If we denote v as $v = \begin{pmatrix} w & x & y & z \end{pmatrix}$, then the above equations become

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

$U^\perp$ is then the set of vectors in $\mathbb{R}^4$ solving this system of equations. The coefficient matrix of this system in RREF is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

which tells us

$$\begin{aligned} w &= -z \\ x &= 0 \\ y &= -z \end{aligned}$$

Thus

$$U^\perp = \left\{ \begin{pmatrix} -z \\ 0 \\ -z \\ z \end{pmatrix} \middle| z \in \mathbb{R} \right\}.$$

13.2 Call this vector v and let u_1, u_2, u_3 be the vectors in the basis of U given in the first problem.

$$\begin{aligned}\text{proj}_U(v) &= \frac{\langle v, u_1\rangle}{\langle u_1, u_1\rangle}u_1 + \frac{\langle v, u_2\rangle}{\langle u_2, u_2\rangle}u_2 + \frac{\langle v, u_3\rangle}{\langle u_3, u_3\rangle}u_3 \\ &= \frac{5}{3}\begin{pmatrix}1\\1\\-1\\0\end{pmatrix} + \frac{12}{7}\begin{pmatrix}2\\1\\-1\\1\end{pmatrix} + \frac{11}{6}\begin{pmatrix}1\\2\\0\\1\end{pmatrix} \\ &= \begin{pmatrix}97/14\\148/21\\-71/21\\149/42\end{pmatrix}\end{aligned}$$

13.3 Suppose $\{u_1, u_2, ..., u_n\}$ is a basis for U. Then

$$\text{proj}_U(v) = \sum_{i=1}^{n} \frac{\langle v, u_i\rangle}{\langle u_i, u_i\rangle}u_i$$

If $v \in U$, then $\frac{\langle v, u_i\rangle}{\langle u_i, u_i\rangle}$ is precisely the scalar that u_i is multiplied by when writing v as a linear combination of $\{u_1, ..., u_n\}$, and so $v = \text{proj}_U(v)$.

If $\text{proj}_U(v) = v$, then since $\text{proj}_U(v)$ is a linear combination of the u_i, $v \in U$.

B.13 Chapter 14

14.1

$$\begin{aligned}\text{dist}(u, w) &= \|u - w\| \\ &= \|u - v + v - w\| \\ &\leq \|u - v\| + \|v - w\| \\ &= \text{dist}(u, v) + \text{dist}(v, w)\end{aligned}$$

14.2 Notice

$$\begin{aligned}\|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \langle u, u + v\rangle + \langle v, u + v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 + 2\langle u, v\rangle + \|v\|^2\end{aligned}$$

Similarly,

$$\begin{aligned}\|u-v\|^2 &= \langle u-v, u-v\rangle \\ &= \langle u, u-v\rangle - \langle v, u-v\rangle \\ &= \langle u, u\rangle - \langle u, v\rangle - \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 - 2\langle u, v\rangle + \|v\|^2\end{aligned}$$

Adding these together gives

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2$$