Math 309 - Linear Algebra

A simple description of the subject of (finite dimensional) Linear Algebra might be the study of matrices and vectors. Matrices are arrays that are (usually) filled with real or complex numbers:

$$A = \begin{pmatrix} 1 & 2 & 3 & \cdots & 4 & 5 \\ 6 & 7 & 8 & \cdots & 9 & 10 \\ \vdots & & \vdots & & \vdots & \vdots \\ 11 & 12 & 13 & \cdots & 14 & 15 \end{pmatrix}$$

Vectors are matrices that have either one column or one row:

$$v = \begin{pmatrix} 1 \\ 2 \\ \vdots \\ 3 \end{pmatrix}, \quad v = \begin{pmatrix} 1 & 2 & \cdots & 3 \end{pmatrix}$$

The reason that we care so much about matrices and vectors is that we can take a system of linear equations like

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

and write them as an equation involving matrices and vectors:

$$Ax = b.$$

Another related problem is to find nonzero solutions to equations like

$$Ax = \lambda x,$$

where λ is a real number.

The main purpose of this class will be to study equations like this. But why do we care about these equations? Because they show up everywhere.

- If you are interested in applied math, then you are interested in differential equations. In virtually every interesting application, the differential equations are too complicated to be solved by hand, so you need to use a computer to get a numerical solution. The usual way to do this is to solve an equation of the form $Ax = b$.
- If you are interested in pure math, then you'll need to know about the algebraic properties of matrices, which will factor prominently in our study of these equations. We'll look closely at matrix multiplication, inversion, etc.
- Problems like $Ax = \lambda x$ are the basis for the Google page rank.
- Images can be represented as matrices, and linear algebra is used in the compression of images.
- Many economic models are expressed in matrices and vectors.
- Many, many, more....

In our study of these equations, we need to answer several questions:

- When do these equations have a solution?
- How do we find it?
- If there is a solution, is it unique?
- Among different methods for solving the system, which ones are best for which situations?

All the things that we study in this class contribute to answering these questions, although many of the things we study are interesting in their own right.

The first thing that we will study is the sets where the solutions to our equations live, which are called vector spaces.

Chapter 1: Vector Spaces

Vector spaces are sets, so it is important that we spend a little bit of time making sure that we know what we mean when we talk about a set. You can be as precise as you want when talking about sets, even going to the extremes that set theorists do. However, for our purposes, we'll just stick with

Definition 1 *A set is a collection of objects, called elements.*

Notation: We denote that an object x is an element of the set S by writing $x \in S$. Sometimes it is convenient to write a set by listing all its elements, such as

$$S = \{2, 3, 5, 7\} \text{ or } X = \{x_1, x_2, \ldots, x_n\}.$$

However, for sets with large or infinite numbers of elements, it is often easier to use set-builder notation. For example, let $\mathbb{N}$ be the set of natural numbers (whole numbers greater than zero). Then the set of even natural numbers can be written as

$$E = \{2, 4, 6, 8, \ldots\}$$

or

$$E = \{x \in \mathbb{N} : x \text{ is divisible by } 2\}.$$

The important thing about sets is which elements they have or don't have. Therefore, we don't care what order the elements are listed in, and we don't care if an individual element is listed multiple times.

Operations on Sets: Using the set-builder notation makes it easy to define the simple operations on sets:

Definition 2 *Given two sets S and T, their union $S \cup T$ is defined by*

$$S \cup T = \{x : x \in S \textit{ or } x \in T\}.$$

Definition 3 *Given two sets S and T, their intersection $S \cap T$ is defined by*

$$S \cap T = \{x : x \in S \textit{ and } x \in T\}.$$

Relations between sets: Since the important thing about sets is which elements they contain, the relationship between sets that is probably the most used is the subset relation.

Definition 4 *If all the elements of a set S are also elements of a set T, then we say that S is a subset of T, and we write $S \subset T$.*

Sometimes it is obvious when one set is a subset of another set, but sometimes it isn't as clear. One strategy that can help in dealing with this is to think of set containment in terms of logic. The statement $S \subset T$ is equivalent to the statement

$$\text{if } x \in S, \text{ then } x \in T.$$

So, in order to show that the statement $S \subset T$ is true, we can show that no matter how we pick $x \in S$, the x we pick is also in T. In symbols this reads

$$x \in S \ \Rightarrow \ x \in T.$$

Logic As has been mentioned, this class is probably the first class where you will be expected to generate your own proofs. A proof is basically a string of logical implications that lead to a conclusion, so in writing proofs, it is very important to understand logic.

The basis of logic is the implication. An implication relates two statements, called the hypothesis and the conclusion. As the name suggests, the hypothesis is something that can be assumed to be true. The conclusion is the result that we are trying to derive once the hypothesis is assumed.

Be careful not to confuse the truth of the implication with the truth of the conclusion. The implication can be true even though the conclusion is not, such as, “When it rains, I bring an umbrella.” Just because I don’t have an umbrella today, doesn’t mean that the statement is false.

Here is a table that shows the truth of the statement $P \Rightarrow Q$ in relation to the logical value (true or false) of the statements P and Q.

P	Q	$P \Rightarrow Q$
False	False	True
False	True	True
True	False	False
True	True	True

Note that an implication is true unless the hypothesis being true can lead to the conclusion being false.

Example: Consider the statement

$$x > 2 \ \Rightarrow \ x^2 > 4.$$

I think we can all agree that this statement is true. Let’s take a look at the logic table:

x	Hypothesis	Conclusion
-3	False	True
-1	False	True
3	True	True

Example: Even though it's pretty clear, let's go through the logical steps in showing that $[2,3] \subset [0,4]$. We have to show that no matter how we choose a point in $[2,3]$, it is also in $[0,4]$. Let $x \in [2,3]$. Then $2 \leq x \leq 3$. But $0 < 2$ and $3 < 4$, so $0 < 3 \leq x \leq 3 < 4$, or $0 \leq x \leq 4$, which is the same as $x \in [0,4]$.

Special Sets: Some sets are used so often that they have a special notation associated with them. Some examples are the natural numbers

$$\mathbb{N} = \{1, 2, 3, 4, \ldots\},$$

the set of all positive whole numbers, the real numbers, denoted by $\mathbb{R}$, which is the set of all numbers that have an ordinary decimal representation, and the empty set $\emptyset = \{\}$, which has no elements.

Example: Suppose S is any set. Show that $\emptyset \subset S$.
Just like before, to show that this is true, we have to show that no matter how we choose $x \in \emptyset$, that choice of x is also in S. But since there are no elements in $\emptyset$, we are done.

Here are two examples illustrating strategies for proving that sets are equal (meaning that they share all the same elements):

Example: Show that $S \cap T = T \cap S$ for all sets S and T.
We have that

$$S \cap T = \{x : x \in S \text{ and } s \in T\} = \{x : x \in T \text{ and } x \in S\} = T \cap S.$$

In this example, we started with one side of the equality we were proving and changed the form until it looked like the other side.

Example: Prove that $S \cap \emptyset = \emptyset$ for any set S.
We have already shown that $\emptyset \subset S \cap \emptyset$. If we can show the reverse inclusion, then they must be the same set. Let $x \in S \cap \emptyset$. This means that $x \in S$ and $x \in \emptyset$, so $S \cap \emptyset \subset \emptyset$. The two conditions $\emptyset \subset S \cap \emptyset$ and $S \cap \emptyset \subset \emptyset$ together imply that $S \cap \emptyset = \emptyset$.

Quick Skip Ahead to Chapter 2

Systems of Linear Equations

We talked briefly last time about how answering fundamental questions about linear systems of equations (existence, uniqueness of a solution, etc.) is one of the main focuses of this class. We'll take some time now to talk in more details about these systems.

You probably have some experience working with systems of equations that have two or three variables. Most likely, you learned a method for solving such equations that involved eliminating variables until you knew the value of one variable, and then substituting the value of that variable back through the proceeding equations until you knew the value of all the variables.

We're going to develop this method carefully so that we can use it on arbitrarily large systems of equations. Also, in using this method, we'll be able to answer a lot of our questions about a given system, such as whether or not it has a solution and whether or not a solution is unique.

Let's begin with an example:

$$\begin{array}{l} x + y - z = -1 \\ 2x + 4y = -4 \\ -x + 3y + 2z = -6 \end{array}.$$

We say that equations such as this are linear because the only operations that are applied to the variables are multiplication by scalars and addition of other such terms. Of course, our goal here is to find values of x, y and z that satisfy the three equations simultaneously.

The set where we need to begin to look is called $\mathbb{R}^3$. In set-builder notation,

$$\mathbb{R}^3 = \{(x, y, z) : x,\ y,\ \text{and } z \in \mathbb{R}\},$$

or in other words, all ordered pairs of three real numbers. But we are looking for order pairs that satisfy our equations, so the **solution set** to our equations is

$$S = \{(x, y, z) \in \mathbb{R}^3 : x + y - z = -1,\ 2x + 4y = -4,\ -x + 3y + 2z = -6\}.$$

Any element of S will be a solution to our system. To check if an element of $\mathbb{R}^3$ is in S, we simply substitute the x, y, and z coordinates of the candidate solution into the left-hand side of the equations, and see if the result matches the right-hand side.

Example: Is $(-4, 1, -2) \in S$?
We have

$$(-4) + (1) - (-2) = -1 \checkmark$$

$$2(-4) + 4(1) = -4 \checkmark$$
$$-(-4) + 3(1) + 2(-2) = 3$$

It is easy to find elements of $\mathbb{R}^3$ that are not in S. But how do we find out if there are any elements in S? The answer is to simplify the system so it becomes more clear just what a solution looks like. We can do anything we want to to simplify this system, as long as we maintain the equalities. What this means is that if we change one side of an equation, we must change the other side as well.

For example, we can replace the third equation with the sum of the first and the third equations, to get the new system

$$\begin{array}{l} x + y - z = -1 \\ 2x + 4y = -4 \\ 4y + z = -7 \end{array} .$$

We have changed the form of the problem, but we haven't changed the solution set S for the problem. Notice that we eliminated x from the last equation. We can continue to simplify by adding -2 times the first equation to the second to get

$$\begin{array}{l} x + y - z = -1 \\ 2y + 2z = -2 \\ 4y + z = -7 \end{array} \xrightarrow{R2=-\frac{1}{2}R2} \begin{array}{l} x + y - z = -1 \\ y + z = -1 \\ 4y + z = -7 \end{array} \xrightarrow{R3=-4R2+R3} \begin{array}{l} x + y - z = -1 \\ y + z = -1 \\ -3z = -3 \end{array}$$

$$\xrightarrow{R3=\frac{1}{3}R3} \begin{array}{l} x + y - z = -1 \\ y + z = -1 \\ z = 1 \end{array} \xrightarrow{R2=R2-R3,\ R1=R1+R3} \begin{array}{l} x + y = 0 \\ y = -2 \\ z = 1 \end{array} \xrightarrow{R1=R1-R2} \begin{array}{l} x = 2 \\ y = -2 \\ z = 1 \end{array}$$

This process of simplifying the equations has shown us that $S = \{(2, -2, 1)\}$. Because we didn't change the equalities as we simplified the equations, we did not change S. The equations we started with and the simple ones we ended with have the same solution.

Also notice that there is no real reason that we need to write out the variables each time we take a step to simplify the equations. All the variables really do is hold our place until we have solved the system. Each variable corresponds to it's own column in the equations. Why not let each variable have it's own column in a matrix? Our original system could be represented as

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

The vertical line here represents the equals signs in the original system of equations. The matrix on the left of the vertical line is the **coefficient matrix** for the system. The vector on the right of the vertical line is the same as the right-hand sides of the original equations. Now, instead of writing out the whole system each time we make a simplification, we can just work with this **augmented matrix**.

We can solve the system again with the following steps:

$$\left(\begin{array}{ccc|c} 1 & 1 & -1 & -1 \\ 2 & 4 & 0 & 4 \\ -1 & 3 & 2 & -6 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & -1 & -1 \\ 0 & 2 & 2 & -2 \\ 0 & 4 & 1 & -7 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|c} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & -3 & -3 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1 \end{array}\right).$$

Since we're working with matrices now, let's take a moment out to establish some standard notation. A matrix A is said to be $m \times n$ if it has m rows and n columns. We locate the elements of a matrix by the row and column in which they are located. If we are interested in the entries of a matrix A, then we denote the (i, j) entry (the entry in the ith row and jth column) by a_{ij}.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m-1,1} & a_{m-1,2} & \cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m1} & a_{m2} & \cdots & a_{m,n-1} & a_{mn} \end{pmatrix}.$$

The operations that we are can perform that will preserve the equalities in a system of equations are called elementary row operations on the augmented matrix.

Definition 5 *An elementary row operation on an $m \times n$ matrix A is any of the following operations:*
1. Interchange rows r and s of A. That is, replace $a_{r1}, a_{r2}, \ldots, a_{rn}$ by $a_{s1}, a_{s2}, \ldots, a_{sn}$ and vice-versa.
2. Multiply row r of A by $c \neq 0$. That is, replace $a_{r1}, a_{r2}, \ldots, a_{rn}$ by $ca_{r1}, ca_{r2}, \ldots, ca_{rn}$.
3. Add d times row r of A to row s of A, where $r \neq s$. That is, replace $a_{s1}, a_{s2}, \ldots, a_{sn}$ by $a_{s1} + da_{r1}, a_{s2} + da_{r2}, \ldots, a_{sn} + da_{rn}$.

We have seen that performing elementary row operations on a matrix does not change the solution set of the system it represents. For this reason, if we can get from one matrix to another by row operations, we call the two matrices are called row equivalent.

Definition 6 *An $m \times n$ matrix A is said to be row equivalent to an $m \times n$ matrix B if B can be obtained by applying a finite sequence of elementary row operations to A.*

Now we know how to apply row operations, but what is our goal? What did we do in the example above to simplify the system?

Definition 7 *An $m \times n$ matrix is said to be in reduced row echelon form when it satisfies the following properties:*
1. All rows consisting entirely of zeros are at the bottom of the matrix.
2. Reading from left to right, the first nonzero entry in each row that does not consist entirely of zeros is a 1, called the leading entry of its row.
3. If rows i and $i+1$ are two successive rows that do not consist entirely of zeros, then the leading entry of row $i+1$ is to the right of the leading entry of row i.
4. If a column contains a leading entry of some row, then all other entries in that column are zero.

Note that there don't *have* to be any rows that are all zeros. But if there are, they must all be at the bottom of a matrix in reduced row echelon form.

The reason that this definition exists is to tell us when we are done with the row reduction process. When a matrix is in reduced row echelon form, then it is (usually) easy to see what the solution to the corresponding system of equations is.

So our approach to solving a system of equations is:

- Form the augmented matrix.
- Row reduce the augmented matrix to reduced row echelon form.
- Determine the solution(s) to the system (or if there are any).

This process is usually called Gaussian elimination or Gauss-Jordan reduction.

Example Find the solution to the linear system

$$\begin{aligned} x + 2y + 3z &= 9 \\ 2x - y + z &= 8 \\ 3x - z &= 3 \end{aligned} .$$

The augmented matrix reduces as

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 2 & -1 & 1 & 8 \\ 3 & 0 & -1 & 3 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & -5 & -5 & 10 \\ 0 & -6 & -10 & -24 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 1 & 5 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -4 & -12 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{array}\right).$$

Thus, the solution to the system is $x = 2$, $y = -1$, and $z = 3$.

Example Find the solution to the linear system

$$\begin{aligned} x + y + 2z - 5w &= 3 \\ 2x + 5y - z - 9w &= -3 \\ 2x + y - z + 3w &= -11 \\ x - 3y + 2z + 7w &= -5 \end{aligned}$$

We again row reduce the augmented matrix:

$$\left(\begin{array}{cccc|c} 1 & 1 & 2 & -5 & 3 \\ 2 & 5 & -1 & -9 & -3 \\ 2 & 1 & -1 & 3 & -11 \\ 1 & -3 & 2 & 7 & -5 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 1 & 2 & -5 & 3 \\ 0 & 3 & -5 & 1 & -9 \\ 0 & -1 & -5 & 13 & -17 \\ 0 & -4 & 0 & 12 & -8 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|c} 1 & 0 & -3 & 8 & -14 \\ 0 & 1 & 5 & -13 & 17 \\ 0 & 0 & -20 & 40 & -60 \\ 0 & 0 & 20 & -40 & 60 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 2 & -5 \\ 0 & 1 & 0 & -3 & 2 \\ 0 & 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Notice that this time we ended up with a row of zeros (located at the bottom of the matrix, as required). If we again interpret this augmented matrix as a system of equations, then the last equation is saying that $0 = 0$, which is hopefully obvious. This means that one of our equations is unnecessary, in the sense that the information it contains is already contained in the other equations. So what we really have is four unknowns, but only three equations.

In terms of the row-reduced matrix, there is a leading one in every column but the last. We can think of a variable having a leading one in it's column as that variable being fixed once values have been chosen for the rest of the variables. In this example, there is not a leading one in the column for w, so it can take on any value. This means that this system has infinitely many solutions.

We write our solution as

$$x = -2w - 5,\ y = 3w + 2,\ z = 2w + 3,$$

where w can be any real number that we choose.

Example: Find the solution to the system

$$\begin{array}{l} x + 2y + 3z + 4w = 5 \\ x + 3y + 5z + 7w = 11 \\ x - z - 2w = -6 \end{array}.$$

$$\left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 7 & 11 \\ 1 & 0 & -1 & -2 & -6 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 2 & 3 & 6 \\ 0 & -2 & -4 & -6 & -11 \end{array}\right)$$
$$\sim \left(\begin{array}{cccc|c} 1 & 0 & -1 & -2 & -7 \\ 0 & 1 & 2 & 3 & 6 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

This time there is a row of zeros in the coefficient matrix, but there is a one on the other side of the vertical line. If we interpret this as and equation, it says that $0 = 1$. This is impossible, so we are forced to conclude that the system of equations has no solutions.

A system of equations that has no solution is called **inconsistent**. If the system has one or more solutions, then it is called **consistent**. We can recognize these different solutions by looking at the reduced row-echelon form of the augmented matrix. If there is a row that is all zeros except the last element, which is not zero, then the system is inconsistent. If the augmented matrix in RREF has one or more rows that consist entirely of zeros, then the system is consistent, but it may have infinitely many solutions. If the augmented matrix in RREF has a leading one in every column, there is a unique solution to the system.

Back to Chapter 1

We have been talking about solution sets, so now is the time to develop some ideas about what the solution sets of some equations look like.

Vectors: You've probably had some previous experience in dealing with vectors. In physics vectors are defined to be a quantity that has both a magnitude and a direction. It makes sense to add them together and to scale them. Graphically, vectors from physics are often represented by a directed line segment. Vectors can be added graphically "tip to tail" and scaled by changing their length by the appropriate factor.

Our approach here is to define sets of vectors that encompass these examples, but also examples that show up elsewhere in mathematics, such as in the study of differential equations. Algebraically speaking, the thing that is important about the vectors from physics is the way in which they are added and scaled. We have a good idea of what it means to perform these operations in that setting, but what if instead of directed line segments, the vectors that we use are something different? Perhaps in that situation we would also need to change what we mean by addition and scaling.

Because of this, we're not going to be very specific about what we mean by addition and scalar multiplication just yet, but we are going to require that these operations follow some rules.

Definition 8 *A vector space is a set V with operations of addition and scalar multiplication. The elements of V are called vectors. The operation of addition combines any two vectors $v \in V$ and $w \in V$ to produce another vector in V denoted $v + w$. The operation of scalar multiplication combines any real number $r \in \mathbb{R}$ and any vector $v \in V$ to produce another vector in V denoted rv. A real number used in this operation is called a scalar. These operations must satisfy the following eight axioms for all $v, w, x \in V$ and $r, s \in \mathbb{R}$.*
1. $v + w = w + v$ (commutative law of addition)
2. $(v + w) + x = v + (w + x)$ (associative law of addition)
3. There is a vector in V, denoted 0, such that $v + 0 = v$. (additive identity law)
4. For each vector v in V, there is a vector in V, denoted $-v$, such that $v+(-v) = 0$. (additive inverse law)
5. $r(v + w) = rv + rw$. (distributive law)
6. $(r + s)v = rv + sv$. (distributive law)
7. $r(sv) = (rs)v$. (associative law of multiplication)
8. $1v = v$. (scalar identity law)

What we are doing here is adding a level of abstraction. We are allowing the vectors to be more general objects, but we are requiring familiar rules of algebra to continue to hold in that more abstract setting.

This is a good place to point out an important part of this definition: both the operations of vector addition and scalar multiplication must produce another element of the vector space. We say that a vector space is closed under the operations of vector addition and scalar multiplication.

There are lots of easy examples of sets that are closed under the usual definitions of addition and multiplication, such as the natural numbers, rational numbers, integers, and real numbers. However, the real numbers are not closed under the operation of division, since we can't divide by zero.

If you think that a set is closed under a certain operation (such as addition), then to prove that this is true, you need to start with an arbitrary element (or elements, depending on the operation in question) of the set, and show that the result of the operation is an element of the set.

If you think that a set is not closed under an operation, you need only find one example of element(s) such that when the operation is applied, the result is not an element of the original set.

Example: Are the sets $[0, 1]$ and $\{0, 1, 2\}$ closed under normal multiplication?
First we have to decide whether we think the answer is yes or no, and then we have to come up with a proof of our assertion. In the case of $[0, 1]$, when you multiply together two of its elements, you get a result that is smaller than either of the original numbers, but still positive (unless one of the factors was zero). So it should still be in $[0, 1]$.

Let's write this out as a proof: Let $x_1, x_2 \in [0, 1]$. Then $x_1x_2 \geq 0$ because x_1 and x_2 are in $[0, 1]$, and $x_1x_2 \leq 1 \cdot x_2 \leq 1$, so $x_1x_2 \in [0, 1]$. This implies that $[0, 1]$ is closed under multiplication.

How about $\{0, 1, 2\}$? A little experimentation yields a counterexample, which is also our proof: The set $\{0, 1, 2\}$ is not closed under multiplication because $2 \cdot 2 = 4 \notin \{0, 1, 2\}$.

Properties of Vector Spaces: We already know about the eight properties that all vector spaces must have by definition. But are there other properties that all vector spaces must have, which can be logically deduced from the definition? The answer is yes, and we'll work out a few here.

Theorem 1 *Suppose V is a vector space. If v is any element of V, then $0 + v = v$.*

This might seem like a no-brainer, but there is one step involved.
Proof : One of our assumptions is that $v + 0 = v$. That is almost what we want,

but it is in the wrong order. However, another of our assumptions tells us that $v + w = w + v$ for all choices of v and w in V. Therefore, $v + 0 = 0 + v = v$. $\square$

Theorem 2 *Suppose V is a vector space. For any $v \in V$ there is only one vector $-v$ in V that satisfies the additive inverse condition of axiom 4 (i.e., such that v+(-v)=0).*

Proof : This is a proof of uniqueness, and as such, a good strategy is to assume the opposite and try to derive a contradiction. If assuming that the additive inverse is not unique leads us to a conclusion that is clearly nonsense, then we know that our assumption must be wrong, and therefore it must be unique.

Assume that there are two different vectors, x and y, such that

$$v + x = 0 \text{ and } v + y = 0.$$

Let's look at the first of these equations, and add y to both sides. Then

$$v + y + x = y \;\Rightarrow\; 0 + x = y \;\Rightarrow\; x = y.$$

Obviously, our original assumption that the vectors were different is wrong. $\square$

Theorem 3 *Suppose V is a vector space. If v is any element of V, then $0v = 0$.*

Proof : Here are two ways to accomplish this proof.

$$\begin{gathered} 0v = (0+0)v = 0v + 0v \\ \Rightarrow\; 0v + (-0v) = (0v + 0v) + (-0v) \\ \Rightarrow\; 0 = 0v + (0v + (-0v)) \\ \Rightarrow\; 0 = 0v \end{gathered}$$

OR

$$\begin{gathered} 0v = 0v + 0 = 0v + (0v + (-0v)) = (0v + 0v) + (-0v) \\ = (0+0)v + (-0v) = 0v + (-0v) = 0. \end{gathered}$$

$\square$

Here are some other useful properties. You will be asked to prove some of them as exercises, but you might want to prove them all for yourself, just for practice.

Theorem 4 *Suppose v is any element of a vector space V, and $r, s \in \mathbb{R}$. The following results hold:*
1. -v+v=0.
2. r0=0.
3. If $rv = 0$, then $r = 0$ or $v = 0$.
4. $(-1)v = -v$.
5. If $-v = v$, then $v = 0$.
6. $-(-v) = v$.
7. $(-rv) = -(rv)$.
8. $r(-v) = -(rv)$.
9. if $v \neq 0$ and $rv = sv$, then $r = s$.

Subtraction and Cancellation Since we have defined vector addition and additive inverses for vectors, we have enough machinery in place to define subtraction.

Definition 9 *let V be a vector space. The operation of subtraction combines two vectors $v \in V$ and $w \in V$ to produce a vector denoted $v - w$ and defined by the formula*

$$v - w = v + (-w).$$

Theorem 5 *Suppose v, w and x are vectors in a vector space V. If $v + x = w + x$, then $v = w$.*

Proof : If $v + x = w + x$, then $v + x + (-x) = (w + x) + (-x)$, which we can rearrange as $v + (x + (-x)) = w + (x + (-x))$, and since $x + (-x) = 0$, we must have that $v = w$.

The Euclidean Spaces

When we talked about solving linear equations, we used the space

$$\mathbb{R}^3 = \{(x, y, z) : x, y, z \in \mathbb{R}\}.$$

In particular, we used this set because the system we were solving at the time had three unknowns. Of course, we might want to solve systems with more than three unknowns, and in this case it makes sense to use vectors with more components.

Definition 10 *The set $\mathbb{R}^n$ is defined as*

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{R}\}$$

for each natural number n.

While it is useful as a set, it would be much more useful to have a vector space. For this reason, we need define the operations of vector addition and scalar multiplication on $\mathbb{R}^n$. Since the entries of the vectors in $\mathbb{R}^n$ are just real numbers, the best way to define these operations is in terms of addition and multiplication of real numbers. That way we can be sure that things like associativity, distributivity, etc. will work like we need them to.

Definition 11 *If $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$ are elements of $\mathbb{R}^n$, we define their sum by*

$$v + w = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n).$$

If $c \in \mathbb{R}$, then we define $cv \in \mathbb{R}^n$ by

$$cv = (cv_1, cv_2, \ldots, cv_n).$$

With these operations defined, $\mathbb{R}^n$ is a vector space. Let's verify a few of the requirements, just to make sure. You may want to verify them all for yourself, just for practice.

Example: Show that addition in $\mathbb{R}^n$ is commutative.
Remember that commutative means that we get the same answer no matter what order we add in. Let $v, w \in \mathbb{R}^n$. Then

$$v + w = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix} = \begin{pmatrix} w_1 + v_1 \\ w_2 + v_2 \\ \vdots \\ w_n + v_n \end{pmatrix} = w + v.$$

We were able to switch the order of addition of the elements because we already know that addition of real numbers is commutative.

For the purposes of exploring the vector space $\mathbb{R}^n$, we don't care if the vectors are represented as rows or as columns. They are the same. However, when we start having vectors and matrices interact via matrix multiplication, then we'll have to be careful whether we're talking about row or column vectors.

Example: Show that each element of $\mathbb{R}^n$ has an additive inverse.
Let $v \in \mathbb{R}^n$. Then

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} -v_1 \\ -v_2 \\ \vdots \\ -v_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In other words, $(-v) = -v$. Notice that the additive identity (zero vector) in $\mathbb{R}^n$ is the one that has all zero entries.

Example: Show that the distributive property $r(v+w) = rv+rw$ holds for $v, w \in \mathbb{R}^n$ and $r \in R$.
We have that

$$r(v+w) = r\begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix} = \begin{pmatrix} r(v_1 + w_1) \\ r(v_2 + w_2) \\ \vdots \\ r(v_n + w_n) \end{pmatrix}$$

$$= \begin{pmatrix} rv_1 + rw_1 \\ rv_2 + rw_2 \\ \vdots \\ rv_n + rw_n \end{pmatrix} = \begin{pmatrix} rv_1 \\ rv_2 \\ \vdots \\ rv_n \end{pmatrix} + \begin{pmatrix} rw_1 \\ rw_2 \\ \vdots \\ rw_n \end{pmatrix} = rv + rw.$$

More on Matrices First, we need to define the sets of matrices that we will talk about.

Definition 12 *Let m and n be positive integers. We denote by $\mathbb{M}(m, n)$ the set of all rectangular arrays of real numbers with m horizontal rows and n vertical columns.*

Since we are able to add vectors and multiply them by scalars, it makes sense that we should be able to perform these operations on matrices as well, since in some sense we can think of matrices as vectors that have been "reshaped."

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \text{ and } B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

be elements of $\mathbb{M}(m, n)$. Then we define their sum by

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}.$$

That is, we add matrices element-wise. Because of this, matrices must have the same size in order to add them. Also, if $r \in \mathbb{R}$ is a scalar, then

$$rA = \begin{pmatrix} ra_{11} & ra_{12} & \cdots & ra_{1n} \\ ra_{21} & ra_{22} & \cdots & ra_{2n} \\ \vdots & \vdots & & \vdots \\ ra_{m1} & ra_{m2} & \cdots & ra_{mn} \end{pmatrix}.$$

Sometimes it is nice to use a shorthand when referring to matrices, especially since the algebraic operations that we have just defined for matrices are applied element-wise. We can refer to a matrix A by its elements a_{ij} by writing $A = [a_{ij}]$. If we add two matrices A and B, then we can write this as

$$[a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}].$$

It shouldn't be a surprise that $\mathbb{M}(m, n)$ is a vector space when endowed with these operations.

Example Show that if $A, B \in \mathbb{M}(m, n)$ and $r \in \mathbb{R}$, then $r(A + B) = rA + rB$. Using our shorthand notation, this is straight forward

$$r(A + B) = r[a_{ij} + b_{ij}] = [r(a_{ij} + rb_{ij})] = [ra_{ij} + rb_{ij}] = [ra_{ij}] + [rb_{ij}] = rA + rB.$$

Function Spaces So far we have talked about vector spaces whose elements are either rows, columns, or arrays of real numbers. In the case of $\mathbb{R}^n$, we can think of each element as a function that takes an integer (the index) as input, and outputs a real number. So it shouldn't be surprising that we can also define vector spaces whose elements are functions.

Definition 13 *A real-valued function f defined on a set X is a rule that assigns to each element $x \in X$ a unique and unambiguous real number, denoted $f(x)$.*

If $X = \mathbb{R}$, then we are just talking about the usual functions that you encountered in previous algebra and calculus classes. If $X = \mathbb{R}^n$, then we are talking about a function of several variables, such as those studied in Calculus III.

A function that is defined on a set X and takes values in a set Y is usually denoted by $f : X \to Y$. In this case, the set X is called the domain of X, and Y is called the range. The set of all $y \in Y$ that can be realized as $f(x)$ for some $x \in X$ is called the image of the function f.

Definition 14 *Let $\mathbb{F}(X)$ be se set of all possible real-valued functions defined on the nonempty set X.*

We will think of the functions in this set (which is very large) as the elements of a vector space. Therefore, we will need to make sure that we understand how to add together functions and multiply them by scalars in such a way that the requirements in the definition of a vector space will hold.

One thing to be careful of is not to confuse the function f with its values $f(x)$. The function f is an element of $\mathbb{F}(X)$, while $f(x)$ is a real number. To know what f is, we have to know *all* of its values. For two functions $f, g \in \mathbb{F}(X)$ to be equal means that they agree at every point in X, i.e., $f(x) = g(x)$ for all $x \in X$.

Definition 15 *Let X be a nonempty set. The sum of two functions $f, g \in \mathbb{F}(X)$ is the function $(f + g) \in \mathbb{F}(X)$ defined by*

$$(f + g)(x) = f(x) + g(x) \text{ for all } x \in X.$$

If $r \in \mathbb{R}$ is a scalar, then the function $rf \in \mathbb{F}(X)$ is defined by

$$(rf)(x) = r \cdot f(x) \text{ for all } x \in X.$$

There should be nothing new about these definitions. We are just adding functions pointwise, just like you should be used to doing from Calculus. Also, scalar multiplication is done pointwise.

Thinking about what function addition looks like graphically depends a lot on what the functions involved are. Scalar multiplication is a little more straightforward, though. When you multiply a function by a scalar, the graph of the function is stretched toward or away from the x-axis. If the scalar is negative, then the function is also flipped around the x-axis.

Subspaces: Remember that a vector space is a set which has operations of vector addition and scalar multiplication that satisfy certain properties. When we talked about sets in general, the important relation between them was the subset relation. Therefore, we can talk about subsets of a vector space. But since a vector space is more than just a set, we are most interested in subsets that inherit the properties of a subspace as well.

Definition 16 *A subspace of a vector space V is a subset $S \subset V$ such that when the addition and scalar multiplication of V are used to add and scalar multiply the elements of S, then S is a vector space.*

Example: Consider the vector space $\mathbb{R}^3$. We know that another vector space is $\mathbb{R}^2$, which is a plane. In fact, $\mathbb{R}^2$ is, for all intents and purposes, the same as the set

$$S = \{(x_1, x_2, 0) : x_1, x_2 \in \mathbb{R}\},$$

which is a subset of $\mathbb{R}^3$. Since it is so closely related to $\mathbb{R}^2$, it seems reasonable that this might be a subspace of $\mathbb{R}^3$. Let's verify this.

According to the definition, all we have to do to check that a subset of a vector space is a subspace is verify that it is closed under vector addition and scalar multiplication. All the other properties of a vector space are inherited from the larger vector space.

To check that S is closed under scalar multiplication, we choose any two elements of S, say v and w. Since, they are in S, we know that $v_3 = w_3 = 0$. Therefore,

$$v + w = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ 0 \end{pmatrix} \in S.$$

To check that S is closed under scalar multiplication, we choose $r \in \mathbb{R}$ and again let v be an arbitrary element of S. Then

$$rv = \begin{pmatrix} rv_1 \\ rv_2 \\ 0 \end{pmatrix} \in S.$$

Since S is closed under vector addition and scalar multiplication, S is a subspace of the vector space $\mathbb{R}^3$.

This is a general method for determining that a subset of a vector space is a subspace, so we'll write it down as a theorem:

Theorem 6 *Suppose a subset S of a vector space V satisfies the following conditions:*
1. S is nonempty.
2. S is closed under addition.
3. S is closed under scalar multiplication.
Then S is a subspace of V.

Proof : Since the identities in the axioms for a vector space hold for all elements of V, they must hold when the elements come form S. That leaves only a couple of things to check. First, we have to check that $0 \in S$. Since S is nonempty, we can choose a vector $v \in S$. Since S is closed under scalar multiplication, we have that $0v = 0 \in S$. The other thing that we need to check is that the elements of S have additive inverses. If $v \in S$, then so is $-1v$, since it is closed under scalar multiplication. We showed previously that $-1v = -v$, so v has an additive inverse in S.

Example: Show that the set of points on the plane $3x + 2y - z = 0$ is a subspace of $\mathbb{R}^3$.
Let $S = \{(x, y, z) : 3x+2y-z = 0\}$. Obviously S is not empty, since $\vec{0} \in S$. Suppose that $\vec{v}, \vec{w} \in S$. Then

$$3v_1 + 2v_2 - v_3 = 0 \text{ and } 3w_1 + 2w_2 - w_3 = 0.$$

To show that S is closed under vector addition, we must show that

$$3(v_1 + w_1) + 2(v_2 + w_2) - (v_3 + w_3) = 0.$$

This is true, because

$$3(v_1 + w_1) + 2(v_2 + w_2) - (v_3 + w_3) = (3v_1 + 2v_2 - v_3) + (3w_1 + 2w_2 - w_3) = 0 + 0 = 0.$$

Also, S is closed under scalar multiplication, since if $r \in \mathbb{R}$,

$$3(rv_1) + 2(rv_2) - (rv_3) = r(3v_1 + 2v_2 - v_3) = r0 = 0.$$

Example: Show that the set

$$S = \{A \in \mathbb{M}(2, 2) : a_{12} = a_{21} = 0\}$$

is a subspace of $\mathcal{M}(2, 2)$.
S is nonempty, since the 2×2 zero matrix is in S. If we take $A, B \in S$, then

$$A + B = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & 0 \\ 0 & a_{22} + b_{22} \end{pmatrix}.$$

Since the $1, 2$ and $2, 1$ elements are zero, $A + B \in S$. Also, if $r \in \mathbb{R}$, then

$$rA = \begin{pmatrix} ra_{11} & 0 \\ 0 & ra_{22} \end{pmatrix},$$

which is also in S. Therefore, by the subspace theorem, S is a subspace of $\mathbb{M}(2, 2)$.

Example: Show that the set $\mathbb{P}_2$ of all polynomials of degree 2 or lower forms a subspace of $\mathbb{F}(\mathbb{R})$.
Obviously $P_2(x)$ is nonempty. Let $p, q \in \mathbb{P}_2$, say $p(x) = p_1x^2 + p_2x + p_3$ and $q(x) = q_1x^2 + q_2x + q_3$. Then

$$(p + q)(x) = (p_1 + q_1)x^2 + (p_2 + q_2)x + p_3 + q_3 \in \mathbb{P}_2.$$

Also, if $r \in \mathbb{R}$,

$$(rp)(x) = rp_1x^2 + rp_2x + rp_3 \in \mathbb{P}_2.$$

Since $\mathbb{P}_2$ is closed under vector addition and scalar multiplication, it is a subspace of $\mathbb{F}(\mathbb{R})$.

Example: Suppose that $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ are elements of a vector space V. Show that the set

$$S = \{c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n : c_1, c_2, \ldots, c_n \in \mathbb{R}\}$$

is a subspace of V.
First, S is clearly not empty, since it contains all the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. Let $\vec{u}, \vec{w} \in S$. Then there exist scalars $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ such that

$$\vec{u} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n$$

and

$$\vec{w} = b_1\vec{v}_1 + b_2\vec{v}_2 + \cdots + b_n\vec{v}_n.$$

It follows that

$$\vec{u} + \vec{w} = (a_1 + b_1)\vec{v}_1 + (a_2 + b_2)\vec{v}_2 + \cdots + (a_n + b_n)\vec{v}_n \in S.$$

Also, if $r \in \mathbb{R}$, then

$$ru = ra_1\vec{v}_1 + ra_2\vec{v}_2 + \cdots + ra_n\vec{v}_n \in S.$$

Therefore, by the subspace theorem, S is a subspace of V.

Here are a few useful subspaces of $\mathbb{F}(X)$.

$$C^{(n)}(X) = \{f \in \mathbb{F}(X) : f \text{ is continuous and has } n \text{ continuous derivatives on } X\}$$

In particular, $C^0(X) = C(X)$ is the set of continuous functions on X.

$$\mathbb{P}_n = \{f \in \mathbb{F}(\mathbb{R}) : f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0\}$$

$$\mathbb{P} = \{f \in \mathbb{F}(\mathbb{R}) : f \text{ is a polynomial of any degree}\}$$

Lines and Planes: Since we have described $\mathbb{R}^3$ as a set of vectors, we can use the idea of vectors to describe the lines and planes in $\mathbb{R}^2$ and $\mathbb{R}^3$.

As we have mentioned before, vectors can be added graphically. We add them tip to tail. When we add two vectors together and connect the tip of the sum with the tip of the two original vectors with line segments, we end up with a parallelogram. This is what is known as the parallelogram law for vector addition.

Scalar multiplication also has a geometric interpretation. If $\vec{v}$ is a vector in $\mathbb{R}^2$ and $r > 1$ is a scalar, then $r\vec{v}$ is a stretched version of $\vec{v}$ that is exactly r times longer. If $0 < r < 1$, then $\vec{v}$ is shortened by a factor r. If r is negative, then $\vec{v}$ is reflected through the origin before it is scaled.

The same geometric interpretations of vector addition and scalar multiplication hold in $\mathbb{R}^3$. We can use these ideas and our intuition from experience working with vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ to get a feeling for how vectors work in $\mathbb{R}^n$ for $n > 3$, where we are unable to visualize things.

The points that lie on the tips of two vectors $\vec{v}$ and $\vec{w}$ always lie on a plane through the origin, and the sum of those two vectors can be used to form a parallelogram, just as in $\mathbb{R}^2$. Similarly, scalar multiples of a vector in higher dimensions can still be thought of as lying on a line through the origin and the point at the tip of the original vector.

Choose any vector $\vec{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$. As we multiply this vector by different scalars, the tips of the resulting vectors trace out a line through the origin that is parallel to $\vec{v}$. In fact, we can describe this line as the set

$$\{r\vec{v} : r \in \mathbb{R}\}.$$

We can use this idea of describing a line as a set of multiples of a given vector in higher dimensions as well, even when it doesn't make sense to "draw" the line. For example, we can think of the set

$$\{r \sin x : r \in \mathbb{R}\}$$

as the line parallel to $\sin(x)$ in $\mathbb{F}(\mathbb{R})$.

Next, let's consider two vectors $\vec{e}_1 = (1, 0)$ and $\vec{e}_2 = (0, 1)$ from $\mathbb{R}^2$. Now, consider another vector with components (r, s). Then we have that

$$r\vec{e}_1 + s\vec{e}_2 = r(1, 0) + s(0, 1) = (r, 0) + (0, s) = (r, s).$$

Thus, we can "build" any vector in $\mathbb{R}^2$ from the two vectors $\vec{e}_1$ and $\vec{e}_2$. Another way of saying this is that $\vec{e}_1$ and $\vec{e}_2$ generate $\mathbb{R}^2$.

The vectors $\vec{e}_1$ and $\vec{e}_2$ are easy to use to generate $\mathbb{R}^2$ because it is easy to see what the choice of scalars should be that generates a given vector. But we could use other pairs of vectors just as well. Basically, using different vectors is the same as using different axes. Instead of using a vertical and a horizontal axis, we use two axes that are not at a right angle. This will work as long as the vectors that we choose are not paralell.

Since we can generate the entire plane that is $\mathbb{R}^2$ by taking scalar multiples of two vectors and adding them, we can do the same thing in $\mathbb{R}^3$. Any two vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$ that are not in the same direction will generate a plane through the origin, which can be described as the set

$$\{r\vec{v} + s\vec{w} : r, s \in \mathbb{R}\}.$$

Since $\vec{v}$, $\vec{w}$, and $\vec{0}$ are all in this set, it is the plane that passes through those three points.

What if we are interested in a line or a plane that does not pass though the origin? Whatever that line or plane is, we can think of it as a line/plane through the origin that has been shifted. In other words, we add a single nonzero vector $\vec{x}$ to all the vectors that lie on the line or plane. For example, the line through the point $\vec{x}$ that is parallel to the vector $\vec{v}$ is the set

$$\{r\vec{v} + \vec{x} : r \in \mathbb{R}\}.$$

Similarly, a plane can be translated in the same way. When we translate the plane $\{r\vec{v} + s\vec{w} : r, s \in \mathbb{R}\}$ by the vector $\vec{x}$, it becomes the plane

$$\{r\vec{v} + s\vec{w} + x : r, s \in \mathbb{R}\}.$$

Example: Consider the plane that passes through the points $(1, 2, 3)$, $(0, 2, 1)$, and $(3, -1, 1)$. We can choose any of these vectors to be the translation vector, so let $\vec{x} = (1, 2, 3)$. Then we need to figure out what the $\vec{v}$ and $\vec{w}$ are for this case, so we subtract $\vec{x}$ from the other two vectors, to get $\vec{v} = (0, 2, 1) - \vec{x} = (-1, 0 - 2)$ and $\vec{w} = (3, -1, 1) - \vec{x} = (2, -2 - 2)$. Therefore, the plane in question can be described as

$$\{r\vec{v} + s\vec{w} + x : r, s \in \mathbb{R}\}.$$

Notice that we get the points that we started with if we make the choices $r = s = 0$, $r = 1,\ s = 0$, and $r = 0,\ s = 1$.

Definition 17 *Suppose V is a vector space. Let $\vec{v}$ be a nonzero vector in V. The line through a vector $\vec{x} \in V$ in the direction $\vec{v}$ is the set $\{rv + x : r \in \mathbb{R}\}$. If $\vec{v}$, $\vec{w}$, and $\vec{x}$ are three vectors in V with neither $\vec{v}$ nor $\vec{w}$ being a multiple of the other, then the plane through $\vec{x}$ in the directions $\vec{v}$ and $\vec{w}$ is the set $\{r\vec{v} + s\vec{w} + x : r, s \in \mathbb{R}\}$.*

Example: Write the set $\{(x, y, z) \in \mathbb{R}^3 : 4x + 2y + 2z = 6\}$ as a plane according to the definition above.
To do this, we have to determine what the vectors $\vec{x}$, $\vec{v}$ and $\vec{w}$ should be. We can reduce this problem to one that we already solved if we can find three points that are in this plane. It is easy to verify that the points

$$(0, 3, 0), \quad (0, 0, 3), \quad \text{and} \quad (1/2, 1, 1)$$

lie on the plane. We now choose one of these vectors to be the shift, say $\vec{x} = (0, 3, 0)$. Then

$$\vec{v} = (0, 0, 3) - (0, 3, 0) = (0, -3, 3) \quad \text{and} \quad \vec{w} = (1/2, 1, 1) - (0, 3, 0) = (1/2, -2, 1).$$

With these choices, we have that

$$\{(x, y, z) \in \mathbb{R}^3 : 4x + 2y + 2z = 6\} = \{r\vec{v} + s\vec{w} + \vec{x} : r, s \in \mathbb{R}\}.$$

Chapter 2: Systems of Linear Equations

Section 2.3: Solving Linear Systems

We have already discussed the solution of systems of linear equations at some length. We have talked about how to write a system of equations as an augmented matrix, and we have discussed how to row reduce the augmented matrix to reduced row echelon form. We also talked about the different possible outcomes of this process.

1. In RREF, the augmented matrix has an equation that says that $0 = 1$. This means that the system has no solution.
2. In RREF, the augmented matrix has leading ones in each column. In this case, the solution to the system of equations is unique.
3. The system is consistent, but there is not a leading one in each column corresponding to a variable in the RREF of the augmented matrix. In this case, the solution is not unique.

The variables whose columns do not contain a leading one are called free variables. The variables whose columns do contain a leading one are called leading variables.

Suppose that a system of equations has an augmented matrix whose RREF is

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

If the unknowns are x_1, x_2, x_3, x_4, x_5, then x_2 and x_5 are free variables and the rest are leading variables. Since x_2 and x_5 are free, we can choose any real value for them. Lets say $x_2 = r$ and $x_5 = s$, where r and s are real numbers. Then we can write the solution to our system as

$$\begin{aligned} x_1 &= -r + 2s + 4 \\ x_2 &= r \\ x_3 &= -5s - 1 \\ x_4 &= -3s \\ x_5 &= s \end{aligned} \quad .$$

If we write the solution as a vector, then we have

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -r + 2s + 4 \\ r \\ -5s - 1 \\ -3s \\ s \end{pmatrix} = r \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 0 \\ -5 \\ -3 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \\ -1 \\ 0 \\ 0 \end{pmatrix}.$$

In other words, the solution set for this system of equations is

$$S = \left\{ r \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 0 \\ -5 \\ -3 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \\ -1 \\ 0 \\ 0 \end{pmatrix} : r, s \in \mathbb{R} \right\}.$$

Notice that this is just like the set that is a plane in $\mathbb{R}^3$, except that the vectors are in $\mathbb{R}^5$. It is customary to call such a set a hyperplane, which is loosely defined to be a plane in dimensions higher than three.

A system of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{12}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned} \quad ,$$

where all the constants on the right-hand side of the equations are zero, is called a homogeneous equation.

The first thing that we should notice is that the zero vector $\vec{0}$ is *always* a solution of a homogeneous solution. For this reason, we call it the trivial solution. The question to answer for a given system of equations is whether or not the solution set to the homogeneous system contains more than the zero vector.

Notice that when we form the augmented matrix for a homogenous system, the rightmost column is all zeros. Also, notice that when we perform elementary row operations, that column of zeros will not change. Therefore, when we solve a homogeneous system, there is no reason to write that column of zeros. Just remember that it is there.

Example: Solve the homogeneous system

$$\begin{array}{l} x - 2y - z = 0 \\ 2x - y + 2z = 0 \\ x + y + 3z = 0 \end{array}\ .$$

The coefficient matrix row reduces as

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

From this, we can see that z is a free variable, so let's say $z = r$, where $r \in \mathbb{R}$ is arbitrary. Then the solution to the system is

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -\frac{5}{3}r \\ -\frac{4}{3}r \\ r \end{pmatrix} = r \begin{pmatrix} -\frac{5}{3} \\ -\frac{4}{3} \\ 1 \end{pmatrix}.$$

Theorem 7 *A homogeneous system of m linear equations in n unknowns with $n > m$ has at least one nontrivial solution.*

Proof : When a matrix is in RREF, there can only be at most one leading one in each row. Therefore, the number of columns containing leading ones must be less than or equal to the number of rows. Since in this case there are more columns than rows, there must be columns that do not contain leading ones. This means that there are free variables, so there are nontrivial solutions. $\square$

Applications: Here are a couple of applications where systems of linear equations must be solved.

Example: Determine a polynomial p whose graph passes through $(0,2)$ with a slope of -1 and through $(1,1)$ with a slope of 2.
Since there are four requirements, it makes sense that a polynomial of degree 3, having four unknown coefficients, should be able to fill the requirements. Therefore, let $p(x) = ax^3 + bx^2 + cx + d$ and $p'(x) = 3ax^2 + 2bx + c$. The first requirement implies that

$$d = 2,$$

and the second implies that

$$c = -1.$$

The requirements at the point $(1,1)$ imply that

$$a + b + c + d = 1 \quad \text{and} \quad 3a + 2b + c = 2.$$

The augmented matrix for this system is

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

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

Therefore $p(x) = 6x^3 - 3x^2 - x + 2$.

Example: A manufacturer makes three kinds of plastics: A, B, and C. Each kilogram of A produces 10 grams of particulate matter discharged into the air and 30 liters of liquid waste discharged into the river. Each kilogram of B produces 20 grams of particulate matter and 50 liters of liquid waste. Each kilogram of C produces 20 grams of particulate matter and 110 liters of liquid waste. The Environmental Protection agency limits the company to 2250 grams of particulate matter per day and 7950 liters of liquid waste discharge per day. Determine the production levels of the three plastics that will result in emission levels that reach the maximum values for both particulate matter and liquid waste.

Let x, y, and z be the number of kilograms produced of plastics A, B, and C, respectively. Then

$$10x + 20y + 20z = 2250 \quad \text{and} \quad 30x + 50y + 110z = 7950.$$

As an augmented system, this is

$$\left(\begin{array}{ccc|c} 10 & 20 & 20 & 2250 \\ 30 & 50 & 110 & 7950 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 2 & 2 & 225 \\ 0 & -10 & 50 & 1200 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 12 & 465 \\ 0 & 1 & -5 & -120 \end{array}\right).$$

Therefore, the solution is

$$\begin{array}{l} x = -12r + 465 \\ y = 5r - 120 \\ z = r \end{array}.$$

We have infinitely many solutions, but only some of them make sense. We have to ensure that the values of all the variables are nonnegative. For x to be nonnegative we must have

$$r \leq 38.75,$$

and for y to be nonnegative, it must be true that

$$r \geq 24.$$

Therefore, the solution set for this system is

$$\left\{ \begin{pmatrix} -12 \\ 5 \\ 1 \end{pmatrix} + r \begin{pmatrix} 465 \\ -120 \\ 0 \end{pmatrix} : 24 \leq r \leq 38.75 \right\}.$$

Example: A patient is injected with 1 gram of a long-lasting drug. At any time thereafter, a portion of the drug is in use by the body, a portion is stored in the lever, and the remainder is in circulation throughout the body. Each day, suppose that of the amount in the circulatory system on the previous day, 20% is stored in the liver, 30% goes into use, and the rest remains in circulation. Suppose that 20% of the amount in the liver and 10% of the amount in use on the previous day are released into the bloodstream, with no direct transfer between the liver and the sites of utilization. Find the amount of the drug in the various locations when the distribution reaches equilibrium.

Let x be the amount in the circulatory system, y be the amount in the liver, and z be the amount in use by the body. Then we have the system of equations

$$\begin{array}{ccccc} x+ & y+ & z & = & 1 \\ 0.5x+ & 0.2y+ & 0.1z & = & x \\ 0.2x+ & 0.8y & & = & y \\ 0.3x+ & & 0.9z & = & z \end{array}.$$

The first equation is obvious, since there is only one gram in the body, and it has to be split up between the three locations. The last two coefficients in the first column and the last two coefficients in the first row were given in the problem, as were the zero coefficients. In the last three equations, each of the columns must add up to one, so that all of the drug in each location is accounted for, which allows us to fill in the rest of the coefficients. Because the arithmetic is tedious, I'll skip the row reduction, but the RREF of the coefficient matrix is

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0.2 \\ 0 & 1 & 0 & 0.2 \\ 0 & 0 & 1 & 0.6 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Section 7.1 Mathematical Induction

Mathematical induction is a very useful tool that is often used in proofs. It will be especially useful when we get to determinants in Chapter 7.

Definition 18 *The natural numbers are the positive integers. The set* $\{1, 2, 3, \ldots\}$ *of all natural numbers is denoted* $\mathbb{N}$.

Mathematical induction is most often used to prove statements that claim that something is true for each value of $n \in \mathbb{N}$

Example: Note that

$$1 = 1$$

$$1 + 3 = 4$$

$$1 + 3 + 5 = 9$$

$$1 + 3 + 5 + 7 = 16$$

$$1 + 3 + 5 + 7 + 9 = 25$$

From this evidence, we might conjecture that

$$\sum_{k=1}^{n} (2k - 1) = n^2.$$

We could certainly check this by hand for a few more examples. We could even write a computer program that would check it for a whole lot of examples. But we can't prove this statement by doing examples, because it would take an endless amount of time, even for a computer.

We can think of the situation this way: for each $n \in \mathbb{N}$ we have a statement, which we'll call S_n. In the case of the example, the statement is "the sum $\sum_{k=1}^{n}(2k-1)$ is equal to n^2." A strategy to prove that S_n is true for every $n \in \mathbb{N}$ might be as follows:

- Show that S_1 is true.
- Show that S_1 being true implies that S_2 is true.
- Show that S_2 being true implies that S_3 is true.
- Show that S_3 being true implies that S_4 is true.
- $\cdots$

This strategy works in theory, but in practice, there are still infinitely many steps. What we need to do is shrink all those steps down into one. We do this by proving that whenever S_n is true, it implies that S_{n+1} is true. Then the steps are

1. Show that S_1 is true.
2. Show that S_n being true implies that S_{n+1} is true.

Let's use this method on our example. First, we must show that S_1 is true, or in other words, we must show that

$$\sum_{k=1}^{1}(2k-1) = 1^2.$$

This is easy, since

$$\sum_{k=1}^{1}(2k-1) = 2(1) - 1 = 1.$$

Now for step 2. We *assume* that

$$\sum_{k=1}^{n}(2k-1) = n^2,$$

and we use this fact to show that

$$\sum_{k=1}^{n+1}(2k-1) = (n+1)^2.$$

We have that

$$\sum_{k=1}^{n+1}(2k+1) = \sum_{k=1}^{n}(2k-1) + 2(n+1) - 1 = n^2 + 2n + 1 = (n+1)^2.$$

The two steps that are involved in completing a proof by induction are called the basis step and the induction step. In the induction step, it is some times more convenient to write it as $S_{n-1} \;\Rightarrow\; S_n$. The only difference here is in notation.

Example: Show that

$$\sum_{k=1}^{n} 2^{k-1} = 2^n - 1$$

for all $n \in \mathbb{N}$.

To proceed by induction, we must first check the basis step. If $n = 1$, then

$$\sum_{k=1}^{1} 2^{k-1} = 2^0 = 1$$

and

$$2^{1-1} = 1.$$

Next, we proceed with the induction step. In this case we'll show that

$$\sum_{k=1}^{n-1} 2^{k-1} = 2^{n-1} - 1$$

implies

$$\sum_{k=1}^{n} 2^{k-1} = 2^n - 1.$$

We have that

$$\sum_{k=1}^{n} 2^{k-1} = \sum_{k=1}^{n-1} 2^{k-1} + 2^{n-1} = 2^{n-1} - 1 + 2^{n-1} = 2 \cdot 2^{n-1} - 1 = 2^n - 1,$$

and this completes the induction step.

The tower of Hanoi problem concerns three pegs and a stack of 64 rings stacked on one of the pegs. The rings are of different sizes, and they are stacked in decreasing size with the largest on the bottom and the smallest on the top. The problem is to find a strategy for transferring all the rings from the peg they are on to another peg, subject to the rules that only one ring can be off the pegs at a time, and they must always be stacked from largest to smallest.

Example: The Tower of Hanoi puzzle with n rings can be solved in $2^n - 1$ moves.

Of course, we proceed by induction. First, we check the base case where $n = 1$. This is simple, since if there is only one ring, we can move it to another peg in $2^1 - 1 = 1$ move. Now, for the induction step, we assume that we can solve a puzzle with n rings in $2^n - 1$ steps, and use this to show that we can solve a puzzle with $n+1$ rings in $2^{n+1} - 1$ steps. At first we ignore the largest ring on the bottom of the stack. Then we use our n ring solution to move the other rings from the original peg to a new peg. Then we use one move to take the largest ring to the unoccupied peg. Then we again perform the n ring solution to move the other rings on top of the largest. This takes a total of $2^n - 1 + 1 + 2^n - 1 = 2^{n+1} - 1$ moves.

Example: Show that $4^n - 1$ is divisible by 3 whenever $n \in \mathbb{N}$.

First we check the basis step where $n = 1$. In this case $4^n - 1 = 3$, which is obviously divisible by 3. Next, we assume that $4^n - 1$ is divisible by 3 and use that to show that $4^{n+1} - 1$ is divisible by 3. We have that

$$4^{n+1} - 1 = 4 \cdot 4^n - 1 = 4 \cdot 4^n - 4 + 3 = 4(4^n - 1) + 3.$$

Since both of the numbers in this sum are divisible by three, it must be true that $4^{n+1} - 1$ is divisible by 3.

Chapter 3: Dimension Theory

We have already made reference to dimension several times in class so far, even though we haven't given a precise definition for what we mean. We can probably all agree that a line is a one dimensional object, and that a plane is two dimensional. But sometimes there are gray areas. For example, if we twist a line into a knot, is the resulting object still one dimensional, or is it three dimensional?

Fortunately, we can develop a very precise definition for what the dimension of a vector space is. The mathematical machinery that we develop on the way to making this definition will be very useful throughout the rest of the course.

Section 3.1: Linear Combinations

Back in Chapter 1, we looked at planes in $\mathbb{R}^3$, and we saw that we were able to build them out of two vectors, as long as the two vectors were not parallel. The way that we built the plane form these vectors was to take scalar multiples of the vectors and add them together. When we did this for all possible combinations of the scalars, we got a plane. This is an example of a linear combination.

Definition 19 *Let $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a finite set of vectors in a vector space V. A linear combination of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is any vector of the form*

$$r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_n\vec{v}_n,$$

where $r_1, r_2, \ldots, r_n \in \mathbb{R}$. The real numbers $r_1, r_2, \ldots, r_n$ are called the coefficients of the linear combination.

If we want to leave any of the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ out of the linear combination, then we simply set its coefficient to 0. Therefore, the vector $\vec{0}$ can be written as a linear combination of any set of vectors. It is also convenient to say that this is true even if the set in question is empty.

It is easy to pick scalars and take linear combinations of a set of vectors. The more interesting question is to determine whether a given vector can be written as a linear combination of a set of vectors. The problem is to find out if coefficients exist, and if so to determine them.

Example: Determine whether $(8, -2, -7)$ is a linear combination of $(1, 1, 1)$, $(-2, 0, 1)$, and $(-1, 3, 5)$.

If this is true, then there exist scalars r_1, r_2 and r_3 such that

$$r_1\begin{pmatrix}1\\1\\1\end{pmatrix} + r_2\begin{pmatrix}-2\\0\\1\end{pmatrix} + r_3\begin{pmatrix}-1\\3\\5\end{pmatrix} = \begin{pmatrix}8\\-2\\-7\end{pmatrix}.$$

This is equivalent to the system of equations

$$\begin{array}{cccccc} r_1 & -2r_2 & -r_3 & = & 8 \\ r_1 & & +3r_3 & = & -2 \\ r_1 & +r_2 & +5r_3 & = & -7 \end{array},$$

which has the augmented matrix

$$\left(\begin{array}{ccc|c} 1 & -2 & -1 & 8 \\ 1 & 0 & 3 & -2 \\ 1 & 1 & 5 & -7 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -2 & -1 & 8 \\ 0 & 2 & 4 & -10 \\ 0 & 3 & 6 & -15 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 3 & -2 \\ 0 & 1 & 2 & -5 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

This system of equations is consistent, but the solution is not unique. This means that the first vector can be written as a linear combination of the other three. In fact, there are infinitely many different choices of coefficients that will do the job.

The method that we have just used also works in vector spaces other than $\mathbb{R}^n$.

Example: Determine whether the polynomial defined by $q(x) = x^2 + x + 2$ is a linear combination of the polynomials defined by $p_1(x) = x^2 + 5$ and $p_2(x) = x^2 + 2x - 1$.

This means that we have to determine whether there are constants r_1 and r_2 such that

$$x^2 + x + 2 = r_1(x^2 + 5) + r_2(x^2 + 2x - 1) = (r_1 + r_2)x^2 + (2r_2)x + (5r_1 - r_2).$$

Since these polynomials are equal, the coefficients must also be equal, so

$$r_1 + r_2 = 1 \quad 2r_2 = 1 \quad 5r_1 - r_2 = 2.$$

The augmented matrix is

$$\left(\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 5 & -1 & 2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & -6 & -3 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{array}\right).$$

This time there is only one solution, $r_1 = r_2 = 1/2$.

Example: Determine whether the exponential function e^x is a linear combination of the functions defined by $f(x) = e^{2x}$ and $g(x) = e^{3x}$.

If this is true, then there exist constants r_1 and r_2 such that

$$e^x = r_1 e^{2x} + r_2 e^{3x}.$$

This equation must hold for every value of x. Let's substitute in a few different values:

$$\begin{aligned} x = 0: &\quad 1 = r_1 + r_2 \\ x = 1: &\quad e = r_1 e^2 + r_2 e^3 \\ x = 2: &\quad e^2 = r_1 e^4 + r_2 e^6 \end{aligned}$$

As an augmented matrix, this is

$$\left(\begin{array}{cc|c} 1 & 1 & 1 \\ e^2 & e^3 & e \\ e^4 & e^6 & e^2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & e^3 - e^2 & e - e^2 \\ 0 & e^6 - e^4 & e^2 - e^4 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 1 - \frac{1-e}{e^2-e} \\ 0 & 1 & \frac{1-e}{e^2-e} \\ 0 & 0 & e^2 - e^4 + (e^4 - e^6)\frac{1-e}{e^2-e} \end{array}\right).$$

In this case, there is no solution. Therefore, e^x is not a linear combination of e^{2x} and e^{3x}.

Section 3.2 Span

Now that we know how to take linear combinations of a set of vectors, we want to know what we get when we take all linear combinations of a set of vectors.

Definition 20 *Let $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a finite set of vectors in a vector space V. The subset of V spanned by $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is the set of all linear combinations of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. This set is called the span of $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ and is denoted by*

$$\mathit{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}.$$

We have already seen that the span of of a single vector is a line and that the span of two nonparallel vectors is a plane. If we take the span of two vectors that are parallel, we get a line. In $\mathbb{R}^3$, the span of three vectors that do not lie in the same plane gives the whole space $\mathbb{R}^3$. If they are in the same plane, then we just get that plane. If all three vectors are parallel, then we just get a line.

Another example that we have seen before is that the vectors $\vec{e}_1 = (1, 0)$ and $\vec{e}_2 = (0, 1)$ span all of $\mathbb{R}^2$. In other words, every vector in $\mathbb{R}^2$ can be written as a linear combination of $\vec{e}_1$ and $\vec{e}_2$. The same thing holds true in $\mathbb{R}^3$: each vector in R^3 can be written as a linear combination of the vectors $\vec{e}_1 = (1, 0, 0)$, $\vec{e}_2 = (0, 1, 0)$, and $\vec{e}_3 = (0, 0, 1)$.

We are often interested in finding a spanning set for a vector space V. That is, we want to find vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ so that $V = \text{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$.

Theorem 8 *The span of any finite set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ of vectors in a vector space V is a subspace of V.*

Proof : We use the subspace theorem. Let $S = \text{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. First, we know that S is nonempty, since $\vec{0} \in S$. Let $\vec{u}, \vec{v} \in S$. The we can write

$$\vec{u} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n$$

and

$$\vec{v} = b_1\vec{v}_1 + b_2\vec{v}_2 + \cdots + b_n\vec{v}_n$$

for some choices of scalars $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$. We have that

$$\vec{u} + \vec{v} = (a_1 + b_1)\vec{v}_1 + (a_2 + b_2)\vec{v}_2 + \cdots + (a_n + b_n)\vec{v}_n \in S.$$

Also, if $r \in \mathbb{R}$, then

$$r\vec{u} = (ra_1)\vec{v}_1 + (ra_2)\vec{v}_2 + \cdots + (ra_n)\vec{v}_n. \in S$$

Thus S is closed under vector addition and scalar multiplication. By the subspace theorem, S is a subspace of V.

The obvious question to answer is when a given set spans the whole vector space in which it lives.

Example: Show that

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

spans the vector space $\mathbb{M}(2,2)$.

We have to show that given an arbitrary element $A \in \mathbb{M}(2,2)$ we can find scalars r_1, r_2, r_3, r_4 such that

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = r_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + r_2 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + r_3 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + r_4 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} r_1 & r_2 \\ r_3 & r_4 \end{pmatrix}.$$

It is clear that this problem always has the solution $r_1 = a$, $r_2 = b$, $r_3 = c$, $r_4 = d$.

Example: Show that the vectors $\{(1,-2,3), (1,0,1), (0,1,-2)\}$ span $\mathbb{R}^3$.

This means that we must show that every vector (x, y, z) can be written as a linear combination of these three vectors. In other words, there must exist scalars r_1, r_2, r_3 such that

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = r_1 \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} + r_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + r_3 \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}.$$

This is equivalent to the system with coefficient matrix

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

always having a solution, no matter what we put on the other side of the equals sign. Then in RREF, A cannot have a row of zeros, because if it did, then we could choose a right-hand side so that the last column of the augmented matrix was not zero, and then the system would not have a solution. We have

$$\begin{pmatrix} 1 & 1 & 0 \\ -2 & 0 & 1 \\ 3 & 1 & -2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & -2 & -2 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since there are no rows of zeros in the RREF of A, we conclude that the given vectors span $\mathbb{R}^3$.

Generally speaking, if we select three random vectors in $\mathbb{R}^3$, they will usually span all of $\mathbb{R}^3$, but this is not always the case.

Example: Write the subspace spanned by $\{(1,1,1),(-2,0,1),(-1,3,5)\}$ in the form of a plane in $\mathbb{R}^3$.

If these vectors span a plane and not all of $\mathbb{R}^3$, then it must be true that one of the vectors can be written as a linear combination of the other two, meaning that one of the vectors lies in the plane generated by the other two. So if the span is a plane, there must be scalars r_1 and r_2 such that

$$r_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + r_2 \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 5 \end{pmatrix},$$

which is equivalent to the augmented system

$$\left(\begin{array}{cc|c} 1 & -2 & -1 \\ 1 & 0 & 3 \\ 1 & 1 & 5 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -2 & -1 \\ 0 & 2 & 4 \\ 0 & 3 & 6 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right).$$

Since the last vector lies in the plane generated by the first two, we can write the plane as

$$\{r_1\vec{v}_1 + r_2\vec{v}_2 : r_1, r_2 \in \mathbb{R}\},$$

where $\vec{v}_1 = (1,1,1)$ and $\vec{v}_2 = (-2,0,1)$.

Example: In a vector space V, suppose that $\vec{v}$ and $\vec{w}$ are both linear combinations of $\vec{x}$ and $\vec{y}$. Prove that if $\{\vec{v},\vec{w}\}$ spans V, then $\{\vec{x},\vec{y}\}$ spans V.

Let $\vec{u} \in V$. We have to show that there exists scalars r_1 and r_2 such that $\vec{u} = r_1\vec{x}+r_2\vec{y}$. Since $\vec{v}$ and $\vec{w}$ are linear combinations of $\vec{x}$ and $\vec{y}$, we can write

$$\vec{v} = a_1\vec{x} + a_2\vec{y}, \;\; \vec{w} = b_1\vec{x} + \vec{b}_2\vec{y},$$

Since $\{\vec{v}, \vec{w}\}$ spans V, we can write

$$\vec{u} = c_1\vec{v} + c_2\vec{w}.$$

Therefore,

$$\vec{u} = c_1\vec{v} + c_2\vec{w} = c_1(a_1\vec{x} + a_2\vec{y}) + c_2(b_1\vec{x} + b_2\vec{y}) = (c_1a_1 + c_2b_1)\vec{x} + (c_1a_2 + c_2b_2)\vec{y},$$

so it must be that $r_1 = c_1a_1 + c_2b_1$ and $r_2 = c_1a_2 + c - 2b_2$.

Section 3.3: Linear Independence

If we take the span of two nonparallel vectors in $\mathbb{R}^3$, we get a plane. But if we add another vector that lies in this plane and take the span, we still just get the same plane. However, if we add a vector not in the plane, the resulting set spans all of $\mathbb{R}^3$. In this section, we want to make precise this idea some vectors add nothing new to a span, while others do.

Definition 21 *A finite set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ of vectors in a vector space is linearly independent if and only if the equation*

$$r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_n\vec{v}_n = \vec{0}$$

implies that $r_1 = r_2 = \cdots = r_n = 0$. If it is possible for the equation to hold when one or more of the coefficients is nonzero, the set is linearly dependent.

As a special case, we say that an empty set of vectors is linearly independent. This is the case where $n = 0$, and this assumption will be convenient when we talk about dimension.

Example: Show that $\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ is a linearly independent subset of $\mathbb{R}^3$.

To do this, we have to take the equation

$$r_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + r_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + r_3 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and determine whether or not the only solution is $r_1 = r_2 = r_3 = 0$. But this equation is the same as the homogeneous equation with coefficient matrix

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

Since there is a leading one in each column, the homogenous system has only the trivial solution. This implies that the vectors are linearly independent.

Example: Determine if the set of matrixes

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

is linearly independent.

We have to examine the equation

$$r_1 \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} + r_2 \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} + r_3 \begin{pmatrix} 0 & 2 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and determine whether it is possible to satisfy this equation without all the coefficients equal to zero. Equating coefficients, we see that the matrix equation above is equivalent to a homogeneous system of equations with coefficient matrix

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

The RREF form of the coefficient matrix has columns that do not contain leading ones, so there are nontrivial solutions. This implies that the matrices are linearly dependent. In particular, if we take $r_1 = -1$, $r_2 = 1$ and $r_3 = 1$ the linear combination is the zero matrix. Notice that if we throw out the last matrix, we will have a linearly independent set. This would be equivalent to removing the last column of the coefficient matrix.

If we are trying to determine if a set of polynomials is linearly independent, we can use their coefficients in the same way that we use the components of vectors in $\mathbb{R}^n$. But when we deal with other functions, we usually have to resort to plugging in values of the independent variable to get a homogeneous system to solve. If the system has only the trivial solution, then the functions are linearly independent. If we can't find such a system, then it is likely that the functions are linearly dependent, and we need to look for an identity that allows a linear combination of the functions to equal zero.

Example: Determine whether the functions defined by $\sin^2 x$, $\cos 2x$, and 1 form a linearly independent set.

We have to determine if the equation

$$r_1 \sin^2 x + r_2 \cos 2x + r_3 = 0$$

has nontrivial solutions. If we evaluate this equation at $x = 0$, $x = \pi/2$, and $x = \pi$, we get

$$r_2 + r_3 = 0 \quad r_1 - r_2 + r_3 = 0 \quad r_2 + r_3 = 0.$$

This homogeneous system has the coefficient matrix

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

Since this equation has trivial solutions, it might be a good indication that the functions are linearly dependent. Therefore, we look for an identity that combines them. Since $\cos 2x = 1 - 2\sin^2 x$, we see that the functions are linearly dependent.

Theorem 9 *The set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ of vectors in a vector space is linearly dependent if and only if one of the vectors can be written as a linear combination of the other vectors in the set.*

Proof : ($\Rightarrow$) Suppose that the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is linearly dependent. That means that there is a nontrivial solution to the equation

$$r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_n\vec{v}_n = \vec{0}.$$

Suppose that $r_k \neq 0$. Then we can write

$$r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_{k-1}\vec{v}_k + r_{k+1}\vec{v}_k + \cdots + r_n\vec{v}_n = -r_k\vec{v}_k$$

$$\Rightarrow \quad \vec{v}_k = \frac{-r_1}{r_k}\vec{v}_1 + \frac{-r_2}{r_k}\vec{v}_2 + \cdots + \frac{-r_{k-1}}{r_k}\vec{v}_{k-1} + \frac{-r_{k+1}}{r_k}\vec{v}_k + \cdots + \frac{-r_n}{r_k}\vec{v}_n$$

Therefore, we have written one of the vectors as a linear combination of the others.

($\Leftarrow$) Suppose one of the vectors in the set can be written as a linear combination of the rest of the vectors in the set, say

$$\vec{v}_k = r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_{k-1}\vec{v}_{k-1} + r_{k+1}\vec{v}_{k+1} + \cdots + r_n\vec{v}_n.$$

Then we can rearrange this equation as

$$r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_{k-1}\vec{v}_{k-1} - \vec{v}_k + r_{k+1}\vec{v}_{k+1} + \cdots + r_n\vec{v}_n = \vec{0}.$$

Since not all of the coefficients on the left are zero, we conclude that the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is linearly dependent.

In the case when the set in question contains only two vectors, this theorem implies that the set is linearly dependent if and only if the vectors are scalar multiples of one another:

$$\left\{ \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \begin{pmatrix} 27 \\ -18 \end{pmatrix} \right\} \text{ is linearly dependent,}$$

but

$$\left\{ \begin{pmatrix} 4 \\ -1 \end{pmatrix}, \begin{pmatrix} 8 \\ -3 \end{pmatrix} \right\} \text{ is linearly independent.}$$

When there are three or more vectors, things are a little more difficult. The set

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

is linearly dependent, even though none of the vectors is a multiple of the others. In this case the third vector is the sum of the first two.

Section 3.4: Basis

Now that we have defined spanning sets and linear independence, we are able to define a basis.

Definition 22 *A finite subset $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ of a vector space V is a basis for V if and only if the set is linearly independent and spans all of V.*

One way to interpret this definition is to say that a basis is a smallest spanning set for a vector space. If we take a basis for V and add any other vector to the basis, the result is a linearly dependent set, because the vector that we added was already in the span of the basis.

Definition 23 *A vector space V has dimension n if and only if there is a basis for V containing exactly n vectors. In this situation, we say that V is n-dimensional and write*

$$dim(V) = n.$$

Example: Show that $\{x+1, x-1\}$ is a basis for $\mathbb{P}_1$.

First, let us show that this set spans $\mathbb{P}_1$. To do this, we show that an arbitrary element of $\mathbb{P}_1$, say $ax+b$ can be written as a linear combination of $x+1$ and $x-1$.

In other words, there must exist scalars r_1 and r_2 such that

$$r_1(x+1)+r_2(x-1)=ax+b. \Rightarrow (r_1+r_2)x+(r_1-r_2)=ax+b.$$

Since these two polynomials are equal, their coefficients are equal. This gives rise to a system of equations with augmented matrix

$$\left(\begin{array}{cc|c} 1 & 1 & a \\ 1 & -1 & b \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & a \\ 0 & -2 & b-a \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & a-\frac{a-b}{2} \\ 0 & 1 & \frac{a-b}{2} \end{array}\right).$$

Notice that this system of equations always has a solution, since there are leading ones in all of the rows of the RREF. Thus the set spans all of $\mathbb{P}_1$. We must also show that the set is linearly independent. But we have already done this, since of we let $a=b=0$ above, this shows that the only solution to the equation

$$r_1(x+1)+r_2(x-1)=0$$

is $r_1=r_2=0$. Therefore $\{x+1, x-1\}$ is a basis for $\mathbb{P}_1$. This also shows that the dimension of $\mathbb{P}_1$ is 2.

Even though we have been saying it for a while, we can finally show that $\mathbb{R}^n$ has dimension n. We need only find a basis for $\mathbb{R}^n$ that has n vectors in it. This is easy to do, since as we have seen before, the vectors $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$, where $\vec{e}_j$ is a vector with a 1 in the jth position and zeros elsewhere span $\mathbb{R}^n$ and are linearly independent.

Another easy example is to show that the set $\{1, x, x^2, \ldots, x^n\}$ is a basis for $\mathbb{P}_n$. Obviously, every polynomial of degree n or less is a linear combination of these functions. And the functions are linearly independent since the only way to have a linear combination of them equal to zero (the zero function) is for the coefficients to be zero (by the fundamental theorem of algebra).

Definition 24 *A vector space is finite-dimensional if and only if it contains a finite basis. Otherwise, the vector space is infinite-dimensional.*

Example: Show that $\{(1,2,3),(0,1,2),(0,0,1),\}$ is a basis for $\mathbb{R}^3$.

We can show this by examining the matrix that has these vectors as columns. We have that

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

Since the RREF of this matrix has a leading one in each row, any system that has this coefficient matrix will be consistent. This implies that the given vectors span $\mathbb{R}^3$. Also, they are linearly independent, since any system with this coefficient matrix will have a unique solution, since the RREF has a leading one in each column. Therefore, we have a basis for $\mathbb{R}^3$.

Section 3.5 Dimension

If you were paying very close attention in the last section, you noticed that we said that the dimension of a vector space was *the* number of vectors in a basis for the space. But conceivably there could be bases that have different numbers of vectors. The following theorem shows that this is not possible.

Theorem 10 *Suppose V is a vector space. If $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ spans V and $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ is a linearly independent subset of V, then $n \geq m$.*

Proof : Let's assume that the conclusion of the theorem is false and try to derive a contradiction. So we assume that $n < m$. Since the $\vec{w}$ vectors span V, we can write each one of the $\vec{v}$ vectors as a linear combination of the $\vec{w}$'s, say

$$\begin{array}{c} a_{11}\vec{w}_1 + a_{12}\vec{w}_2 + \cdots + a_{1n}\vec{w}_n = \vec{v}_1 \\ a_{21}\vec{w}_1 + a_{22}\vec{w}_2 + \cdots + a_{2n}\vec{w}_n = \vec{v}_2 \\ \vdots \\ a_{m1}\vec{w}_1 + a_{m2}\vec{w}_2 + \cdots + a_{mn}\vec{w}_n = \vec{v}_m \end{array}.$$

The system of equations

$$\begin{array}{c} a_{11}c_1 + a_{21}c_2 + \cdots + a_{m1}c_m = 0 \\ a_{12}c_1 + a_{22}c_2 + \cdots + a_{m2}c_m = 0 \\ \vdots \\ a_{1n}c_1 + a_{2n}c_2 + \cdots + a_{mn}c_m = 0 \end{array}$$

will have nontrivial solutions because it is homogeneous with more columns than rows. Suppose we have such a nontrivial solution $c_1, c_2, \ldots, c_m$. Then we have that

$$\begin{array}{c} c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n \\ = c_1(a_{11}\vec{w}_1 + a_{12}\vec{w}_2 + \cdots + a_{1n}\vec{w}_n) \\ +c_2(a_{21}\vec{w}_1 + a_{22}\vec{w}_2 + \cdots + a_{2n}\vec{w}_n) + \cdots \\ +c_m(a_{m1}\vec{w}_1 + a_{m2}\vec{w}_2 + \cdots + a_{mn}\vec{w}_n) \end{array}$$

$$
\begin{aligned}
&= (c_1a_{11} + c_2a_{21} + \cdots + c_ma_{m1})\vec{w}_1 \\
&+(c_1a_{12} + c_2a_{22} + \cdots + c_ma_{m2})\vec{w}_2 + \cdots \\
&+(c_1a_{1n} + c_2a_{2n} + \cdots + c_ma_{mn})\vec{w}_n \\
&= 0\vec{w}_1 + 0\vec{w}_2 + \cdots + 0\vec{w}_n = \vec{0}.
\end{aligned}
$$

This means that we have found a nontrivial linear combination of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ that is zero. This contradicts the assumption that the $\vec{v}$s are linearly independent, so it must be that the assumption $n < m$ is false.

This theorem was the first step in proving that the number of vectors in a basis is unique. Here is part two:

Theorem 11 *If $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ and $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ are bases for a vector space V, then $m = n$.*

Proof : Since the $\vec{v}$s span V and the $\vec{w}$s are linearly independent, it must be that $n \leq m$. Also, since the $\vec{w}$s span V and the $\vec{v}$s are linearly independent, we have $m \leq n$. Therefore $m = n$.

As we have mentioned before, a basis is a smallest possible spanning set. The next theorem tells us how we can take a spanning set an whittle it down to a basis.

Theorem 12 *Supose $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ spans a vector space V. Then some subset of $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis for V.*

Proof : If $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is already linearly independent, then we are done. If not, then there is a vector, say $\vec{v}_i$ such that

$$\vec{v}_i = r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_{i-1}\vec{v}_{i-1} + r_{i+1}\vec{v}_{i+1} + \cdots + r_n\vec{v}_n.$$

Now we need to show that the subset $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_n\}$ still spans V. This is true since if we take any $\vec{v} \in V$, we can write it as

$$\vec{v} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n.$$

We use the expansion of $\vec{v}_i$ in terms of the rest of the vectors to write

$$\vec{v} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_{i-1}\vec{v}_{i-1}$$

$$+a_i(r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_{i-1}\vec{v}_{i-1} + r_{i+1}\vec{v}_{i+1} + \cdots + r_n\vec{v}_n) + a_{i+1}\vec{v}_{i+1} + \cdots + a_n\vec{v}_n.$$

$$= (a_1 + a_ir_1)\vec{v}_1 + \cdots + (a_{i-1} + a_ir_{i-1})\vec{v}_{i-1} + (a_{i+1} + a_ir_{i+1})\vec{v}_{i+1} + \cdots + (a_n + a_ir_n)\vec{v}_n$$

So we can throw out $\vec{v}_i$ and still have a spanning set. If the remaining vectors are linearly independent, then we are done. If not, we can repeat this process to remove another vector. Eventually we will arrive at a linearly independent set, even if we have to go all the way to the empty set, which is linearly independent by convention.

Example: An easy corollary of this result is that a vector space is finite dimensional if and only if it has a finite spanning set. This is easily seen, since if we start with a finite spanning set, the theorem above implies that we can reduce it to a basis, which obviously has a finite number of vectors. On the other hand, if we know the vector space is finite dimensional, then it has a basis with a finite number of elements, which is itself a spanning set.

Notice that the proof of the theorem above did more than just prove the theorem, it actually provided an algorithm for constructing a basis from a spanning set.

Example: Let $S = \text{span}\{(1,2,-1),(3,1,2),(2,-1,3),(1,-3,4)\}$. Find a subset of this set that is a basis for S.

In the proof of the theorem, we started with a spanning set, and we found vectors in the spanning set that could be written as a linear combination of the remaining vectors. Then we removed these vectors. To see how this can be done for our current example, let's form the matrix that has the given vectors as columns and row reduce it:

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

For a moment, ignore the last column and imagine that there is a line before the third column. Then this is telling us how we can write the third vector as a linear combination of the first two. If we ignore the third column, then it tells us how the last vector can be written as a linear combination of the first two. Therefore, we throw out the last two vectors to get a linearly independent set that still spans S.

The previous theorem tells us how to take a spanning set and reduce it to a basis. The next lemma and theorem tell us how to start with a linearly independent subset of a vector space and expand it to a basis.

Lemma 1 *Suppose $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a linearly independent subset of a vector space V. Consider a vector $\vec{v}_{n+1} \in V$. If $\vec{v}_{n+1} \notin \text{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, then $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{v}_{n+1}\}$ is linearly independent.*

Proof : To determine if the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n+1}\}$ is linearly independent, we must consider the equation

$$r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_{n+1}\vec{v}_{n+1} = \vec{0}$$

and decide whether it has a nontrivial solution. If $r_{n+1} \neq 0$, we can solve for $\vec{v}_{n+1}$ in the equation above, which would imply that the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n+1}\}$ is linearly dependent. If $r_{n+1} = 0$, then we end up with the equation

$$r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_n\vec{v}_n = \vec{0},$$

and since the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is linearly independent, this implies that $r_1 = r_2 = \cdots = r_n = 0$. Thus, all of the coefficients must be zero, which implies that the set is linearly independent.

Theorem 13 *Suppose $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ is a linearly independent subset of a finite-dimensional vector space V. Then we can adjoin vectors $\vec{v}_{m+1}, \ldots, \vec{v}_{m+k}$ so that $\{\vec{v}_1, \ldots, \vec{v}_m, \vec{v}_{m+1}, \ldots, \vec{v}_{m+k}\}$ is a basis for V.*

Proof : If the set $\{\vec{v}_1, \ldots, \vec{v}_m\}$ spans V, then we are done. If not, then there is some vector $\vec{v}_{m+1} \in V$ that is not in $\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_m\}$. By the lemma, the set $\{\vec{v}_1, \ldots, \vec{v}_{m+1}\}$ is linearly independent. If this set does not span all of V, then we repeat the same process again. Eventually the process will terminate because V is finite dimensional.

Example: Find a basis for $\mathbb{R}^4$ that contains the vectors $\vec{v}_1 = (1, 1, 1, 1)$ and $\vec{v}_2 = (0, 0, 1, 1)$.

A simple way to approach this problem is to use the vectors from the standard basis $\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4$ to fill out the basis. We have two vectors for our basis, and since $\dim(\mathbb{R}^4) = 4$, we need two more. Consider the matrix

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

$$\sim \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

When we have a basis, the matrix that is formed using the basis vectors as columns row reduces to one that has a leading one in each column. In the matrix above, the fourth column, which corresponds to $\vec{e}_2$ doesn't have a leading one, so we throw $\vec{e}_2$ out. There is no way to tell which of the vectors $\vec{e}_j$ we have to add so that we get a basis until we do the row reduction, so it is a good idea to add several to the matrix before we row reduce.

The desired basis is $\{\vec{v}_1, \vec{v}_2, \vec{e}_1, \vec{e}_3\}$.

Theorem 14 *A subspace S of a finite-dimensional vector space V is finite dimensional, and $\dim(S) \leq \dim(V)$.*

Theorem 15 *Suppose V is an n-dimensional vector space. If $\{\vec{v}_1, \ldots, \vec{v}_n\}$ spans V, then it is also linearly independent. If $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is linearly independent, then it spans V.*

Section 3.6: Coordinates

Consider the sets

$$E = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \quad \text{and} \quad B = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}.$$

These are both bases for $\mathbb{R}^2$, and therefore we can express any vector in $\mathbb{R}^2$ as a linear combination of either basis. For example

$$\begin{pmatrix} 6 \\ 9 \end{pmatrix} = 6 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 9 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and

$$\begin{pmatrix} 6 \\ 9 \end{pmatrix} = 5 \begin{pmatrix} 2 \\ 1 \end{pmatrix} - 4 \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

In the first basis, the coefficients are just the entries of the vector. In the second basis, it is not as easy to see what the coefficients should be.

In this section, for the definition that follows, it is important that the vectors in a basis be given a certain order.

Definition 25 *Let $B = \{\vec{u}_1, \ldots, \vec{u}_n\}$ be an ordered basis for a vector space V. For any $\vec{v} \in V$, write $\vec{v} = r_1\vec{u}_1 + \cdots + r_n\vec{u}_n$. The coordinate vector for $\vec{v}$ with respect to B is the element of $\mathbb{R}^n$ denoted $[\vec{v}]_B$ and defined by*

$$[\vec{v}]_B = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}.$$

If we change the order in which the vectors in the basis are listed, we get a different coordinate vector.

In our previous example, we have

$$\begin{pmatrix} 6 \\ 9 \end{pmatrix}_E = \begin{pmatrix} 6 \\ 9 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 6 \\ 9 \end{pmatrix}_B = \begin{pmatrix} 5 \\ -4 \end{pmatrix}.$$

In order to find the coordinates of a vector with respect to a basis, we have to solve a system of equations. To find a vector given the coordinates and a basis is much easier. For example, if

$$\begin{pmatrix} x \\ y \end{pmatrix}_B = \begin{pmatrix} 8 \\ 3 \end{pmatrix},$$

then

$$\begin{pmatrix} x \\ y \end{pmatrix} = 8\begin{pmatrix} 2 \\ 1 \end{pmatrix} + 3\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 19 \\ 5 \end{pmatrix}.$$

Theorem 16 *Suppose $B = \{\vec{u}_1, \ldots, \vec{u}_n\}$ is an ordered basis for a vector space V. For any $\vec{v} \in V$, there is a unique list of scalars $r_1, \ldots, r_n$ such that $\vec{v} = r_1\vec{u}_1 + \cdots + r_n\vec{u}_n$.*

Proof : Since B is a basis, it spans V, so we can write $\vec{v} = r_1\vec{u}_1 + \cdots + r_n\vec{u}_n$. This shows that coordinates with respect to B exist. Suppose there is another expansion of $\vec{v}$ with respect to B, say

$$\vec{v} = s_1\vec{u}_1 + \cdots + s_n\vec{u}_n.$$

Then

$$\vec{0} = \vec{v} - \vec{v} = (r_1 - s_1)\vec{u}_1 + \cdots + (r_n - s_n)\vec{u}_n.$$

Because B is linearly independent, we know that $r_1 - s_1 = \cdots = r_n - s_n = 0$, so the expansions are the same and the coordinate vector is unique.

Example: Find the coordinate vector

$$\begin{pmatrix} 3 \\ -8 \\ 7 \end{pmatrix}_B$$

with respect to the basis

$$B = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

If $B = \{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$, then we are seeking scalars r_1, r_2, r_3 such that

$$\begin{pmatrix} 3 \\ -8 \\ 7 \end{pmatrix} = r_1\vec{u}_1 + r_2\vec{u}_2 + r_3\vec{u}_3.$$

As an augmented matrix, this is

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & -8 \\ 0 & 0 & 1 & 7 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 11 \\ 0 & 1 & 1 & -8 \\ 0 & 0 & 1 & 7 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 11 \\ 0 & 1 & 0 & -15 \\ 0 & 0 & 1 & 7 \end{array}\right).$$

This implies that

$$\begin{pmatrix} 3 \\ -8 \\ 7 \end{pmatrix}_B = \begin{pmatrix} 11 \\ -15 \\ 7 \end{pmatrix}.$$

Example: Find the coordinate vector for $x^2 - x + 4$ with respect to the basis

$$B = \{x^2 + 1, 2x^2 + x - 1, x^2 + x\}$$

for $\mathbb{P}_2$.

We seek scalars r_1, r_2, r_3 such that

$$r_1(x^2 + 1) + r_2(2x^2 + x - 1) + r_3(x^2 + x) = x^2 - x + 4.$$

Gathering like terms, we find

$$(r_1 + 2r_2 + r_3)x^2 + (r_2 + r_3)x + (r_1 - r_2) = x^2 - x + 4.$$

Upon equating the coefficients, we arrive at the system of equations

$$\begin{array}{cccccc} r_1 & +2r_2 & +r_3 & = & 1 \\ & r_2 & r_3 & = & -1 \\ r_1 & -r_2 & & = & 4 \end{array}.$$

As an augmented system, this is

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

$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

Therefore,

$$[x^2 - x + 4]_B = \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix}.$$

Example: Write the coordinates of sinh with respect to the basis $B = \{e^x, e^{-x}\}$.

Since $\sinh(x) = (e^x - e^{-x})/2$, we have that

$$[\sinh x]_B = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}.$$

Chapter 5: Matrices

Section 5.1: Matrix Algebra

How should we multiply matrices together?

We could multiply them together element-wise, just like we add them. But it ends up that if we use a different (and unfortunately more complicated) definition, matrix multiplication becomes a very useful modeling tool.

Example: A trucking company has the following tables which show the number of routes between different citites:

	Jackson	Battle Creek
Albion	3	2
Homer	2	0
Parma	1	1

	Chicago	Detroit	Indianapolis	Cleveland
Jackson	2	1	1	3
Battle Creek	4	2	1	1

The number of ways to ship from Albion to Chicago is $3 \cdot 2 + 2 \cdot 4 = 14$. The number of ways to ship from Homer to Chicago is $2 \cdot 2 + 0 \cdot 4 = 4$. By the same reasoning, we can find the number of ways to get from any of the small cities to any of the large cities.

	Chicago	Detroit	Indianapolis	Cleveland
Albion	14	7	5	11
Homer	4	2	2	6
Parma	6	3	2	4

Before we re-formulate this in terms of matrices, let's recall a definition that you might have seen in multivariable calculus:

Definition 26 *Given two vectors $\vec{x}$ and $\vec{y}$ from $\mathbb{R}^n$, their dot product is defined by*

$$\vec{x} \cdot \vec{y} = \sum_{k=1}^{n} x_k y_k = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

The matrix representation of the tables are

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 0 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 & 1 & 3 \\ 4 & 2 & 1 & 1 \end{pmatrix}, \quad \text{and} \quad C = \begin{pmatrix} 14 & 7 & 5 & 11 \\ 4 & 2 & 2 & 6 \\ 6 & 3 & 2 & 4 \end{pmatrix}.$$

Notice that to get the i, j entry of C, we take the dot product of the ith row of A with the jth column of B.

Definition 27 *The matrix product of an $m \times n$ matrix $A = [a_{ij}]$ with an $n \times p$ matrix $B = [b_{jk}]$ is the $m \times p$ matrix, denoted AB, whose ik entry is the dot product*

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

of the ith row of A with the kth column of B.

Notice: The matrix product is only defined when the inner dimensions of the matrices agree. If we multiply an $m \times n$ matrix with an $n \times p$ matrix, the result is $m \times p$–it inherits its number of rows from the first matrix and its number of columns from the second matrix.

Example:

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

$$\begin{pmatrix} 2 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 0 & 2 \end{pmatrix} = \text{undefined}.$$

Notice that we are able to compute the matrix product column by column. Suppose that A is $n \times m$ and B is $m \times p$, and the columns of B are the vectors $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_p$. Then

$$AB = A \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_p \end{pmatrix} = \begin{pmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_p \end{pmatrix}.$$

Each of the matrix-vector products are defined since A is $n \times n$ and the columns are $m \times 1$.

Example: Compute the second row and the third column of the product

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

The second row of the product is

$$\left((3\ \ -2\ 1) \cdot \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} \quad (3\ \ -2\ 1) \cdot \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix} \quad (3\ \ -2\ 1) \cdot \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} \right) = (11\ \ -3\ \ -4),$$

and the third column is

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

We can use the definition of matrix multiplication to write systems of linear equations as matrix equations. If $A = [a_{ij}]$ is the coefficient matrix and $\vec{x}$ is the vector of unknowns for a system of linear equations, then

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

Therefore, if the right-hand sides of the equations are $b_1, b_2, \ldots, b_n$, we can write our system of equations as

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

Notice that usually when the product AB is defined, the product BA is not. The only time when both products is defined is when A is $m \times n$ and B is $n \times m$. A special case is when the number of rows and columns is equal.

Definition 28 *A square matrix is a matrix in which the number of rows equals the number of columns.*

When we defined vector spaces, one of the important properties was that there be an additive identity (zero vector), which had the property that whenever it was added to another vector, the result was the other vector. For matrix multiplication, we have a multiplicative identity that has a similar property.

Definition 29 *For any positive integer n, the $n \times n$ identity matrix is the matrix*

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix},$$

whose entries are ones along the diagonal running from upper left to lower right and zeros elsewhere. If the size of the matrix is understood, such a matrix is denoted simply by I.

Theorem 17 *Suppose $A, A' \in \mathbb{M}(m,n)$, $B, B' \in \mathbb{M}(n,p)$, $C \in \mathbb{M}(p,q)$, and $r \in \mathbb{R}$. Then*
(a) $(AB)C = A(BC)$ (associative)
(b) $(A + A')B = AB + A'B$ (right distributive law)
(c) $A(B + B') = AB + AB'$ (left distributive law)
(d) $(rA)B = r(AB) = A(rB)$ (associativity of scalar/matrix multiplication)
(e) $AI_n - A = I_m A$ (multiplicative identity)

Proof : (of (b)) Using the compact matrix notation introduced before, we have

$$(A+A')B = ([a_{ij}]+[a'_{ij}])[b_{jk}] = [a_{ij}+a'_{ij}][b_{jk}] = \left[\sum_{j=1}^{n}(a_{ij} + a'_{ij})b_{jk}\right] = \left[\sum_{j=1}^{n}(a_{ij}b_{jk} + a'_{ij}b_{jk})\right]$$

$$= \left[\sum_{j=1}^{n} a_{ij}b_{jk} + \sum_{j=1}^{n} a'_{ij}b_{jk}\right] = \left[\sum_{j=1}^{n} a_{ij}b_{jk}\right] + \left[\sum_{j=1}^{n} a'_{ij}b_{jk}\right] = [a_{ij}][b_{jk}]+[a'_{ij}][b_{jk}] = AB+A'B.$$

Notice that there is no commutative law for matrix multiplication. For example,

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{but} \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

From this example we also see that in matrix multiplication it is possible to multiply together two nonzero matrices to get the zero matrix.

When it is possible to switch the order in matrix multiplication and still get the same matrix as a result, we say that the two matrices commute. One easy example of this is when taking powers of a matrix:

$$A^0 = I, \quad A^1 = A, \quad A^2 = AA, \quad A^3 = AAA, \quad \text{etc.}$$

Then $A^m A^n = A^n A^m = A^{n+m}$.

Theorem 18 *Suppose that A is an $m \times n$ matrix and B is a $n \times q$ matrix. Let A' be the result of applying an elementary row operation to A. Then $A'B$ is the result of applying the same elementary row operation to AB.*

Proof : (for a row interchange) The rows of A and A' are the same, they are just in a different order. Therefore, the rows of AB and $A'B$ are the rows of $A'B$ are rearranged in the same manner as those in A'.

Section 5.2: Inverses

As we saw in the last section, a system of linear equations can be written in the form

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

If A, $\vec{x}$, and $\vec{b}$ were real numbers and not matrices/vectors, then we would solve this equation by dividing both sides of the equation by A, or equivalently, we would multiply both sides by the multiplicative inverse of A:

$$A\vec{x} = \vec{b} \;\Rightarrow\; A^{-1}A\vec{x} = A^{-1}\vec{b} \;\Rightarrow\; \vec{x} = A^{-1}\vec{b}.$$

Definition 30 *Let $A \in \mathbb{M}(m,n)$. The matrix $B \in \mathbb{M}(n,m)$ is a multiplicative inverse of A if and only if $AB = I_m$ and $BA = I_n$. If the first equation holds, we say that B is a right inverse of A. If the second holds, we say that B is a left inverse of A. If A has an inverse, we say that A is invertible or nonsingular.*

Example:

$$\begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus, the inverse of

$$\begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} \text{ is } \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}$$

and vice-versa.

Theorem 19 *An invertible matrix has a unique inverse.*

Proof : Suppose that A has two inverses, B, and C. Then

$$B = BI = BAC = IC = C.$$

Theorem 20 *If a matrix A is invertible, then A^{-1} is also invertible. In this case, $(A^{-1})^{-1} = A$.*

Proof : Since A^{-1} is the inverse of A, we have that $AA^{-1} = I$ and $A^{-1}A = I$. But these are exactly the conditions necessary for A to be the inverse of A^{-1}.

Theorem 21 *Suppose $A, B \in \mathbb{M}(n,n)$. If A and B are invertible, then so is AB, and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof : We have that

$$ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I \text{ and } B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I,$$

so $(AB)^{-1} = B^{-1}A^{-1}$.

So far we have defined the inverse of a matrix, but we haven't said anything about how to find it. For simplicity of notation, let's write $C = A^{-1}$. Then we want to find C such that $AC = I$. By our previous remarks about matrix multiplication, we have that

$$AC = A(\vec{c}_1\ \vec{c}_2 \cdots \vec{c}_n) = (A\vec{c}_1\ A\vec{c}_2\ \cdots\ A\vec{c}_n).$$

So if $AC = I$, then the columns of C satisfy

$$A\vec{c}_1 = \vec{e}_1,\ A\vec{c}_2 = \vec{e}_2, \ldots, A\vec{c}_n = \vec{e}_n.$$

This means that we can find the columns of C by solving these n systems of linear equations. Since each of the systems has the same coefficient matrix, we can solve them all at the same time by row reducing the augmented matrix $(A|I) \sim (I|A^{-1})$.

Example: Find the inverse of

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

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ -1 & 3 & 2 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 4 & 3 & 1 & 1 & 0 \\ 0 & -1 & -1 & -2 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 1 & 1 & 2 & 0 & -1 \\ 0 & 0 & -1 & -7 & 1 & 4 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -5 & 1 & 3 \\ 0 & 0 & 1 & 7 & -1 & -4 \end{array}\right) \Rightarrow A^{-1} = \begin{pmatrix} -1 & 0 & 1 \\ -5 & 1 & 3 \\ 7 & -1 & -4 \end{pmatrix}$$

Definition 31 *The rank of a matrix A, denoted rank(A), is the number of leading ones in the reduced row-echelon form of the matrix.*

Example: Find the ranks of the $m \times n$ zero matrix, I_n, and

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

The zero matrix and I_n are already in RREF, so it is clear that $\text{rank}(0) = 0$ and $\text{rank}(I_n) = n$. For A, we have

$$A = \begin{pmatrix} 1 & 3 & 1 & 1 & -2 \\ 2 & 6 & 2 & 5 & 1 \\ 0 & 0 & 0 & 3 & 2 \\ -3 & -9 & -3 & 1 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 3 & 1 & 1 & -2 \\ 0 & 0 & 0 & 3 & 5 \\ 0 & 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 4 & -2 \end{pmatrix} \sim \begin{pmatrix} 1 & 3 & 1 & 0 & -\frac{1}{6} \\ 0 & 0 & 0 & 1 & \frac{5}{3} \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & -\frac{26}{3} \end{pmatrix}.$$

At this point it is clear that A will have three columns with leading ones, so $\text{rank}(A) = 3$.

Suppose that $A \in \mathbb{M}(m,n)$ has rank r. Let S be the solution set to the homogeneous system of equations $A\vec{x} = \vec{0}$. Then S is a subspace of $\mathbb{R}^n$, since it is nonempty ($\vec{0} \in S$), closed under vector addition:

$$\vec{x}, \vec{y} \in S \;\Rightarrow\; A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} = \vec{0} + \vec{0} = \vec{0},$$

and closed under scalar multiplication:

$$r \in \mathbb{R} \Rightarrow A(r\vec{x}) = rA\vec{x} = r\vec{0} = \vec{0}.$$

We can find a basis for this subspace easily. We just solve the homogeneous system by row reduction and determine which are the free variables. Then we write the solution as a linear combination using the free variables as coefficients. The vectors in the linear combination are a basis for S. Obviously they span S, and they are also linearly independent (to see this, look at the last rows of the matrix formed by using these vectors as columns).

We get a basis element for S for each column of the RREF of A that doesn't have a leading one, and each column that does have a leading one contributes to the rank. Therefore,

$$\dim(S) + \operatorname{rank}(A) = n.$$

Theorem 22 *An $n \times n$ matrix A has a right inverse C if and only if rank$(A) = n$. In this case the right inverse is unique.*

Proof : ($\Leftarrow$) Suppose $\operatorname{rank}(A) = n$. Then $A \sim I_n$, and using the same sequence of elementary row operations, we have $(A|I) \sim (I|C)$ for some $n \times n$ matrix C. Let $C = (\vec{c}_1 \ \vec{c}_2 \ \cdots \ \vec{c}_n)$. Then

$$AC = A(\vec{c}_1 \ \vec{c}_2 \ \cdots \ \vec{c}_n) = (A\vec{c}_1 \ A\vec{c}_2 \ \cdots \ A\vec{c}_n) = (\vec{e}_1 \ \vec{e}_2 \ \cdots \ \vec{e}_n) = I.$$

($\Rightarrow$) Suppose that $\operatorname{rank}(A) \neq n$. Then we need to show that there is no C such that $AC = I$. Since $\operatorname{rank}(A) < n$, the RREF of A must have rows of zeros. If we row-reduce the augmented matrix $(A|I)$, the result will be a matrix that has some rows that start with n zeros. But there are no rows that are all zeros, for otherwise we would have $\operatorname{rank}(I) < n$. But this implies that some systems $A\vec{x} = \vec{e}_j$ have no solution, so there can be no matrix C such that $AC = I$.

Theorem 23 *An $n \times n$ matrix A has an inverse C if and only if rank$(A) = n$.*

Proof : ($\Leftarrow$) Suppose $\operatorname{rank}(A) = n$. Then, as proved in the last theorem, $(A|I) \sim (I|C)$. Reversing the sequence of elementary row operations, we find that $(C|I) \sim (I|A)$, which implies that $CA = I$, and therefore C is the inverse of A.

($\Leftarrow$) Suppose A has inverse C. Then, in particular, C is a right inverse of A, so $\operatorname{rank}(A) = n$ by the previous theorem.

Theorem 24 *If A and C are $n \times n$ matrices with $AC = I$, then $CA = I$.*

Proof : If $AC = I$, then A has a right inverse, and therefore its rank is n. Since $\text{rank}(A) = n$, A has a two-sided inverse, which is a right inverse in particular, and the right inverse is unique, so it must be C. Therefore, C is a two-sided inverse for A.

Before moving on, let's take a closer look at matrix/vector multiplication. If $A \in \mathbb{M}(m,n)$, then the product $A\vec{x}$ is defined for any vector $\vec{x} \in \mathbb{R}^n$. Consider the set

$$R(A) = \{\vec{y} : \vec{y} = A\vec{x} \text{ for } \vec{x} \in \mathbb{R}^n\}.$$

This is the range of the matrix A. Since $A\vec{0} = \vec{0}$, we see that $\vec{0} \in R(A)$. If $\vec{x}, \vec{y} \in R(A)$, then there are vectors $\vec{u}$ and $\vec{v}$ such that $\vec{x} = A\vec{u}$ and $\vec{y} = A\vec{v}$. Then $A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v} = \vec{x} + \vec{y}$, which implies that $R(A)$ is closed under vector addition. If $r \in \mathbb{R}$, then $A(r\vec{u}) = rA\vec{u} = r\vec{x}$, so $R(A)$ is also closed under scalar multiplication. Therefore $R(A)$ is a subspace of $\mathbb{R}^m$. Since it is a subspace, it has a basis. As we discussed previously in the chapter, one way to find a basis is to start with a spanning set and throw out unnecessary vectors until we have a basis. Notice that if we write $A = (\vec{a}_1 \ \vec{a}_2 \ \cdots \ \vec{a}_n)$ and given a vector $\vec{x} \in \mathbb{R}^n$, we have

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

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

This shows that whenever we form the matrix product $A\vec{x}$, we are actually taking a linear combination of the columns of A using the entries of $\vec{x}$ as the coefficients. This means that $R(A)$ can also be thought of as the span of the columns of A. If we need to find a basis for $R(A)$, then we can do so by reducing the spanning set (the columns of A) to a linearly independent set that still spans $R(A)$. As seen before, we accomplish this by row reducing A and finding out which columns have leading ones in RREF. We then use the corresponding columns from A as our basis for $R(A)$

Section 5.3 Markov Chains

Example: A simple migration problem: Suppose the population of a closed society is composed of a rural segment (initially 60%) and an urban segment (initially 40%). Suppose that each year 2/10 of the people on farms move to cities, with the remaining 8/10 staying on farms, and 1/10 of the people in cities move to farms, with the remaining 9/10 staying in cities.

As a table:

	From farm	From city
To farm	8/10	1/10
To city	2/10	9/10

As a matrix:

$$P = \begin{pmatrix} .8 & .1 \\ .2 & .9 \end{pmatrix}.$$

The initial distribution of the population is given by the vector

$$\vec{v}_0 = \begin{pmatrix} .6 \\ .4 \end{pmatrix}.$$

The population distribution after one year is

$$\vec{v}_1 = P\vec{v}_0 = \begin{pmatrix} .8 & .1 \\ .2 & .9 \end{pmatrix} \begin{pmatrix} .6 \\ .4 \end{pmatrix} = \begin{pmatrix} .52 \\ .48 \end{pmatrix}.$$

Similarly,

$$\vec{v}_2 = P\vec{v}_1 = \begin{pmatrix} .8 & .1 \\ .2 & .9 \end{pmatrix} \begin{pmatrix} .52 \\ .48 \end{pmatrix} = \begin{pmatrix} .464 \\ .536 \end{pmatrix}.$$

On the other hand, we can write

$$\vec{v}_2 = P\vec{v}_1 = PP\vec{v}_0 = P^2\vec{v}_0,$$

and in general, we have

$$\vec{v}_n = P\vec{v}_{n-1} = P^2\vec{v}_{n-2} = \cdots = P^n\vec{v}_0.$$

See the text for a table that shows the evolution of the population distribution. Eventually it approaches the distribution of 2/3 urban, 1/3 rural.

In the example, we can think of the percentages as probabilities that a randomly chosen individual lives in a rural or urban area.

Definition 32 *A Markov chain consists of*
1. A list of a finite number r of states,
2. A transition matrix $P = [p_{ij}] \in \mathbb{M}(r, r)$, and
3. An initial distribution vector $\vec{v}_0 \in \mathbb{R}^r$.

In our example, $r = 2$ with the states being rural and urban,

$$P = \begin{pmatrix} .8 & .1 \\ .2 & .9 \end{pmatrix} \quad \text{and} \quad \vec{v}_0 = \begin{pmatrix} .6 \\ .4 \end{pmatrix}.$$

Example: A community access television station has four sponsors: a clothing store, a hardware store, and two grocery stores. The station starts the day with an ad for one of the grocery stores (chosen by the flip of a coin). At the end of each program it runs an add for one of the sponsors. Although the station never runs an and for the same type of store twice in a row, all other sequences of ads are equally likely. Set up a Markov chain to model this situation.

The states are the adds for the different types of stores. Let's call them C, H, G_1 and G_2. The transition matrix is

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 \end{pmatrix} \quad \text{and} \quad \vec{v}_0 = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Each of the entries of the transition matrix is a probability, so $0 \leq p_{ij} \leq 1$. The jth column contains the probabilities for all possible states of the system, so

$$\sum_{i=1}^{r} p_{ij} = 1, \quad j = 1, \ldots, r.$$

Also, the entries of the initial distribution vector must add up to 1.

Theorem 25 *Suppose a Markov chain with r states has transition matrix P and initial distribution vector $\vec{v}_0$. The probability of moving from the jth state to the ith state in n steps is the ij-entry of P^n. The probability taht the system is in the ith state after n steps is the ith entry of the probability distribution vector $P^n\vec{v}_0$.*

An equilibrium of a Markov chain is a vector whose entries specify a probability distribution that will not change as the system evolves. If $\vec{s}$ is an equilibrium, then $P\vec{s} = \vec{s}$. Given our experience with the demographic model, it seems reasonable to expect that the distribution vectors might approach an equilibrium, that is $\vec{v}_n = P\vec{v}_0$ converges to $\vec{s}$ for any initial distribution vector $\vec{v}_0$. In particular, this holds true for the standard basis vectors $\vec{e}_j$ and the columns of P.

In order to be sure that a Markov chain converges to an equilibrium, there is one condition.

Definition 33 *A Markov chain is regular if and only if some power of the transition matrix has only positive entries.*

This means that after n steps, it is possible to enter any state from any starting state.

Theorem 26 *Suppose P is the transition matrix for a regular Markov chain. Then there is a unique vector $\vec{s}$ whose components add up to 1 and that satisfies $P\vec{s} = \vec{s}$. The entries of the columns of P^n converge to the corresponding entries of s as n increases to infinity.*

We can find the equilibrium by solving the system of equations $P\vec{s} = \vec{s}$.

Example: In the demographic model developed in this section, determine the equilibrium vector and the limiting matrix for P^n as n increases to infinity.

The system of equations $P\vec{s} = \vec{s}$ is equivalent to the system with augmented matrix

$$\begin{pmatrix} .8-1 & .1 \\ .2 & .9-1 \end{pmatrix} = \begin{pmatrix} -.2 & .1 \\ .2 & -.1 \end{pmatrix} \sim \begin{pmatrix} -2 & 1 \\ 2 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{pmatrix}.$$

This implies that $s_1 = \frac{1}{2}s_2$. We also have the constraint that $s_1 + s_2 = 1$. Therefore, we end up with $s_1 = \frac{1}{2}(1-s_1) \;\Rightarrow\; s_1 = \frac{1}{3}$ and $s_2 = \frac{2}{3}$. This implies that P^n converges to

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Chapter 6: Linearity

We have defined many different vector spaces, and we have made particular use of the vector spaces $\mathbb{R}^n$. In this chapter we will talk about functions that relate different vector spaces while still preserving the vector space properties and operations that are important.

Section 6.1 Linear Functions

Definition 34 *A function $T : V \to W$ from a vector space V to a vector space W is linear if and only if for all $\vec{v}, \vec{w} \in V$ and $r \in \mathbb{R}$, we have*

$$T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w}) \quad \textit{and} \quad T(r\vec{v}) = rT(\vec{v}).$$

These two conditions are known as additivity and homogeneity. The domain and range of a linear function are assumed to be vector spaces, although this is not always made explicit. Other names for a linear function are linear map and linear transformation. A linear function whose domain and range are the same is sometimes called a linear operator.

The ultimate example of a linear function is a matrix acting on column vectors. Let

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

We define a mapping by $T : \mathbb{R}^3 \to \mathbb{R}^2$ by $T(\vec{v}) = Av$, then

$$T\left(\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}\right) = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 3v_1 + v_3 \\ v_2 - v_3 \end{pmatrix}.$$

We have actually already shown that matrix multiplication is linear in Chapter 5. If $A \in \mathbb{M}(m, n)$, $\vec{u}, \vec{v} \in \mathbb{R}^n$, and $r \in \mathbb{R}$, then

$$A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v}$$

and

$$A(r\vec{u}) = rA\vec{u}.$$

Another example of a linear transformation that we have seen before is the mapping from a vector to its coordinate vector with respect to an ordered basis. Let $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ be an ordered basis for a vector space V. Then if we define $C(\vec{v}) = [\vec{v}]_B$. Then if $\vec{v}, \vec{u} \in V$, we can write

$$\vec{v} = r_1\vec{u}_1 + r_2\vec{u}_2 + \cdots + r_n\vec{u}_n \quad \text{and} \quad \vec{u} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_n\vec{u}_n.$$

$$\Rightarrow \ [\vec{v}]_B = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}, \quad [\vec{u}]_B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}.$$

Therefore

$$\vec{v} + \vec{u} = (r_1 + c_1)\vec{u}_1 + (r_2 + c_2)\vec{u}_2 + \cdots + (r_n + c_n)\vec{u}_n,$$

which implies that

$$[\vec{v} + \vec{u}]_B = \begin{pmatrix} r_1 + c_1 \\ r_2 + c_2 \\ \vdots \\ r_n + c_n \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = [\vec{v}]_B + [\vec{u}]_B$$

Also if $c \in R$, then

$$c\vec{v} = cr_1\vec{u}_1 + cr_2\vec{u}_2 + \cdots + cr_n\vec{u}_n \;\Rightarrow\; [c\vec{v}]_B = \begin{pmatrix} cr_1 \\ cr_2 \\ \vdots \\ cr_n \end{pmatrix} = c \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} = c[\vec{v}]_B.$$

Also, the inverse of this process is a linear operation. If we define a mapping $L : \mathbb{R}^n \to v$ by

$$L\left(\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}\right) = r_1\vec{u}_1 + r_2\vec{u}_2 + \cdots + r_n\vec{u}_n,$$

Then L is linear. In fact, we can write

$$L(\vec{r}) = r_1\vec{u}_1 + r_2\vec{u}_2 + \cdots + r_n\vec{u}_n = (\vec{u}_1 \;\vec{u}_2\; \cdots \;\vec{u}_n)\vec{r} = U\vec{r},$$

where U is the matrix whose columns are the vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$. Since this operation is just matrix/vector multiplication, it is linear.

Theorem 27 *Suppose $T : V \to W$ is linear. Then*
(a) $T(\vec{0}_V) = \vec{0}_W$,
(b) $T(-\vec{v}) = -T(\vec{v})$ for any $\vec{v} \in V$.
(c) $T(r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_n\vec{v}_n) = r_1T(\vec{v}_1) + r_2T(\vec{v}_2) + \cdots + r_nT(\vec{v}_n)$ for any $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \in V$ and $r_1, r_2, \ldots, r_n \in \mathbb{R}$.

Proof : (a) We have

$$T(\vec{0}_v) = T(0\vec{0}_V) = 0T(\vec{0}_V) = \vec{0}_W.$$

(b)

$$T(-\vec{v}) = T((-1)\vec{v}) = (-1)T(\vec{v}) = -T(\vec{v}).$$

(c) This follows from repeated use of the additivity and homogeneity assumptions:

$$\begin{aligned} T(r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_n\vec{v}_n) &= T(r_1\vec{v}_1) + T(r_2\vec{v}_3) + \cdots + T(r_n\vec{v}_n) \\ &= r_1T(\vec{v}_1) + r_2T(\vec{v}_2) + \cdots + r_nT(\vec{v}_n). \end{aligned}$$

The next theorem tells us that the action of a linear map is completely determined by its action on a spanning set. This means that if we want to prove something about a linear map, we can most likely prove it by just showing that it works for the spanning set.

Theorem 28 *Suppose* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ *spans a vector space* V*. Suppose* $T : V \to W$ *and* $T' : V \to W$ *are linear maps such that* $T(\vec{v}_i) = T'(\vec{v}_i)$ *for* $i = 1, 2, \ldots n$*. Then* $T = T'$*.*

Proof : To show that two functions are equal, we must show that their values are equal at every point in their common domain. Let $\vec{v} \in V$. Since the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ spans V, there are constants $r_1, r_2, \ldots, r_n$ such that

$$\vec{v} = r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_n\vec{v}_n.$$

Then

$$T(\vec{v}) = T(r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_n\vec{v}_n) = r_1T(\vec{v}_1) + r_2T(\vec{v}_2) + \cdots + r_nT(\vec{v}_n)$$
$$= r_1T'(\vec{v}_1) + r_2T'(\vec{v}_2) + \cdots + r_nT'(\vec{v}_n) = T'(r_1\vec{v}_1 + r_2\vec{v}_2 + \cdots + r_n\vec{v}_n) = T'(v).$$

Example: Suppose that the linear map $T : \mathbb{R}^3 \to \mathbb{R}^4$ satisfies

$$T(0,0,1) = \vec{0}, \quad T(0,1,1) = \vec{0}, \quad T(1,1,1) = \vec{0}.$$

Show that $T(\vec{v}) = \vec{0}$ for all vectors $\vec{v} \in \mathbb{R}^3$.

By the theorem just proved, we only have to show that

$$\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

is a spanning set. This is clear, since

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

implies that any system of equations with this coefficient matrix will have a solution. Therefore, since T agrees with the zero function on the spanning set, it agrees with it on all of $V = \mathbb{R}^3$.

Section 6.2: Compositions and Inverses

Definition 35 *Suppose* $f : X \to Y$ *is a function from the set* X *to the set* Y*. The function* f *is one-to-one if and only if for any* $x_1, x_2 \in X$*,*

$$f(x_1) = f(x_2) \quad \textit{implies} \quad x_1 = x_2.$$

The function f *is onto if and only if for every* $y \in Y$ *there is* $x \in X$ *with* $f(x) = y$*.*

The idea of a function being one-to-one means that the output of the function being the same at two points means that the two points must be the same. A logically equivalent way of saying this is

$$x_1 \neq x_1 \;\Rightarrow\; f(x_1) \neq f(x_2).$$

If $f : X \to Y$ is a function, then Y is the range. The image of f is the subset

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

of Y. For f to be onto, these two sets must be equal.

Example: Consider the linear function $T : \mathbb{R}^2 \to R^2$ defined by

$$T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 2x + y \\ -3x \end{pmatrix}.$$

Show that T is one-to-one and onto.

First, note that

$$T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 2 & 1 \\ -3 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Let's call this matrix A.

T being one-to-one is the same as saying that the equation

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

has a unique solution whenever it has a solution. Since

$$\begin{pmatrix} 2 & 1 \\ -3 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{3}{2} \end{pmatrix} \sim \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

we see that there are no free variables in this system, so the solution is unique. To show that T is onto, we must show that the equation $A\vec{x} = \vec{b}$ has a solution regardless of how we choose $\vec{b}$. But this is true since there are no rows of zeros in the RREF of A.

Note: When the linear mapping T is defined by multiplication with a matrix, we can reformulate the questions about T being one-to-one and onto by asking if $A\vec{x} = \vec{b}$ always has a solution, and if that solution is unique.

Definition 36 *Consider two functions $f : X \to Y$ and $g : Y \to Z$. The composition of f followed by the function g is the function $g \circ f : X \to Z$ defined by $(g \circ f)(x) = g(f(x))$. The identity function on any set X is the function $id_X : X \to X$ defined by $id_X(x) = x$. A function $g : Y \to X$ is an inverse of the function $f : X \to Y$ if*

$$g \circ f = id_X \quad \textit{and} \quad f \circ g = id_Y.$$

Note: Even when both compositions are defined, the functions $g \circ f$ and $f \circ g$ need not be equal. For example,

$$\sin(x^2) \neq \sin^2 x.$$

If a function f has an inverse, that inverse must be unique. The inverse of a function f is denoted by f^{-1}.

Theorem 29 *A function $f : X \to Y$ has an inverse if and only if f is one-to-one and onto.*

Proof : ($\Rightarrow$) Suppose that g is an inverse of f. If $f(x_1) = f(x_2)$, then

$$x_1 = g(f(x_1)) = g(f(x_2)) = x_2,$$

which implies that f is one-to-one. Let $y \in Y$. Then $g(y)$ is mapped to Y, since

$$f(g(y)) = y.$$

($\Leftarrow$) Suppose that f is one-to-one and onto. Then we can define a function $g : Y \to X$ by $g(y) = x$ if and only if $f(x) = y$. Because f is one-to-one and onto, no matter how y is chosen, there is a unique x such that $f(x) = y$, so g is well defined.

Theorem 30 *Suppose $S : U \to V$ and $T : V \to W$ are linear maps. Then the composition $T \circ S : U \to W$ is linear.*

Proof : Let $\vec{x}, \vec{y} \in U$. Then

$$(T \circ S)(\vec{x} + \vec{y}) = T(S((\vec{x} + \vec{y})) = T(S(\vec{x}) + S(\vec{y}))$$

$$= T(S(\vec{x})) + T(S(\vec{y})) = (T \circ S)(\vec{x}) + (T \circ S)(\vec{y}).$$

Also, if $r \in R$, then

$$(T \circ S)(r\vec{x}) = T(S(r\vec{x})) = T(rS(\vec{x})) = rT(S(\vec{x})) = r(T \circ S)(\vec{x}).$$

Thus $T \circ S$ is linear.

Theorem 31 *Suppose the linear function $T : V \to W$ has an inverse function. Then $T^{-1} : W \to V$ is linear.*

Proof : Let $\vec{w}_1, \vec{w}_2 \in W$ and $r \in \mathbb{R}$. Define $\vec{v}_1 = T^{-1}(\vec{w}_1)$ and $v_2 = T^{-1}(\vec{w}_2)$. Then

$$T(\vec{v}_1) = T(T^{-1}(\vec{w}_1) = \vec{w}_1 \text{ and } T(\vec{v}_2) = T(T^{-1}(\vec{w}_2)) = \vec{w}_2.$$

Therefore,

$$T^{-1}(\vec{w}_1 + \vec{w}_2) = T^{-1}(T(\vec{v}_1) + T(\vec{v}_2)) = T^{-1}(T(\vec{v}_1 + \vec{v}_2))$$

$$= \vec{v}_1 + \vec{v}_2 = T^{-1}(\vec{w}_1) + T^{-1}(\vec{w}_2)$$

and

$$T^{-1}(r\vec{w}_1) = T^{-1}(rT(\vec{v}_1)) = T^{-1}(T(r\vec{v}_1)) = r\vec{v}_1 = rT^{-1}(\vec{w}_1).$$

Theorem 32 *Suppose $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis for a vector space V. Then for any elements $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n$ of a vector space W, there is a unique linear map $T : V \to W$ such that $T(\vec{v}_i) = \vec{w}_i$ for $i = 1, 2, \ldots, n$.*

Proof : Define a function $T : V \to W$ by

$$T(\vec{v}) = r_1\vec{w}_1 + r_2\vec{w}_2 + \cdots + r_n\vec{w}_n,$$

where $r_1, r_2, \ldots, r_n$ are the coordinates of $\vec{v}$ with respect to the basis B. Define $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$. Then $T(\vec{v}) = L_{B'}([\vec{v}]_B)$, where $L_{B'}$ was defined in Section 6.1. Since T is a composition of linear functions, it is linear. Also,

$$T(\vec{v}_i) = L_{B'}([\vec{v}_i]_B) = L_{B'}(e_i) = 0\vec{w}_1 + \cdots + 1\vec{w}_i + \cdots + 0\vec{w}_n = \vec{w}_i.$$

We proved previously that a linear map's action on a spanning set is unique.

Section 6.3: Matrix of a Linear Function

Example: Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map with

$$T\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 9 \\ 4 \end{pmatrix} \text{ and } T\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 2 \\ -5 \end{pmatrix}.$$

Compute $T((5, 6))$ and give a formula for $T((a, b))$ in terms of matrix multiplication.

We have

$$T\left(\begin{pmatrix} 5 \\ 6 \end{pmatrix}\right) = T\left(5\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 6\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = 5T\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + 6T\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$$

$$= 5\begin{pmatrix} 9 \\ 4 \end{pmatrix} + 6\begin{pmatrix} 2 \\ -5 \end{pmatrix} = \begin{pmatrix} 57 \\ -10 \end{pmatrix}.$$

In general,

$$T\left(\begin{pmatrix} a \\ b \end{pmatrix}\right) = T\left(a\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = aT\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + bT\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$$

$$= a\begin{pmatrix} 9 \\ 4 \end{pmatrix} + b\begin{pmatrix} 2 \\ -5 \end{pmatrix} = \begin{pmatrix} 9a + 2b \\ 4a - 5b \end{pmatrix}.$$

In terms of matrix multiplication, we can write

$$T\left(\begin{pmatrix} a \\ b \end{pmatrix}\right) = \begin{pmatrix} 9 & 2 \\ 4 & -5 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}.$$

Example: Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the function that rotates each point in $\mathbb{R}^2$ counterclockwise about the origin through an angle of θ. Compute $R_\theta((1,0))$ and $R_\theta((0,1))$. Find a formula for $R_\theta((a,b))$ in terms of matrix multiplication.

Simple geometry tells us that

$$R_\theta\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \quad \text{and} \quad R_\theta\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

Therefore,

$$R_\theta\left(\begin{pmatrix} a \\ b \end{pmatrix}\right) = R_\theta\left(a\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = aR_\theta\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + bR_\theta\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$$

$$= a\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} + b\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = \begin{pmatrix} a\cos\theta - b\sin\theta \\ a\sin\theta + b\cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}$$

Theorem 33 *Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear. Let $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ be the standard basis for $\mathbb{R}^n$. There is a unique $m \times n$ matrix A such that $T(\vec{v}) = A\vec{v}$ for all $\vec{v} \in \mathbb{R}^n$. The columns of this matrix are the m-tuples $T(\vec{e}_1), T(\vec{e}_2), \ldots, T(\vec{e}_n)$ in that order.*

Proof : Since $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ is a basis, there are constants $r_1, r_2, \ldots, r_n$ such that

$$\vec{v} = r_1\vec{e}_1 + r_2\vec{e}_2 + \cdots + r_n\vec{e}_n.$$

Therefore,

$$\begin{aligned} T(\vec{v}) = T(r_1\vec{e}_1 + r_2\vec{e}_2 + \cdots + r_n\vec{e}_n) &= r_1T(\vec{e}_1) + r_2T(\vec{e}_2) + \cdots + r_nT(\vec{e}_n) \\ &= (T(\vec{e}_1)\ T(\vec{e}_2)\ \cdots\ T(\vec{e}_n)) \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}. \end{aligned}$$

Therefore, the columns of A are the vectors $T(\vec{e}_1), T(\vec{e}_2), \ldots, T(\vec{e}_n)$.

It is nice to know that we can represent linear functions between Euclidean spaces as a matrix, but where this becomes especially useful is when we are working with finite dimensional subspaces of vector spaces that are not Euclidean spaces. As we have seen, we can represent vectors in these spaces in terms of a coordinate vector once we have specified a basis.

Example: The function $T : \mathbb{P}_2 \to \mathbb{P}_3$ defined by $T(ax^2+bx+c) = a(x-1)^2 + b(x-1) + c$ is linear. $B = \{1, x, x^2\}$ is a basis for $\mathbb{P}_2$ and $B' = \{1, x, x^2, x^3\}$ is a basis for $\mathbb{P}_3$. We want a matrix A such that for any $p \in \mathbb{P}_2$ we have

$$[T(p)]_{B'} = A[p]_B.$$

$$A\vec{e}_1 = A[1]_B = [T(1)]_{B'} = [1]_{B'} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$A\vec{e}_2 = A[x]_B = [T(x)]_{B'} = [x-1]_{B'} = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

$$A\vec{e}_3 = A[x^2]_B = [T(x^2)]_{B'} = [(x-1)^2]_{B'} = [x^2 - 2x + 1]_{B'} = \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}.$$

This shows that

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

Theorem 34 *Suppose $B = \{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_n\}$ is a basis for a vector space V and $B' = \{\vec{u}'_1, \vec{u}'_2, \ldots, u'_m\}$ is a basis for a vector space V'. Suppose $T : V \to V'$ is linear. There is a unique $m \times n$ matrix A such that $[T(\vec{v})]_{B'} = A[\vec{v}]_B$ for all $\vec{v} \in V$. The columns of this matrix are the m-tuples $[T(\vec{u}_1)]_{B'}, [T(\vec{u}_2)]_{B'}, \ldots, [T(\vec{u}_n)]_{B'}$ in that order.*

Proof : We have that

$$A\vec{e}_j = A[\vec{u}_j]_B = [T(\vec{u}_j)]_{B'},$$

so the matrix A is determined uniquely. Let A be the matrix

$$A = \left([T(\vec{u}_1)]_{B'} \; [T(\vec{u}_2)]_{B'} \; \cdots \; [T(\vec{u}_n)]_{B'}\right).$$

Then

$$[T(\vec{u}_j)]_{B'} = A\vec{e}_j = A[\vec{u}_j]_B, \quad j = 1, 2, \ldots, n.$$

This means that the two linear functions $[T(\cdot)]_{B'}$ and $A[\cdot]_B$ agree on the basis B. Therefore, they must be equal.

Definition 37 *With the notation as in the previous theorem, A is the matrix of T relative to the bases B and B'.*

Section 6.4: Matrices of Compositions and Inverses

We know that functions that are the composition of linear functions are linear and that the inverse of a linear function is linear. The next question we need to answer is how the matrices representing composites and inverses are related to the matrices representing the functions from which they are built.

Theorem 35 *Suppose $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is a basis for a vector space V, and $B' = \{\vec{u}'_1, \vec{u}'_2, \ldots, \vec{u}'_m\}$ is a basis for a vector space V', and $B'' = \{\vec{u}''_1, \vec{u}''_2, \ldots, \vec{u}''_l\}$ is a basis for a vector space V''. Suppose A' is the matrix of a linear map $T : V \to V'$ relative to the bases B and B'. Suppose A' is the matrix of a linear map $T' : V' \to V''$ relative to the bases B' and B''. Then $A'A$ is the matrix of the linear map $T \circ T : V \to V''$ relative to the bases B and B''.*

Proof : We know that

$$[T(\vec{v})]_{b'} = A[\vec{v}]_B \text{ for all } v \in V$$

and

$$[T'(\vec{v'})]_{B''} = A'[\vec{v'}]_{B'} \text{ for all } v' \in V'.$$

Therefore,

$$[(T' \circ T)(\vec{v})]_{B''} = [T'(T(\vec{v}))]_{B''} = A'[T(\vec{v})]_{B'} = A'(A[\vec{v}]_B) = A'A[\vec{v}]_B.$$

This shows why matrix multiplication is defined the way that it is–because when defined this way, it represents a composition of linear functions.

Example: Relative to the bases $\{1, x, x^2\}$ for $\mathbb{P}_2$ and $\{1, x, x^2, x^3\}$ for $\mathbb{P}_3$, we know from work in the previous section that $T : \mathbb{P}_2 \to \mathbb{P}_3$ defined by $T(ax^2 + bx + c) = a(x-1)^2 + b(x-1) + c$ has matrix

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

and the differentiation operator $D : \mathbb{P}_3 \to \mathbb{P}_3$ has matrix

$$A' = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Compute the matrix for the composition $D \circ T$ directly and verify that it equals the product $A'A$.

First, we'll compute the matrix of the composition. We do this by computing its action on each of the vectors in the basis $\{1, x, x^2\}$, and then find the coordinate vector of the result with respect to the basis $B = \{1, x, x^2, x^3\}$. We have that

$$D(T(1)) = D(1) = 0, \quad [0]_B = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$D(T(x)) = D(x-1) = 1, \quad [1]_B = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

and

$$D(T(x^2)) = D((x-1)^2) = D(x^2 - 2x + 1) = 2x - 2, \quad [2x-2]_B = \begin{pmatrix} -2 \\ 2 \\ 0 \\ 0 \end{pmatrix}.$$

Therefore, the matrix is

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

Also, we see that

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

It should be clear that we can reverse this process, and we can construct a mapping between any vector spaces V of dimension n and V'' of dimension m using a $m \times n$ matrix A by the formula

$$T(\vec{v}) = L_{B'}(A[\vec{v}]_B),$$

where $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is a basis for V, $B' = \{\vec{u}'_1, \vec{u}'_2, \ldots, \vec{u}'_m\}$ is a basis for V' and $L_{B'}$ is the function that uses the entries of an element of $\mathbb{R}^m$ as the coefficients of the elements of B'. Because each of the maps in the composition is linear, L is linear. Since $L_{B'}$ is the inverse of the coordinate vector function $[\cdot]_{B'}$,

$$[T(\vec{v})]_{B'} = [L_{B'}A([\vec{v}]_B)]_{B'} = A[\vec{v}]_B.$$

This shows that A is the matrix of T relative to the bases B and B'.

Theorem 36 *Suppose $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is a basis for a vector space V and $B' = \{\vec{u}'_1, \vec{u}'_2, \ldots, \vec{u}'_m\}$ is a basis for a vector space $\vec{V}'$. Suppose A is the matrix of a linear map $T : V \to V'$ relative to the bases B and B'. Then T has an inverse function if and only if A is an invertible matrix. In this case, A^{-1} is the matrix of T^{-1} relative to the bases B' and B.*

Proof : ($\Rightarrow$) Suppose that T has an inverse, and let $A' \in \mathbb{M}(n,m)$ be the matrix of T^{-1} relative to the bases B' and B. Then $A'A$ is the matrix of $T^{-1} \circ T = \mathrm{id}_V$ relative to B, and therefore, $A'A = I$. Similarly, the matrix for $T \circ T^{-1}$ is $AA' = I$. This shows that A is invertible and $A^{-1} = A'$.

($\Leftarrow$) Suppose A is invertible. Let $T' : V' \to V$ be the map with matrix A^{-1}. For $v \in V$,

$$[T'(T(\vec{v}))]_B = A^{-1}[T(\vec{v})]_{B'} = A^{-1}A[\vec{v}]_B = [\vec{v}]_B,$$

and for any $\vec{v}' \in V'$ we have

$$[T(T'(\vec{v}'))]_{B'} = A[T'(\vec{v}')]_B = AA^{-1}[\vec{v}']_{B'} = [\vec{v}']_{B'}.$$

The coordinate functions $[\cdot]_B$ and $[\cdot]_{B'}$ are one-to-one, so $T'(T(\vec{v})) = \vec{v}$ for all $\vec{v} \in V$ and $T(T'(\vec{v}')) = \vec{v}'$ for all $\vec{v}' \in V'$.

Example: The matrix of the linear function $T : \mathbb{P}_3 \to \mathbb{P}_3$ defined by $T(p(x)) = p(2x+1)$ with respect to the basis $B = \{1, x, x^2, x^3\}$ is

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 4 & 6 \\ 0 & 0 & 5 & 12 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

The inverse of T is defined by $T^{-1}(p(x)) = p(\frac{1}{2}(x-1))$. Find the matrix A' of T^{-1} relative to B and show that this matrix is the inverse of A.

To find A', we calculate the action of T^{-1} on each of the elements of B, then compute the coordinates of the results relative to B. We have that

$$T^{-1}(1) = 1,\ T^{-1}(x) = \frac{1}{2}(x-1),\ T^{-1}(x^2) = \left(\frac{1}{2}(x-1)\right)^2 = \frac{1}{4}(x^2-2x+1) = \frac{1}{4}x^2-\frac{1}{2}x+\frac{1}{4},$$

$$T^{-1}(x^3) = \left(\frac{1}{2}(x-1)\right)^3 = \frac{1}{8}(x^3-3x^2+3x-1) = \frac{1}{8}x^3 - \frac{3}{8}x^2 + \frac{3}{8}x - \frac{1}{8}.$$

Therefore,

$$[T^{-1}(1)]_B = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},\ [T^{-1}(x)]_B = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 0 \end{pmatrix},\ [T^{-1}(x^2)]_B = \begin{pmatrix} \frac{1}{4} \\ -\frac{1}{2} \\ \frac{1}{4} \\ 0 \end{pmatrix},\ [T^{-1}(x^3)]_B = \begin{pmatrix} -\frac{1}{8} \\ \frac{3}{8} \\ -\frac{3}{8} \\ \frac{1}{8} \end{pmatrix},$$

which implies that

$$A' = \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{8} \\ 0 & \frac{1}{2} & -\frac{1}{2} & \frac{3}{8} \\ 0 & 0 & \frac{1}{4} & -\frac{3}{8} \\ 0 & 0 & 0 & \frac{1}{8} \end{pmatrix}.$$

Simple calculation shows that

$$AA' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 4 & 6 \\ 0 & 0 & 5 & 12 \\ 0 & 0 & 0 & 8 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{8} \\ 0 & \frac{1}{2} & -\frac{1}{2} & \frac{3}{8} \\ 0 & 0 & \frac{1}{4} & -\frac{3}{8} \\ 0 & 0 & 0 & \frac{1}{8} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

This shows that $A' = A^{-1}$.

Section 6.5: Change of Basis

At this point, it should be clear that when we find the matrix that represents a linear function, the result that we get depends on the bases that we use to represent the vectors in the domain and range spaces of the linear function.

In the special case where $T : V \to V$, i.e., the domain and range spaces are the same, it is convenient to use the same basis to represent V as both the domain and range space.

The obvious question to ask is, given two bases B and B' for V, how are the matrices representing T with respect to these two bases related?

Theorem 37 *Suppose $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ and $B' = \{\vec{u}'_1, \vec{u}'_2, \ldots, \vec{u}'_n\}$ are bases for a vector space V. There is a unique $n \times n$ matrix P such that $[\vec{v}]_B = P[\vec{v}]_{B'}$ for all $\vec{v} \in V$. The columns of this matrix are the n-tuples $[\vec{u}'_1]_B, [\vec{u}'_2]_B, \ldots, [\vec{u}'_n]_B$ in that order.*

Proof : Let P be the matrix of the identity map $\mathrm{id}_V : V \to V$ relative to the bases B' and B. Then the jth column of P is $[\mathrm{id}_V(\vec{u}'_j)]_B = [\vec{u}'_j]_B$. As we proved before, P is the unique matrix satisfying $[\vec{v}]_B = [\mathrm{id}_V(\vec{v})]_B = P[\vec{v}]_{B'}$ for all $\vec{v} \in V$.

Definition 38 *The matrix P as described in the previous theorem is the change of basis matrix from the basis B' to the basis B.*

Example: Find the change of basis matrix P for changing from the basis

$$B' = \left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ -1 \end{pmatrix} \right\} \text{ to the basis } B = \left\{ \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}$$

in $\mathbb{R}^2$. Verify directly that

$$P\begin{pmatrix} 3 \\ -2 \end{pmatrix}_{B'} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}_B.$$

As directed in the theorem, we compute

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

These are the solutions to the systems of equations

$$\left(\begin{array}{cc|cc} 2 & 0 & 2 & 4 \\ 0 & -1 & 2 & -1 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 1 \end{array}\right) \Rightarrow \begin{pmatrix} 2 \\ 2 \end{pmatrix}_B = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

$$\text{and} \quad \begin{pmatrix} 4 \\ -1 \end{pmatrix}_B = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Therefore,

$$P = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}.$$

We have that

$$\left(\begin{array}{cc|c} 2 & 4 & 3 \\ 2 & -1 & -2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2 & \frac{3}{2} \\ 0 & -5 & -5 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 1 \end{array}\right) \Rightarrow \begin{pmatrix} 3 \\ -2 \end{pmatrix}_{B'} = \begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix}$$

and

$$\left(\begin{array}{cc|c} 2 & 0 & 3 \\ 0 & -1 & -2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & \frac{3}{2} \\ 0 & 1 & 2 \end{array}\right) \Rightarrow \begin{pmatrix} 3 \\ -2 \end{pmatrix}_B = \begin{pmatrix} \frac{3}{2} \\ 2 \end{pmatrix}.$$

We see that

$$P\begin{pmatrix} 3 \\ -2 \end{pmatrix}_{B'} = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}_B.$$

Theorem 38 *Suppose P is the change of basis matrix for changing from a basis B' to a basis B. Then P is invertible, and P^{-1} is the change of basis matrix for changing from the basis B to the basis B'.*

Proof : Suppose we have a linear function $T : \mathbb{R}^2 \to \mathbb{R}^2$ whose matrix relative to a basis B is given. If we want to find the matrix of T relative to another basis B', then we can think of T as

$$\mathrm{id}_V \circ T \circ \mathrm{id}_V.$$

Recall that if P is the matrix representation of the identity function relative to the bases B' and B, then P^{-1} is the matrix of id_V relative to B and B'. Since composition of functions corresponds to matrix multiplication of their matrix representations, we have that $P^{-1}AP$ is the representation of T relative to the bases B' and B.

Example: Suppose the matrix of a linear function T relative to the basis B from the last example is

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

Verify that

$$\left[T\left(\begin{pmatrix} 3 \\ -2 \end{pmatrix} \right) \right]_{B'} = P^{-1}AP \begin{pmatrix} 3 \\ -2 \end{pmatrix}_{B'}.$$

First, we compute

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ -2 & 1 & 0 & 1 \end{array} \right) \sim \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 5 & 2 & 1 \end{array} \right) \sim \left(\begin{array}{cc|cc} 1 & 0 & \frac{1}{5} & -\frac{2}{5} \\ 0 & 1 & \frac{2}{5} & \frac{1}{5} \end{array} \right),$$

so

$$P^{-1} = \begin{pmatrix} \frac{1}{5} & -\frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{pmatrix}.$$

Therefore,

$$P^{-1}AP \begin{pmatrix} 3 \\ -2 \end{pmatrix}_{B'} = \begin{pmatrix} \frac{1}{5} & -\frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} 1 & -4 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{5} & -\frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} \frac{3}{2} \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} & -\frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} -\frac{13}{2} \\ \frac{17}{2} \end{pmatrix} = \begin{pmatrix} -\frac{47}{10} \\ -\frac{9}{10} \end{pmatrix}.$$

We also have that

$$\left[T\left(\begin{pmatrix} 3 \\ -2 \end{pmatrix} \right) \right]_{B} = \begin{pmatrix} 1 & -4 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} \frac{3}{2} \\ 2 \end{pmatrix} = \begin{pmatrix} -\frac{13}{2} \\ \frac{17}{2} \end{pmatrix},$$

and therefore,

$$T\left(\begin{pmatrix} 3 \\ 2 \end{pmatrix} \right) = -\frac{13}{2} \begin{pmatrix} 2 \\ 0 \end{pmatrix} - \frac{17}{2} \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -13 \\ -\frac{17}{2} \end{pmatrix}.$$

Since

$$-\frac{47}{10} \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \frac{9}{10} \begin{pmatrix} 4 \\ -1 \end{pmatrix} = \begin{pmatrix} -13 \\ -\frac{17}{2} \end{pmatrix}$$

we see that

$$\left[T\left(\begin{pmatrix} 3 \\ -2 \end{pmatrix} \right) \right]_{B'} = \begin{pmatrix} -\frac{47}{10} \\ -\frac{9}{10} \end{pmatrix}.$$

Theorem 39 *Suppose $B = \{\vec{u}_1, \ldots, \vec{u}_n\}$ and $B' = \{\vec{u}'_1, \ldots, \vec{u}'_n\}$ are bases for a vector space V. Let P be the change of basis matrix from B' to B. Suppose A is the matrix of a linear map $T : V \to V$ relative to the basis B. Then $P^{-1}AP$ is the matrix of T relative to the basis B'.*

Proof : By the definitions of the matrix of T and the change of basis matrix P, we know that

$$[T(\vec{v})]_B = A[\vec{v}]_B \text{ and } [\vec{v}]_B = P[\vec{v}]_{B'}$$

for all $\vec{v} \in V$. From these two equations, we find that

$$[T(\vec{v})]_{B'} = P^{-1}[T(\vec{v})]_B.$$

Therefore,

$$[T(\vec{v})]_{B'} = P^{-1}[T(\vec{v})]_B = P^{-1}A[\vec{v}]_B = P^{-1}AP[\vec{v}]_{B'},$$

as desired.

Section 6.6: Image and Kernel

We spent a lot of time in Chapter 3 studying vector spaces and subspaces. In this chapter, we will study two important subspaces associated with every linear function.

Definition 39 *Suppose $T : V \to W$ is linear. The kernel (or null space) of T is denoted ker(T) and is defined by*

$$ker(T) = \{\vec{v} \in V : T(\vec{v}) = \vec{0}\}.$$

The image of T is denoted im(T) and is defined by

$$im(T) = \{\vec{w} \in W : \vec{w} = T(\vec{v}) \text{ for some } \vec{v} \in V\}.$$

Theorem 40 *Suppose $T : V \to W$ is linear. Then ker(T) is a subspace of V and im(T) is a subspace of W.*

The proof of this theorem is a straightforward application of the subspace theorem.

Example: Consider the matrix

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

Find bases for the kernel and image of the associated linear map $\mu(\vec{x}) = A\vec{x}$.

At this point, it should not be surprising that we can find out all the information that we need to know about the matrix A by looking at its reduced row echelon form:

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

Let's tackle the image first. Recall that by the definition of matrix-vector multiplication, the product $A\vec{x}$ is a linear combination of the columns of A using the elements of $\vec{x}$ as the coefficients. Therefore, the image of A is the same as the span of the columns of A. A theorem that we have previously proved gave us an algorithm for finding a basis from a spanning set: we throw out the vectors that can be written as a linear combination of the other vectors. In the case of A, the RREF tells us that the third column can be written as a linear combination of the first two, but that the first two columns are not linear combinations of one another. Therefore, a basis for $\mathrm{im}(\mu)$ is

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

Notice that $\ker(A)$ is exactly the solution set to the homogeneous equation $A\vec{x} = \vec{0}$. The RREF of A tells us that x_3 is a free variable. If we let $x_3 = r$, then the solution to $A\vec{x} = \vec{0}$ is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}r \\ -\frac{3}{2}r \\ r \end{pmatrix} = r \begin{pmatrix} \frac{1}{2} \\ -\frac{3}{2} \\ 1 \end{pmatrix}.$$

In other words, $\ker(A) = \mathrm{span}\{(1/2, -3/2, 1)\}$. Since this spanning set consists of only one vector, it is also linearly independent, so $\{(1/2, -3/2, 1)\}$ is a basis for $\ker(A)$.

Definition 40 *The column space of an $m \times n$ matrix A is the subspace of $\mathbb{R}^m$ spanned by the n columns of A considered as elements of $\mathbb{R}^m$.*

Definition 41 *The row space of an $n \times n$ matrix A is the subspace of $\mathbb{R}^n$ spanned by the m rows of A considered as elements of R^n.*

Theorem 41 *Row operations do not change the row space of a matrix.*

Proof : Let $\vec{v}_1, \ldots, \vec{v}_m$ be the rows of an $m \times n$ matrix A. Then any element of the row space of A can be expressed as

$$r_1\vec{v}_1 + \cdots + r_m\vec{v}_m$$

for some choice of scalars $r_1, \ldots, r_m \in \mathbb{R}$. If we interchange two rows of A, then all that does is change the order in which the columns of A are written; it does not change the span. If row i is multiplied by a nonzero constant c, then

$$r_1\vec{v}_1 + \cdots + r_i\vec{v}_i + \cdots + r_m\vec{v}_m = r_1\vec{v}_1 + \cdots + \frac{r_i}{c}(c\vec{v}_i) + \cdots + r_m\vec{v}_m,$$

so any vector that was in $\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_n\}$ is in $\operatorname{span}\{\vec{v}_1, \ldots, c\vec{v}_i, \ldots, \vec{v}_m\}$. If we add c times row i to row j, then

$$r_1\vec{v}_1+\cdots+r_i\vec{v}_i+\cdots+r_j\vec{v}_j+\cdots+r_m\vec{v}_m = r_1\vec{v}_1+\cdots+(r_i-cr_j)\vec{v}_i+\cdots+r_j(c\vec{v}_i+\vec{v}_j)+\cdots+r_m\vec{v}_m,$$

so $\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_m\} = \operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_1, \ldots, (c\vec{v}_i + \vec{v}_j), \ldots, \vec{v}_m\}$. This shows that if a matrix B can be obtained from A by elementary row operations, then the row spaces are the same.

This gives us a simple method for finding a basis for the row space of a matrix. We reduce the matrix to RREF. In doing so, we do not change the row space. But it is easy to find a basis for the row space of a matrix in RREF. Either the rows are all zero, or there is a leading one. The rows with leading ones are linearly independent, since the entries above and below each leading one are zero. Therefore, we obtain a basis for the row space of a matrix from the nonzero rows of the RREF of the matrix.

Example: Find bases for the row space and the column space of the matrix

$$A = \begin{pmatrix} 2 & -2 & -1 & -2 & 8 \\ -4 & 4 & -1 & -3 & -8 \\ 1 & -1 & -1 & -2 & 5 \\ 3 & -3 & -2 & 1 & 3 \end{pmatrix}.$$

First, we row reduce A:

$$\begin{pmatrix} 2 & -2 & -1 & -2 & 8 \\ -4 & 4 & -1 & -3 & -8 \\ 1 & -1 & -1 & -2 & 5 \\ 3 & -3 & -2 & 1 & 3 \end{pmatrix} \sim \begin{pmatrix} 1 & -1 & -1 & -2 & 5 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & -5 & -11 & 12 \\ 0 & 0 & 1 & 7 & -12 \end{pmatrix} \sim \begin{pmatrix} 1 & -1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 5 & -10 \end{pmatrix}$$

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

Therefore, a basis for $\text{im}(A)$ is

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

and a basis for the row space of A is

$$\{(1,-1,0,0,3),(0,0,1,0,2),(0,0,0,1,-2)\}.$$

Theorem 42 *The dimension of the row space of any matrix is equal to the dimension of its column space.*

Proof : Both are determined by the number of leading ones in the RREF of the matrix.

Section 6.7: Rank and Nullity

Definition 42 *Suppose $T : V \to W$ is linear. If ker(T) is a finite-dimensional subspace of V, then the dimension of ker(T) is called the nullity of T. If im(T) is a finite-dimensional subspace of W, then the dimension of im(T) is called the rank of T.*

Notice that the rank of a matrix was defined previously to be the number of leading ones in the RREF of the matrix. In the case that the linear transformation is defined by matrix-vector multiplication, these two ideas of rank agree.

Theorem 43 (Dimension Theorem) *Suppose $T : V \to W$ is a linear map whose domain V is a finite-dimensional vector space. Then*

$$\mathit{rank}(T) + \mathit{nullity}(T) = \mathit{dim}(V).$$

Proof : Since $\ker(T)$ is finite dimensional, it has a basis, say $\{\vec{v}_1, \dots, \vec{v}_m\}$. By the expansion theorem, there are vectors $\vec{v}_{m+1}, \dots, \vec{v}_{m+k}$ in V so that

$$\{\vec{v}_1, \dots, \vec{v}_{m+k}\}$$

is a basis for V. We claim that $\{T(\vec{v}_{m+1}, \ldots, T(\vec{v}_{m+k})\}$ is a basis for $\mathrm{im}(T)$. Let $\vec{w} \in \mathrm{im}(T)$. Then there is a $\vec{v} \in V$ such that $\vec{w} = T(\vec{v})$. There are coefficients $r_1, \ldots, r_{m+k}$ such that

$$\vec{v} = r_1\vec{v}_1 + \cdots + r_{m+k}\vec{v}_{m+k}.$$

Therefore,

$$\vec{w} = T(\vec{v}) = r_1T(\vec{v}_1) + \cdots + r_{m+k}T(\vec{v}_{m+k}) = r_{m+1}T(\vec{v}_{m+1}) + \cdots + r_{m+k}T(\vec{v}_{m+k}).$$

This shows that $\{T(\vec{v}_{m+1}), \ldots, T(\vec{v}_{m+k})\}$ spans $\mathrm{im}(T)$. Suppose that

$$r_{m+1}T(\vec{v}_{m+1}) + \cdots + r_{m+k}T(\vec{v}_{m+k}) = \vec{0} \quad \Rightarrow \quad T(r_{m+1}\vec{v}_{m+1} + \cdots + r_{m+k}\vec{v}_{m+k}) = \vec{0}$$

$$\Rightarrow \; r_{m+1}\vec{v}_{m+1} + \cdots + r_{m+k}\vec{v}_{m+k} \in \ker(T).$$

This implies that there exists scalars $r_1, \ldots, r_m$ so that

$$r_{m+1}\vec{v}_{m+1} + \cdots + r_{m+k}\vec{v}_{m+k} = r_1\vec{v}_1 + \cdots + r_m\vec{v}_m,$$

or in other words,

$$-r_1\vec{v}_1 - \cdots - r_m\vec{v}_m + r_{m+1}\vec{v}_{m+1} + \cdots + r_{m+k}\vec{v}_{m+k} = \vec{0}.$$

Since the vectors in the linear combination above are linearly independent, $r_{m+1} = \cdots = r_{m+k} = 0$, which implies that the set $\{T(\vec{v}_{m+1}), \ldots, T(\vec{v}_{m+k})\}$ is linearly independent, and therefore the rank of T is k and the nullity is m, and $k + m$ is the dimension of V. In other words,

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = k + m = \dim(V).$$

Theorem 44 *A linear map $T : V \to W$ is one-to-one if and only if ker$(T) = \{\vec{0}\}$.*

Proof : $(\Rightarrow)$ Assume T is one-to-one, and let $\vec{v} \in \ker(T)$. Then $T(\vec{v}) = \vec{0}$, but we also know that $T(\vec{0}) = \vec{0}$, and therefore $\vec{0} = \vec{v}$.

$(\Leftarrow)$ Suppose that $\ker(T) = \{\vec{0}\}$. Take two vectors $\vec{v}_1, \vec{v}_2 \in V$ such that $T(\vec{v}_1) = T(\vec{v}_2)$. Then $T(\vec{v}_1 - \vec{v}_2) = \vec{0}$, which implies that $\vec{v}_1 - \vec{v}_2 \in \ker(T)$, so it must be that $\vec{v}_1 - \vec{v}_2 = \vec{0}$, or $\vec{v}_1 = \vec{v}_2$. This shows that T is one-to-one.

Example: Prove that a linear map $T : \mathbb{R}^3 \to \mathbb{R}^2$ cannot be one-to-one.

If T is one-to-one, then $\mathrm{nullity}(T) = 0$, which means that $\mathrm{rank}(T) = 3$. In other words, $\mathrm{im}(T)$ is a 3-dimensional subspace of a 2-dimensional vector space ($\mathbb{R}^2$). But since this is impossible, it must be that T is not one-to-one.

Example: Prove that a linear map $T : \mathbb{M}(2,2) \to \mathbb{P}_4$ cannot be onto.

For T to be onto would mean that $\text{im}(T) = \mathbb{P}_4$, which has dimension 5. The dimension of $\mathbb{M}(2,2)$ is 4, so we must have that

$$\text{nullity}(T) + 5 = 4,$$

which is impossible.

Example: Suppose a linear function $T : \mathbb{P}_4 \to \mathbb{R}^5$ is onto. Prove that T is one-to-one.

If T is onto, then $\text{rank}(T) = 5$. Therefore,

$$\text{nullity}(T) + 5 = \dim(\mathbb{P}_4) = 5,$$

so $\text{nullity}(T) = 0$, or in other words, $\ker(T) = \{\vec{0}\}$, which implies that T is one-to-one.

A good approach to showing that a linear function T is one-to-one and onto is to first show that $\ker(T) = \{\vec{0}\}$ (which involves solving the homogeneous equation $T(\vec{v}) = \vec{0}$), and then using the rank-nullity theorem.

Section 6.8: Isomorphism

We have now seen several examples of situations where two vector spaces are essentially the same, such as the set

$$\{\vec{x} \in \mathbb{R}^3 : x_3 = 0\}$$

and $\mathbb{R}^2$. Another example is a finite dimensional vector space V and its image under the coordinate transformation relative to a given basis.

Consider the example where we pair the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{M}(2,2)$$

with the polynomial $ax^3 + bx^2 + cx + d \in \mathbb{P}_3$. Let's call this mapping T. Then

1. T is a well-defined function.
2. T is defined on all of $\mathbb{M}(2,2)$.
3. T is one-to-one.

4. T is onto.

5. T is additive.

6. T is homogeneous.

Therefore, T is one-to-one, onto, and linear. This means that T preserves the vector space structure between these two spaces, and puts them in a one-to-one correspondence.

Definition 43 *An isomorphism is a linear map between vector spaces that is one-to-one and onto. A vector space V is isomorphic to a vector space W if and only if there is an isomorphism from V to W.*

Example: Show that $\mathbb{R}^2$ is isomorphic to the subspace $P = \{\vec{v} \in \mathbb{R}^3 : (2,3,5)\cdot\vec{v} = 0\}$ of $\mathbb{R}^3$.

If $(2,3,5) \cdot \vec{v} = 0$, then

$$2\vec{v}_1 + 3\vec{v}_2 + 5\vec{v}_3 = 0.$$

In other words,

$$v_1 = -\frac{3}{2}v_2 + 5v_3,$$

so

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} -\frac{3}{2}v_2 - \frac{5}{2}v_3 \\ v_2 \\ v_3 \end{pmatrix} = v_2 \begin{pmatrix} -\frac{3}{2} \\ 1 \\ 0 \end{pmatrix} + v_3 \begin{pmatrix} -\frac{5}{2} \\ 0 \\ 1 \end{pmatrix}.$$

These two vectors form a basis for P. One mapping from P to $\mathbb{R}^2$ is the coordinate transformation relative to this basis. The coordinate transformation is clearly one-to-one, onto, and linear.

Isomorphism allows us to say when two vector spaces are essentially the same. In fact, we can say that the only distinguishing characteristic of finite dimensional vector spaces is their dimension, since all vector spaces of a given dimension are isomorphic.

Theorem 45 *Suppose V and V' are finite dimensional vector spaces. Then V is isomorphic to V' if and only if $dim(V) = dim(V')$.*

Proof : $(\Rightarrow)$ Suppose that V and V' are isomorphic, and let $T : V \to V'$ be an isomorphism. Because T is one-to-one, $\text{nullity}(T) = 0$, and since T is onto, $\text{rank}(T) = \dim(V')$. Therefore,

$$\dim(V) = \text{rank}(T) + \text{nullity}(T) = \dim(V').$$

($\Leftarrow$) Suppose that $\dim(V) = \dim(V')$. Let $V = \{\vec{u}_1, \ldots, \vec{u}_n\}$ be a basis for V, and let $V' = \{\vec{u}'_1, \ldots, \vec{u}'_n\}$ be a basis for V'. We proved in a previous theorem that there is a unique linear map $T : V \to V'$ that satisfies $T(\vec{u}_i) = \vec{u}'_i$ for $i = 1, \ldots, n$. Any element of V can be written as $r_1\vec{u}_1 + \cdots + r_n\vec{u}_n$ for some scalars $r_1, \ldots, r_n$. We have that

$$T(r_1\vec{u}_1 + \cdots + r_n\vec{u}_n) = r_1T(\vec{u}_1) + \cdots + r_nT(\vec{u}_n) = r_1\vec{u}'_1 + \cdots + r_n\vec{u}'_n.$$

This shows that T is onto. Also,

$$\text{nullity}(T) = \dim(V) - \text{rank}(T) = \dim(V) - \dim(V') = 0,$$

which shows that T is one-to-one. Therefore, T is an isomorphism.

Example: Use the classification theorem for finite dimensional vector spaces to show that $\mathbb{R}^2$ is isomorphic to the subspace $P = \{\vec{v} \in \mathbb{R}^3 : (2, 3, 5) \cdot \vec{v} = 0\}$ of $\mathbb{R}^3$.

We showed that $\dim(P) = 2$, and $\dim(\mathbb{R}^2) = 2$, so they are isomorphic.

Chapter 7: Determinants

Section 7.2: Definition

One of our past homework problems was to show that the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is row equivalent to I_2 if and only if $ad - bc \neq 0$. If $A \sim I_2$, then A is invertible, which also means that the system $A\vec{x} = \vec{b}$ has a unique solution for every choice of $\vec{b}$. The next definition generalizes this idea to square matrices of arbitrary size.

Definition 44 *The determinant of an $n \times n$ matrix A, denoted $det(A)$, is defined inductively as follows:*

- *For a 1×1 matrix $A = [a_{11}]$, $det(A) = a_{11}$.*
- *For an $n \times n$ matrix A with $n > 1$, let A_{ij} denote the $(n-1) \times (n-1)$ submatrix obtained from A by deleting the ith row and the jth column. Then*

$$det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{1j} \, det(A_{1j}).$$

What this means is that we compute a determinant by "expanding" along the top row. Notice that to compute the determinant of an $n \times n$ matrix, we have to be able to compute determinants of $(n-1) \times (n-1)$ matrices, which in turn means that we have to be able to compute $(n-2) \times (n-2)$ determinants, etc. Eventually, we get down to 1×1 determinants, which are easy to compute.

For a 2×2 matrix,

$$\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = a \det(d) - b \det(c) = ad - bc,$$

just as we expected.

For a 3×3 matrix, we have

$$\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a \det\begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det\begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det\begin{pmatrix} d & e \\ g & h \end{pmatrix}$$

$$= a[ei - fh] - b[di - fg] + c[dh - eg] = aei - afh - bdi + bfy + cdh - ceg.$$

Example: Compute

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

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

$$-4 \det\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 4 & 2 & 1 \end{pmatrix} = 2\left[\det\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} - \det\begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix}\right] - \left[-\det\begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix}\right]$$

$$+3\left[-\det\begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix}\right] - 4\left[-\det\begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} + \det\begin{pmatrix} 1 & 0 \\ 4 & 2 \end{pmatrix}\right]$$

$$= 2[-2+2] - [5] + 3[5] - 4[3+2] = 10 - 20 = -10.$$

Hopefully this example makes clear that the amount of work necessary to compute an $n \times n$ determinant grows very quickly with n. However, there are some properties of determinants that can help simplify things a little.

Theorem 46 *Suppose A is an $n \times n$ matrix.*

1. *If A has a row of zeros, then $det(A) = 0$.*
2. *If two rows of A are interchanged, then the determinant of the resulting matrix is $-det(A)$.*
3. *If one row of A is multiplied by a constant c, then the determinant of the resulting matrix is $c\ det(A)$.*
4. *If two rows of A are identical, then $det(A) = 0$.*
5. *If a multiple of one row of A is added to another row of A, the determinant of the resulting matrix is $det(A)$.*
6. *$det(I_n) = 1$.*

Example: Compute

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

We have that

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

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

$$= -10\ \det\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = -10.$$

Section 7.3: Properties of Determinants

If you think a little about the definition of a determinant that we gave in the last section, you will probably ask yourself why it is so important that we expand around the first row of a matrix when we find the determinant. The answer is that it is not important at all. In fact, we can expand around any row or any column of the matrix. This is very nice, in that it lets us pick the row or column that contains the most zeros, thus reducing the amount of work we have to do to find the determinant.

In the last section we showed how the determinant works with the elementary row operations. In this section we'll show how it works with the other operations we have defined between matrices, such as matrix multiplication and transpose.

Notation: Recall from the last section that if A is an $n \times n$ matrix, then A_{ij} is the $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and jth column of A. Also, let $A_{ij,rk}$ denote the $(n-2) \times (n-2)$ matrix obtained by deleting rows i and r and columns j and k from A. See page 278 in the text for a useful diagram describing how the indexing in A changes after deleting rows and columns.

Theorem 47 *The determinant of a square matrix can be computed by expansion along any row. That is, if A is an $n \times n$ matrix, and $r \in \{1, \ldots, n\}$, then*

$$det(A) = \sum_{k=1}^{n} (-1)^{r+k} a_{rk} det(A_{rk}).$$

Proof : We finally use induction. For our base case, we check that this is true for a 2×2 matrix. Expanding along the first row, we have that

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Expansion along the second row gives

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = -a_{21}a_{12} + a_{22}a_{11},$$

so the result holds for 2×2 matrices. Suppose that the result holds for matrices of size $(n-1) \times (n-1)$, and let A be of size $n \times n$. We can assume that $r \neq 1$ (since that is the definition). We expand the determinant of A along the first row, but then in the determinants of A_{1k}, we can expand along row $r-1$ (which corresponds to row r of A). Therefore,

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \left(\sum_{k=1}^{j-1} (-1)^{(r-1)+k} a_{rk} \det(A_{1j,rk}) \right.$$

$$+\sum_{k=j+1}^{n}(-1)^{(r-1)+(k-1)}a_{rk}\det(A_{1j,rk})\Bigg) = \sum_{k<j}(-1)^{j+r+k}a_{1j}a_{rk}\ \det(A_{1j,rk})$$

$$+\sum_{k>j}(-1)^{j+r+k-1}a_{1j}a_{rk}\ \det(A_{1j,rk}).$$

Now let's start working from the right-hand side of the equation in the statement of the theorem. We have

$$\sum_{k=1}^{n}(-1)^{r+k}a_{rk}\ \det(A_{rk}) = \sum_{k=1}^{n}(-1)^{r+k}a_{rk}\left(\sum_{j=1}^{k-1}(-1)^{1+j}a_{1j}\ \det(A_{rk,1j})\right.$$

$$\left.+\sum_{j=k+1}^{n}(-1)^{1+(j-1)}a_{1j}\ \det(A_{rk,1j})\right) = \sum_{j<k}(-1)^{r+k+j+1}a_{rk}a_{ij}\det(A_{rk,1j})$$

$$+\sum_{j>k}(-1)^{r+k+j}a_{rk}a_{1j}\ \det(A_{rk,1i}).$$

Since $A_{1j,rk} = A_{rk,1l}$, these expressions are the same.

The following result is proved (somewhat) similarly:

Theorem 48 *The determinant of a square matrix can be computed by expansion along any column. That is, if A is an $n \times n$ matrix and $c \in \{1, \ldots, n\}$, then*

$$det(A) = \sum_{i=1}^{n}(-1)^{i+c}a_{ic}\ det(A_{ic}).$$

Theorem 49 *Suppose A and B are $n \times n$ matrices. Then*

$$det(AB) = det(A)\ det(B).$$

Proof : There is some sequence of elementary row operations that transforms A to A', which is the RREF of A. By our previous theorem, we know that

$$\det(A) = K\ \det(A'),$$

where K is some constant. If we apply the same sequence of row operations to AB, we'll get $A'B$ (we proved this in Chapter 5). Therefore,

$$\det(AB) = K\ \det(A'B).$$

Because A is a square matrix, A' is either the identity matrix, or it has a row of zeros. In the first case $\det(A) = K\ \det(I) = K$, so

$$\det(AB) = K\ \det(A'B) = K\ \det(IB) = K\ \det(B) = \det(A)\det(B).$$

In the second case, $A'B$ will have at least one row of zeros, so

$$\det(AB) = K\det(A'B) = K \cdot 0 = 0 = 0 \cdot \det(B) = \det(A)\det(B).$$

Definition 45 *The transpose of an $m \times n$ matrix $A = [a_{ij}]$ is the $n \times m$ matrix, denoted A^T, whose i, j entry is A_{ji}. That is,*

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

Theorem 50 *Suppose $A, A' \in \mathbb{M}(m, n)$, $B \in \mathbb{M}(n, p)$, and $r \in \mathbb{R}$. Then*
a. $(A^T)^T = A$
b. $(rA)^T = r(A^T)$
c. $(A + A')T = A^T + (A')^T$
d. $(AB)^T = B^T A^T$
e. If A is a square invertible matrix, then $(A^{-1})^T = (A^T)^{-1}$

Proof : (of part d) The i, j entry of $(AB)^T$ is the dot product of row j of A with column i of B. The i, j entry of $B^T A^T$ is the dot product of the ith row of B^T with the jth column of A^T. But since the ith row of B^T is the ith column of B and The jth column of A^T is the jth row of A, this is the same as the i, j entry of $(AB)^T$.

Theorem 51 *Suppose A is an $n \times n$ matrix. Then $det(A) = det(A^T)$.*

Proof : Easy.

Section 7.4: Cramer's Rule

Theorem 52 *For an $n \times n$ matrix A, the following conditions are equivalent:*
a. A is nonsingular
b. A has a right inverse
c. A has a left inverse
d. $rank(A) = n$

e. A can be row reduced to I. f. For any $\vec{b} \in \mathbb{R}^n$*, the equation* $A\vec{v} = \vec{b}$ *has a unique solution* $\vec{v} \in \mathbb{R}^n$*.*
g. det$(A) \neq 0$.

Let's consider the problem of finding the inverse of the 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where $a \neq 0$:

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array} \right) \sim \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & d - \frac{bc}{a} & -\frac{c}{a} & 1 \end{array} \right) \sim \left(\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array} \right).$$

This means that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

or in other words, we can write the inverse of this matrix as one over the determinant times a new matrix. Questions we should ask ourselves are:

- Is this the case for any invertible square matrix?
- How do we find the "new matrix" to be able to write down the inverse?

Trying to do the same thing for a 3×3 or larger matrix to see if there is a pattern gets out of hand very quickly (trust me), so instead we'll write down what the answer to the second question is, and then we'll verify that it gives us a formula for the inverse.

Definition 46 *Suppose A is an* $n \times n$ *matrix. Then for integers i and j between 1 and n, the ij-cofactor of A is the real number*

$$(-1)^{i+j} det(A_{ij}).$$

The adjoint of A is the $n \times n$ *matrix, denoted adj*(A)*, whose* i, j*-entry is the ji-cofactor of A.*

Let's go back to our 2×2 example and see if this definition tells us where the new matrix comes from. It ends up that

$$\text{adj}(A) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Theorem 53 *Suppose the square matrix A is nonsingular. Then*

$$A^{-1} = \frac{1}{det(A)}\ adj(A).$$

Proof : Let c_{ij} be the ij-cofactor of A. If we take $\det(A)$ by expanding along row i, we get

$$\det(A) = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} \det(A_{ik}) = \sum_{k=1}^{n} a_{ik} c_{ik}.$$

For $i \neq j$, let B be the matrix obtained from A by replacing row j by row i. Since B has two identical rows, we know that $\det(B) = 0$. But, if we calculate $\det(B)$ by expanding along row j, we get

$$0 = \det(B) = \sum_{k=1}^{n} (-1)^{j+k} a_{ik} \det(A_{jk}) = \sum_{k=1}^{n} a_{ik} c_{jk}.$$

We can think of these two sums as being the entries of a matrix product, and what they tell us is that if $C = [c_{ij}]$, then

$$AC^T = \det(A) I.$$

But c_{ij} is just the ji-entry of $\mathrm{adj}(A)$, or in other words, $C^T = \mathrm{adj}(A)$. Therefore, we can rearrange to find that

$$A\left(\frac{1}{\det(A)}\ \mathrm{adj}(A)\right) = I.$$

Before you get too excited about this formula for the inverse of an invertible matrix, you should realize that it involves a lot of computation. For example, to use this formula for a 3×3 matrix means computing one 3×3 determinant and nine 2×2 determinant. And things get quickly worse as the size increases.

Example: Use the cofactor formula to compute the inverse if the matrix

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

First, let's compute the determinants of the submatrices:

$$\det(A_{11}) = \det\begin{pmatrix} 3 & 0 \\ 3 & 1 \end{pmatrix} = 3 \ \det(A_{12}) = \det\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = -1 \ \det(A_{13}) = \det\begin{pmatrix} -1 & 3 \\ 0 & 3 \end{pmatrix} = -3$$

$$\det(A_{21}) = \det\begin{pmatrix} 0 & 2 \\ 3 & 1 \end{pmatrix} = -6 \ \det(A_{22}) = \det\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = 1 \ \det(A_{23}) = \det\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} = 3$$

$$\det(A_{31}) = \det\begin{pmatrix} 0 & 2 \\ 3 & 0 \end{pmatrix} = -6 \ \det(A_{32}) = \det\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = 2 \ \det(A_{33}) = \det\begin{pmatrix} 1 & 0 \\ -1 & 3 \end{pmatrix} = 3.$$

Therefore,

$$\mathrm{adj}(A) = \begin{pmatrix} 3 & 6 & -6 \\ 1 & 1 & -2 \\ -3 & -3 & 3 \end{pmatrix}.$$

Also we have that

$$\det(A) = \det\begin{pmatrix} 1 & 0 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -1 \end{pmatrix} = -3.$$

This gives us

$$A^{-1} - \frac{1}{-3}\begin{pmatrix} 3 & 6 & -6 \\ 1 & 1 & -2 \\ -3 & -3 & 3 \end{pmatrix} = \begin{pmatrix} -1 & -2 & 2 \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \\ 1 & 1 & -1 \end{pmatrix}.$$

As you might imagine, having a formula for the inverse of an invertible matrix leads to a new method for solving a system of linear equations.

Theorem 54 (Cramer's Rule) *Suppose A is a nonsingular $n \times n$ matrix, and $\vec{b} = (b_1, \ldots, b_n)^T \in \mathbb{R}^n$. Then the solution $\vec{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ of the system $A\vec{x} = \vec{b}$ is given by*

$$x_j = \frac{1}{det(A)}\ det(A_j),$$

where A_j is the matrix obtained by replacing the jth column of A with $\vec{b}$.

Proof : Let $C = [c_{ij}]$ be equal to $\mathrm{adj}(A)^T$. Then

$$\vec{x} = A^{-1}\vec{b} = \frac{1}{\det(A)} C^T \vec{b}$$

and therefore

$$x_j = \frac{1}{\det(A)} \sum_{i=1}^{n} c_{ij} b_i = \frac{1}{\det(A)} \sum_{i=1}^{n} (-1)^{i+j} b_i \det(A_{ij}) = \frac{1}{\det(A)} \det(A_j).$$

Example: Use Cramer's rule to find the solution to the system

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

The coefficient matrix is

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

so $\det A = -8$. Cramer's rule says that

$$x = -\frac{1}{8}\det\begin{pmatrix} 4 & 1 \\ 1 & -2 \end{pmatrix} = -\frac{1}{8}(-9) = \frac{9}{8},$$

and

$$y = -\frac{1}{8}\det\begin{pmatrix} 3 & 4 \\ 2 & 1 \end{pmatrix} = -\frac{1}{8}(-5) = \frac{5}{8}.$$

As has been mentioned, the drawback of this method is that it takes a lot of computation, but another problem is that it only applies to systems of equations that have a square coefficient matrix. However, if for some reason you need an explicit formula for one of the unknowns in a square system, Cramer's rule can give you that.

Example: Estimate the error in the solution x of the following system if the coefficient a is known to be within .01 of 6.8. What if a is known to be within .01 of -6.8?

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

The coefficient matrix of the system is

$$A = \begin{pmatrix} 4 & 1 & -2 \\ -1 & 0 & 1 \\ a & -3 & 1 \end{pmatrix} \Rightarrow \det(A) = \det\begin{pmatrix} 4 & 1 & -2 \\ -1 & 0 & 1 \\ a+12 & 0 & 7 \end{pmatrix} = -\det\begin{pmatrix} -1 & 1 \\ a+12 & -5 \end{pmatrix}$$

$$= -5 + a + 12 = a + 7.$$

We must also find the determinant of the matrix that we get when we replace the first column of A with the right-hand side:

$$\det\begin{pmatrix} 3 & 1 & -2 \\ 2 & 0 & 1 \\ 0 & -3 & 1 \end{pmatrix} = \det\begin{pmatrix} 3 & 1 & -2 \\ 2 & 0 & 1 \\ 9 & 0 & -5 \end{pmatrix} = -\det\begin{pmatrix} 2 & 1 \\ 9 & -5 \end{pmatrix} = 10 + 9 = 19.$$

Therefore,

$$x = \frac{19}{a+7} \;\Rightarrow\; \frac{dx}{da} = \frac{-19}{(a+7)^2}.$$

In other words, if we make a small change in a, say Δa, then we can expect Δx to be approximately

$$\Delta x \approx \frac{dx}{da}\Delta a = \frac{-19}{(a+7)^2}\Delta a.$$

in our problem, $a = 6.8$ with variation $\Delta a = .01$, so

$$\Delta x = \frac{-19}{(6.8+7)^2}(.01) \approx -.000998.$$

If a is within .01 of -6.8, then

$$\Delta x \approx \frac{-19}{(-6.8+7)^2}(.01) \approx -4.75.$$

Section 7.5: Cross Product

In this section, we will use the notation

$$\vec{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ \vec{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},\ \vec{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

which is standard in Calculus and Physics classes.

Definition 47 *Suppose $\vec{v} = (v_1, v_2, v_3)$ and $\vec{w} = (w_1, w_2, w_3)$ are vectors in $\mathbb{R}^3$. The cross product of $\vec{v}$ with $\vec{w}$ is the vector in $\mathbb{R}^3$ denoted $\vec{v} \times \vec{w}$ and defined by*

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

This definition is easy to remember if we think of it as a determinant:

$$\det \begin{pmatrix} \vec{i} & \vec{j} & \vec{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix} = \vec{i} \det \begin{pmatrix} v_2 & v_3 \\ w_2 & w_3 \end{pmatrix} - \vec{j} \det \begin{pmatrix} v_1 & v_3 \\ w_1 & w_3 \end{pmatrix} + \vec{k} \det \begin{pmatrix} v_1 & v_2 \\ w_1 & w_2 \end{pmatrix}$$

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

Some of the familiar laws, like the associative law, don't hold for the cross product. Here's a list of some important properties that do hold:

Theorem 55 *Suppose $\vec{v} = (v_1, v_2, v_3)$ and $\vec{w} = (w_1, w_2, w_3)$, and $\vec{x} = (x_1, x_2, x_3)$ are elements of $\mathbb{R}^3$, and r is a real number. Then*
a. $\vec{v} \times \vec{w} = -(\vec{w} \times \vec{v})$.
b. $\vec{v} \times (\vec{w} + \vec{x}) = (\vec{v} \times \vec{w}) + (\vec{v} \times \vec{x})$
c. $(\vec{v} + \vec{w}) \times \vec{x} = (\vec{v} \times \vec{x}) + (\vec{w} \times \vec{x})$
d. $r(\vec{v} + \vec{w}) = (r\vec{v}) \times \vec{w} = \vec{v} \times (r\vec{w})$
e. $\vec{v} \times \vec{0} = \vec{0} \times \vec{v} = \vec{0}$
f. $\vec{v} \times \vec{v} = \vec{0}$
g. $\vec{v} \cdot (\vec{w} \times \vec{x}) = (\vec{v} \times \vec{w}) \cdot x = det \begin{pmatrix} v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \\ x_1 & x_2 & x_3 \end{pmatrix}$

Proof : (of a)

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

$$= -\det \begin{pmatrix} w_2 & w_3 \\ v_2 & v_3 \end{pmatrix} \vec{i} + \det \begin{pmatrix} w_1 & w_3 \\ v_1 & v_3 \end{pmatrix} \vec{j} - \det \begin{pmatrix} w_1 & w_2 \\ v_1 & v_2 \end{pmatrix} = -(\vec{w} \times \vec{v}).$$

Even more important than these algebraic properties of the cross product are the geometric properties:

Theorem 56 *Suppose $\vec{v}$ and $\vec{w}$ are vectors is $\mathbb{R}^3$. Let θ be the angle between $\vec{v}$ and $\vec{w}$. Then*
a. $\vec{v} \cdot (\vec{v} \times \vec{w}) = 0$
b. $\vec{w} \cdot (\vec{v} \times \vec{w}) = 0$
c. $\|\vec{v} \times \vec{w}\|^2 = \|\vec{v}\|^2 \|\vec{w}\|^2 - (\vec{v} \cdot \vec{w})^2$
d. $\|\vec{v} \times \vec{w}\| = \|\vec{v}\| \|\vec{w}\| \sin \theta$

If you're familiar with the properties of the dot product, you know that the first two properties say that the result of a cross product is orthogonal (at a 90° angle to) the original vectors. The last two parts deal with the magnitude of vectors. If $\vec{u} = (u_1, u_2)$, then

$$\|\vec{u}\| = \sqrt{u_1^2 + u_2^2}.$$

Section 7.6: Orientation

Definition 48 *An ordered basis $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ for $\mathbb{R}^3$ is right-handed if and only if det$(\vec{u}_1\ \vec{u}_2\ \vec{u}_3) > 0$. It is left-handed if and only if det$(\vec{u}_1\ \vec{u}_2\ \vec{u}_3) < 0$. An orientation of $\mathbb{R}^3$ is the selection of a right-handed or a left-handed ordered basis.*

Example: Show that the standard basis $\{\vec{i}, \vec{j}, \vec{k}\}$ for $\mathbb{R}^3$ is right-handed.

Let $\vec{v}$ and $\vec{w}$ be any vectors in $\mathbb{R}^3$ with $\vec{v} \times \vec{w} \neq 0$. Show that $\{\vec{v}, \vec{w}, \vec{v} \times \vec{w}\}$ is right-handed.

The first part is clear, since

$$\det(\vec{i}\ \vec{j}\ \vec{k}) = \det(I_3) = 1 > 0.$$

Also, it is not difficult to verify that

$$\det(\vec{v}\ \vec{w}\ \vec{v} \times \vec{w}) = (\vec{v} \times \vec{w}) \cdot (\vec{v} \times \vec{w}) > 0.$$

Chapter 8: Eigenvalues and Eigenvectors

Section 8.1: Definitions

Example: Consider the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 23 & -12 \\ 40 & -21 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

It is hard to get an idea of what this transformation does to vectors by looking at this formula. However, note that

$$T\begin{pmatrix} 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 9 \\ 15 \end{pmatrix} = 3\begin{pmatrix} 3 \\ 5 \end{pmatrix} \quad \text{and} \quad T\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \end{pmatrix} = -1\begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

In other words, the linear transformation T acts like scalar multiplication on these vectors.

If you're paying close attention, you'll notice that the set

$$B = \left\{\begin{pmatrix} 3 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right\}$$

is a basis for $\mathbb{R}^2$. Let's look at the matrix of T relative to the basis B. Since we already know what we get when we apply T to the basis vectors, we just have to find the coordinates of the results with respect to B. In this case, it is easy to see that

$$\left[T\begin{pmatrix} 3 \\ 5 \end{pmatrix}\right]_B = \begin{pmatrix} 3 \\ 0 \end{pmatrix} \quad \text{and} \quad \left[T\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right]_B = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Therefore, the matrix of T relative to the the basis B is

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

What this is saying is that if we represent the vectors in $\mathbb{R}^2$ with in the basis B, the action of T is greatly simplified–it becomes a diagonal matrix.

So we were able to do this for T, but can it always be done? That is one of the questions that we will answer as we go throughout this chapter.

Definition 49 *Suppose $T : V \to V$ is a linear operator. Suppose λ is a real number and $\vec{v} \in V$ is a nonzero vector such that*

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

Then λ is an eigenvalue of T, and $\vec{v}$ is an eigenvector of T associated with λ. The set $E_T(\lambda) = \{\vec{v} \in V : T(\vec{v}) = \lambda\vec{v}\}$ is the eigenspace of T associated with λ.

In the example above, T has an eigenvalue 3 with eigenvector $(3, 5)^T$ and another eigenvalue -1 with eigenvector $(1, 2)^T$.

Theorem 57 *Suppose A is an $n \times n$ matrix and $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $T(\vec{v}) = A\vec{v}$. Then the real number λ is an eigenvalue of T if and only if $det(\lambda I - A) = 0$.*

Proof : In this case, we can handle both "directions" at the same time: λ is an eigenvalue of $T \Leftrightarrow T(\vec{v}) = \lambda\vec{v}$ for some nonzero $\vec{v} \in \mathbb{R}^n \Leftrightarrow A\vec{v} = \lambda\vec{v} \Leftrightarrow \lambda\vec{v} - A\vec{v} = \vec{0}$ $\Leftrightarrow (\lambda I - A)\vec{v} = \vec{0}$. This last statement is equivalent to saying that there is a nonzero vector in $\ker(\lambda I - A)$, which means this matrix is singular, which is equivalent to $\det(\lambda I - A) = 0$.

What this is saying is that if the linear operator in question is defined by matrix multiplication ($T(\vec{v}) = A\vec{v}$), then we can find its eigenvalues by solving the equation $\det(\lambda I - A) = 0$. It ends up that $\det(\lambda I - A)$ is a nth degree polynomial in λ.

Definition 50 *Suppose A is an $n \times n$ matrix. The nth degree polynomial in the variable λ defined by*

$$det(\lambda I - A)$$

is the characteristic polynomial of A. The real zeros of the characteristic polynomial (real numbers that are solutions of the characteristic equation $det(\lambda I - A) = 0$) are called the eigenvalues of the matrix A. A nonzero vector $\vec{v} \in \mathbb{R}^n$ such that $(\lambda I - A)\vec{v} = \vec{0}$ is an eigenvector of A associated with λ. The solution space

$$E_A(\lambda) = \{\vec{v} \in \mathbb{R}^n : (\lambda I - A)\vec{v} = \vec{0}\}$$

of this homogeneous system is the eigenspace of A associated with λ.

This is the definition that we'll use for this class, but please be aware that this definition is **not standard**. Most people are perfectly happy to say that a matrix has complex eigenvalues. However, in this book, a matrix that has only complex conjugate solutions to the characteristic equation will be said to have no eigenvalues.

Example: If

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

then the characteristic polynomial is

$$\det \begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix} = \lambda^2 + 1,$$

which has solution $\lambda = \pm i$. Therefore, we say that A has no eigenvalues (even though most of the rest of the world would say that A has eigenvalues $\pm i$ with corresponding complex eigenvectors).

Example: Find the eigenvalues of the matrix

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

For each eigenvalue, find a basis for the associated eigenspace.

To find the eigenvalues, we solve the characteristic equation

$$0 = \det \begin{pmatrix} \lambda - 1 & -2 & 1 \\ -1 & \lambda & -1 \\ -4 & 4 & \lambda - 5 \end{pmatrix} = (\lambda - 1)\det \begin{pmatrix} \lambda & -1 \\ 4 & \lambda - 5 \end{pmatrix} + 2 \det \begin{pmatrix} -1 & -1 \\ -4 & \lambda - 5 \end{pmatrix}$$

$$+\det \begin{pmatrix} -1 & \lambda \\ -4 & 4 \end{pmatrix} = (\lambda - 1)(\lambda(\lambda - 5) + 4) + 2(-(\lambda - 5) - 4) + (-4 + 4\lambda)$$

$$= (\lambda-1)(\lambda^2-5\lambda+4)+2(-\lambda+1)+(-4+4\lambda) = \lambda^3-5\lambda^2+4\lambda-\lambda^2+5\lambda-4-2\lambda+2-4+4\lambda$$

$$= \lambda^3 - 6\lambda^2 + 11\lambda - 6.$$

Our best bet here is to try to find a root of the characteristic polynomial by trying the factors of the first and last coefficients (use the rational roots theorem). If we try $\lambda = 1$, we find

$$(1)^3 - 6(1) + 11(1) - 6 = 1 - 6 + 11 - 6 = 0.$$

Thus $(\lambda - 1)$ is a factor of the characteristic polynomial, and upon performing long division, we find

$$\begin{array}{rl} & \lambda^2 - 5\lambda + 6 \\ \cline{2-2} \lambda - 1 &)\lambda^3 - 6\lambda^2 + 11\lambda - 6 \\ & \underline{-\lambda^3 + \lambda^2} \\ & -5\lambda^2 + 11\lambda - 6 \\ & \underline{5\lambda^2 - 5\lambda} \\ & 6\lambda - 6 \\ & \underline{-6\lambda + 6} \\ & 0 \end{array}.$$

If $\lambda^2 - 5\lambda + 6 = 0$, then

$$\lambda^2 - 5\lambda + \frac{25}{4} = -6\frac{25}{4} \;\Rightarrow\; \left(\lambda - \frac{5}{2}\right)^2 = \frac{1}{4} \;\Rightarrow\; \lambda - \frac{5}{2} = \pm\frac{1}{2} \;\Rightarrow\; \lambda = \frac{5}{2} \pm \frac{1}{2} = 3 \text{ or } 2.$$

Thus A has three eigenvalues 3, 1, and 2. If $\lambda = 1$, then

$$\lambda I - A = \begin{pmatrix} 0 & -2 & 1 \\ -1 & 1 & -1 \\ -4 & 4 & -4 \end{pmatrix} \sim \begin{pmatrix} 1 & -1 & 1 \\ 0 & -2 & 1 \\ 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

This implies that in the homogeneous system of equations $(\lambda I - A)\vec{x} = \vec{0}$, there is only one free variable. Therefore, a basis for the null space of the matrix $(\lambda I - A)$ with $\lambda = 1$ is any nonzero solution to the homogeneous system. One such solution is

$$\left\{ \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} \right\}$$

For $\lambda = 2$, we have

$$\lambda I - A = \begin{pmatrix} 1 & -2 & 1 \\ -1 & 2 & -1 \\ -4 & 4 & -3 \end{pmatrix} \sim \begin{pmatrix} 1 & -2 & 1 \\ 0 & 0 & 0 \\ 0 & -4 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & -\frac{1}{4} \\ 0 & 0 & 0 \end{pmatrix}.$$

In this case, a basis for the null space is

$$\left\{ \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix} \right\}.$$

If $\lambda = 3$, then

$$\lambda I - A = \begin{pmatrix} 2 & -2 & 1 \\ -1 & 3 & -1 \\ -4 & 4 & -2 \end{pmatrix} \sim \begin{pmatrix} 1 & -3 & 1 \\ 0 & 4 & -1 \\ 0 & -8 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & \frac{1}{4} \\ 0 & 1 & -\frac{1}{4} \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus, a basis for this eigenspace is

$$\left\{ \begin{pmatrix} -1 \\ 1 \\ 4 \end{pmatrix} \right\}.$$

Section 8.2: Similarity

As we saw in the last chapter, we can find a matrix representation for any linear operator $T : V \to V$. Also, we can construct the change of basis matrix so that we can start with the representation with respect to one basis and find the representation with respect to another basis.

One question that we might ask ourselves is that if we are given such an operator T, which basis should we choose to represent it?

Definition 51 *An $n \times n$ matrix A is similar to an $n \times n$ matrix B if and only if $A = P^{-1}BP$ for some nonsingular $n \times n$ matrix P. This relation is denoted*

$$A \sim B.$$

Here are some simple properties of this relation:

Theorem 58 *For any $n \times n$ matrices A, B, and C,*
a. $A \sim A$ (reflexivity)
b. if $A \sim B$, then $B \sim A$ (symmetry)
c. if $A \sim B$ and $B \sim C$, then $A \sim C$ (transitivity)

For a given $n \times n$ matrix A, we call the set of all matrices that are similar to A its similarity class. The properties above show that any two matrices in the similarity class of a given matrix must be similar to each other.

It is easy to take a matrix A and a nonsingular matrix P and use these to create a matrix that is similar to A: we just find $P^{-1}AP$. But it is generally much more difficult to take two square matrices of the same size and determine whether or not they are similar.

One test for similarity is to take the determinant. Suppose that A is $n \times n$ and P is $n \times n$ and nonsingular such that $B = P^{-1}AP$. Then

$$P^{-1}P = I \Rightarrow 1 = \det(I) = \det(P^{-1}P) = \det(P^{-1})\det(P).$$

Therefore,

$$\det(B) = \det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) = \det(A).$$

This tells us that if two matrices are similar, they must have the same determinant. Another way of saying this is that if we take the determinants of two $n \times n$ matrices and we do not get the same result, we know that they are not similar.

Definition 52 *The main diagonal of an $n \times n$ matrix $A = [a_{ij}]$ consists of those positions where the row index is equal to the column index. The trace of A, denoted by $tr(A)$, is the sum of the entries on the main diagonal of A. That is,*

$$tr(A) = \sum_{k=1}^{n} a_{kk}.$$

Theorem 59 *Suppose A and B are $n \times n$ matrices, P is an invertible $n \times n$ matrix, and r is a real number. Then*
a. $tr(A + B) = tr(A) + tr(B)$
b. $tr(rA) = r\ tr(A)$
c. $tr(AB) = tr(BA)$
d. $tr(A) = tr(P^{-1}AP)$

One of the things that this theorem says is that the trace is a linear function. Another thing that the theorem says is that if two matrices are similar, they have the same trace. However, even if two matrices have the same determinant and the same trace, that is not enough to say that they are similar. An even better test for similarity is

Theorem 60 *If two matrices are similar, then they have the same characteristic polynomial.*

Proof : If A is a square matrix and $B = P^{-1}AP$, then

$$\det(\lambda I - B) = \det(\lambda I - P^{-1}AP) = \det(P^{-1}(\lambda I)P - P^{-1}AP)$$

$$= \det(P^{-1}(\lambda I - A)P) = \det(\lambda I - A).$$

Note that this is another way to test that two matrices are *not* similar, but it is still possible to find matrices that share the same characteristic polynomial but are not similar.

We can use the idea of a matrix representation to extend the idea of a characteristic polynomial to any linear operator $T : V \to V$ operating on a finite dimensional vector space V.

Definition 53 *Suppose $T : V \to V$ is a linear operator on a finite-dimensional vector space V. The characteristic polynomial of T is the characteristic polynomial of a matrix that represents T relative to some basis for V.*

Definition 54 *A diagonal matrix is a square matrix whose only nonzero entries occur on the main diagonal of the matrix.*

Definition 55 *An $n \times n$ matrix A is diagonalizable if and only if there is a diagonal matrix D with $D \sim A$.*

We will continue to study which matrices are diagonalizable and how this relates to the eigenvalues and eigenvectors of the matrix. Of course, if an $n \times n$ matrix A is diagonalizable, then so is every matrix that is similar to A.

Section 8.3: Diagonalization

Recall that our goal is, given a matrix A, find an invertible matrix P such that $P^{-1}AP$ is invertible. The next two theorems give the answer.

Theorem 61 *Suppose $B = \{\vec{u}_1, \ldots, \vec{u}_n\}$ is an ordered basis for a vector space V and $T : V \to V$ is a linear operator. Let A be the matrix of T relative to B. Then $\vec{u}_j$ is an eigenvector of T if and only if the jth column of A is equal to $\lambda\vec{e}_j$ for some real number λ. In this case, λ is the eigenvalue that $\vec{u}_j$ is associated with.*

Proof : ($\Rightarrow$) Suppose that $\vec{u}_j$ is an eigenvector of T associated with the eigenvalue λ. Since A is the matrix of T,

$$[T(\vec{v})]_B = A[\vec{v}]_B$$

for all $\vec{v} \in V$. The jth column of A is

$$A\vec{e}_j = A[\vec{u}_j]_B = [T(\vec{u}_j)]_B = [\lambda\vec{u}_j]_B = \lambda\vec{e}_j.$$

($\Leftarrow$) Suppose that the jth column of A is $\lambda\vec{e}_j$ for some $\lambda \in \mathbb{R}$. Then

$$[T(\vec{u}_j)]_B = A[\vec{u}_j]_B = A\vec{e}_j = \lambda\vec{e}_j = [\lambda u_j]_B.$$

Since the coordinate mapping $[\cdot]_B$ is one-to-one, we have that

$$T(\vec{u}_j) = \lambda u_j.$$

Theorem 62 *A linear operator $T : V \to V$ defined on a finite dimensional vector space V can be represented by a diagonal matrix if and only if V has an ordered basis $\{\vec{u}_1, \dots, \vec{u}_n\}$ with all elements being eigenvectors of T. In this situation, the entries on the main diagonal of the matrix are the eigenvalues of T and $\vec{u}_j$ is an eigenvector associated with the jj-entry of the matrix.*

Proof : ($\Rightarrow$) Suppose that T can be represented by a diagonal matrix $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ relative to a basis $B = \{\vec{u}_1, \dots, \vec{u}_n\}$. Then

$$[T(\vec{v})]_B = D[\vec{v}]_B \text{ for all } \vec{v} \in V.$$

Then

$$[T(\vec{u}_j)]_B = D[\vec{u}_j]_B = D\vec{e}_j = \lambda_j \vec{e}_j = \lambda_j [\vec{u}_j]_B.$$

Again, since $[\cdot]_B$ is one-to-one, we must have $T(\vec{u}_j) = \lambda \vec{u}_j$.

($\Leftarrow$) Suppose that $B = \{\vec{u}_1, \dots, \vec{u}_n\}$ are eigenvectors of T corresponding to eigenvalues $\lambda_1, \dots, \lambda_n$, and B is a basis for V. Suppose also that A is the matrix representation of T with respect to the basis B. Then

$$A\vec{e}_j = A[\vec{u}_j]_B = [T(\vec{u}_j)]_B = [\lambda_j \vec{u}_j]_B = \lambda_j \vec{e}_j.$$

Therefore, the jth column of A is the jth column of the identity matrix multiplied by λ_j, or in other words $A = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$.

Notice that a set of eigenvectors that is a basis cannot always be found. Sometimes a matrix/operator does not have enough linearly independent eigenvectors to form a basis.

We can be assured that we can find a basis of eigenvectors for an n-dimensional space V if the characteristic polynomial has n real roots, and if for each distinct root, the dimension of the corresponding eigenspace is equal to the multiplicity of the root.

Theorem 63 *Suppose $\lambda_1, \dots, \lambda_n$ are distinct eigenvalues of a linear operator $T : V \to V$. For $i = 1, \dots, n$, suppose $\{\vec{u}_{i1}, \dots, \vec{u}_{ir_i}\}$ is a linearly independent subset of $E_T(\lambda_i)$. Then*

$$\{\vec{u}_{11}, \dots, \vec{u}_{1r_1}, \vec{u}_{21}, \dots, \vec{u}_{2r_2}, \dots, \vec{u}_{n1}, \dots, \vec{u}_{nr_n}\}$$

is a linearly independent set.

Proof : First let's do the special case where we have one eigenvector $\vec{u}_i$ associated with each eigenvalue λ_i. We will use mathematical induction on n, the number of distinct eigenvalues. For the base case $n = 1$, then $\{u_1\}$ is a linearly independent set, so we're done. Suppose that for some $n \geq 1$, the set $\{\vec{u}_1, \dots, \vec{u}_n\}$ is linearly independent. We want to show that the set $\{\vec{u}_1, \dots, \vec{u}_{n+1}\}$ is also linearly independent. Suppose that

$$a_1\vec{u}_1 + \cdots + a_{n+1}\vec{u}_{n+1} = \vec{0}.$$

Applying T to both sides of this equation, we get

$$\vec{0} = T(\vec{0}) = T(a_1\vec{u}_1 + \cdots + a_{n+1}\vec{u}_{n+1}) = a_1T(\vec{u}_1) + \cdots + a_{n+1}T(\vec{u}_{n+1})$$

$$= a_1\lambda_1\vec{u}_1 + \cdots + a_{n+1}\lambda_{n+1}\vec{u}_{n+1}.$$

Now we multiply our first equation by λ_{n+1} to obtain

$$\vec{0} = \lambda_{n+1}\vec{0} = \lambda_{n+1}(a_1\vec{u}_1 + \cdots + a_{n+1}\vec{u}_{n+1}) = a_1\lambda_{n+1}\vec{u}_1 + \cdots + a_{n+1}\lambda_{n+1}\vec{u}_{n+1}.$$

Subtracting the last two equations, we get

$$\vec{0} = a_1(\lambda_1 - \lambda_{n+1})\vec{u}_1 + \cdots + a_n(\lambda_n - \lambda_{n+1})\vec{u}_n + a_{n+1}(\lambda_{n+1} - \lambda_{n+1})\vec{u}_{n+1}$$

$$= a_1(\lambda_1 - \lambda_{n+1})\vec{u}_1 + \cdots + a_n(\lambda_n - \lambda_{n+1})\vec{u}_n.$$

Since $\{\vec{u}_1, \dots, \vec{u}_n\}$ is linearly independent, we must have

$$a_1(\lambda_1 - \lambda_{n+1}) = \cdots = a_n(\lambda_n - \lambda_{n+1}) = 0.$$

Because we have distinct eigenvalues, this implies that

$$a_1 = \ldots = a_n = 0,$$

so our original equation becomes

$$a_{n+1}\vec{u}_{n+1} = \vec{0}.$$

The vector $\vec{u}_{n+1}$ is nonzero (because it is an eigenvector), so $a_{n+1} = 0$ as well. Therefore, $\{\vec{u}_1, \dots, \vec{u}_{n+1}\}$ is linearly independent.

Now let's assume that for each $i = 1, \dots, n$ we have a linearly independent subset $\{\vec{u}_{i1}, \dots, \vec{u}_{ir_i}\}$ of $E_T(\lambda_i)$. As usual, we take a linear combination of all the vectors and set it equal to zero:

$$(a_{11}\vec{u}_{11} + \cdots + a_{1r_1}\vec{u}_{1r_1}) + \cdots + (a_{n1}\vec{u}_{n1} + \cdots + a_{nr_n}\vec{u}_{nr_n}) = \vec{0}.$$

Let's define

$$\vec{u}_j := a_{j1}\vec{u}_{j1} + \cdots + a_{jr_j}\vec{u}_{jr_j}, \quad j = 1, \ldots, n.$$

Then $\vec{u}_j \in E_T(\lambda_j)$. With this notation, our linear combination above can be written as

$$1\vec{u}_1 + \cdots + 1\vec{u}_n = \vec{0}.$$

In this equation, we will drop any terms where $\vec{u}_i = \vec{0}$. Then with the remaining terms we have a linear combination of eigenvectors of T corresponding to distinct eigenvalues, where all of the coefficients are equal to one. The first part of our proof shows that these eigenvectors are linearly independent. Therefore, when we were dropping zero terms, we must have dropped *all* the terms. In other words,

$$\vec{0} = \vec{u}_j = a_{j1}\vec{u}_{j1} + \cdots + a_{jr_j}\vec{u}_{jr_j}, \quad j = 1, \ldots, n,$$

and the assumption that each of these sets of vectors is linearly independent shows that all of the coefficients must be zero. Therefore, the set

$$\{\vec{u}_{11}, \ldots, \vec{u}_{1r_1}, \vec{u}_{21}, \ldots, \vec{u}_{2r_2}, \ldots, \vec{u}_{n1}, \ldots, \vec{u}_{nr_n}\}$$

is linearly independent.

Example: Which of the following matrices can be diagonalized?

$$A = \begin{pmatrix} 4 & 1 & 1 \\ 0 & 7 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 1 & 1 \\ 0 & 7 & 1 \\ 0 & 0 & 7 \end{pmatrix}, \quad C = \begin{pmatrix} 4 & 1 & 1 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

It is easy to see that the characteristic polynomial of A is

$$\det(\lambda I - A) = (\lambda - 4)(\lambda - 7)(\lambda - 2),$$

so the eigenvalues of A are 4, 7, and 2. As we have seen, eigenvectors corresponding to distinct eigenvalues form a linearly independent set, so A can be diagonalized.

For B, the characteristic polynomial is

$$\det(\lambda I - B) = (\lambda - 4)(\lambda - 7)(\lambda - 7),$$

so we can't use our result about eigenvectors corresponding to distinct eigenvalues. We have to see if the eigenspace corresponding to $\lambda = 7$ has dimension one or dimension two. Since

$$7I - B = \begin{pmatrix} -3 & -1 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

the eigenspace, which is just the solution space of the homogeneous equation with coefficient matrix $(7I - B)$. Since there is only one free variable, there is only one linearly independent vector in this set. Therefore B cannot be diagonalized.

For matrix C, we get the same characteristic polynomial as we did with B, so the eigenvalues are the same. However, in this case

$$7I - C = \begin{pmatrix} -3 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

so there will be two free variables this time. Therefore, we can come up with a basis for $\mathbb{R}^3$ by taking two linearly independent vectors from the eigenspace for $\lambda = 7$ and one from the eigenspace for $\lambda = 4$. Thus C can be diagonalized.

Theorem 64 *A linear operator $T : V \to V$ defined on a finite dimensional vector space V can be represented by a diagonal matrix if and only if the sum of the dimensions of all the eigenspaces is equal to dim(V).*

Proof : $(\Rightarrow)$ Suppose that T can be represented by diagonal matrix, and that the unique diagonal entries are $\lambda_1, \ldots, \lambda_r$. As we have proved previously, this implies that $\lambda_1, \ldots, \lambda_r$ are eigenvalues of T, and there exists a basis for V that consists of eigenvectors corresponding to these eigenvalues. The eigenvectors in this basis that correspond to λ_i must span $E_T(\lambda_i)$, for otherwise we can add more vectors from $E_T(\lambda_i)$ until we had a basis for that subspace. But then we could add the same vectors to our basis, and we would have a linearly independent set with more vectors than $\dim(V)$. Hence,

$$\sum_{i=1}^{r} \dim(E_T(\lambda_i)) = \dim(V).$$

By following the same reasoning, we conclude that there cannot be any other eigenspaces. Therefore, the sum of all the dimensions of the eigenspaces of T is equal to the dimension of V.

$(\Leftarrow)$ Suppose that the sum of all the eigenspaces is equal to $\dim(V)$. Then we can find a basis for each eigenspace, and the union of all these bases will be linearly independent, since each basis corresponds to a distinct eigenvalue. Since this set is linearly independent and contains $\dim(V)$ elements, it is a basis for V. Since V has a basis of consisting of eigenvectors of T, T can be diagonalized.

Corollary 1 *Suppose $T : V \to V$ is a linear operator defined on an n-dimensional vector space V. If T has n distinct eigenvalues, then T can be represented by a* *diagonal matrix.*

Section 8.4: Symmetric Matrices

Definition 56 *A matrix is symmetric if and only if it equals its transpose.*

Most of the results in this section will use the dot product on $\mathbb{R}^n$. Recall

Definition 57 *Given two vectors $\vec{x}$ and $\vec{y}$ from $\mathbb{R}^n$, their dot product is defined by*

$$\vec{x} \cdot \vec{y} = \sum_{k=1}^{n} x_k y_k = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

In terms of matrix multiplication, we can write the dot product between two column vectors in $\mathbb{R}^n$ as

$$\vec{x} \cdot \vec{y} = x^T y.$$

Definition 58 *Vectors $\vec{x}$ and $\vec{y}$ from $\mathbb{R}^n$ are said to be orthogonal if*

$$\vec{x} \cdot \vec{y} = 0$$

Definition 59 *A set of vectors $\{\vec{u}_1, \ldots, \vec{u}_k\} \subset \mathbb{R}^n$ is said to be orthogonal if*

$$\vec{u}_i \cdot \vec{u}_j = 0$$

whenever $i \neq j$. The set of vectors is orthonormal if it is orthogonal and

$$\vec{u}_i \cdot \vec{u}_i = 1$$

for $i = 1, \ldots, k$.

Theorem 65 *Suppose A is an $n \times n$ matrix. For any vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^n$,*

$$(A\vec{v}) \cdot \vec{w} = \vec{v} \cdot (A^T \vec{w}).$$

Proof : We have that

$$(A\vec{v}) \cdot \vec{w} = (A\vec{v})^T \vec{w} = \vec{v}^T A^T \vec{w} = \vec{v}^T (A^T \vec{w}) = \vec{v} \cdot (A^T \vec{w}).$$

Theorem 66 *An $n \times n$ matrix A is symmetric if and only if*

$$(A\vec{v}) \cdot \vec{w} = \vec{v} \cdot (A\vec{w})$$

for all vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$.

Proof : ($\Rightarrow$) Since A is symmetric, $A^T = A$, and therefore,

$$(A\vec{v}) \cdot \vec{w} = \vec{v} \cdot (A^T \vec{w}) = \vec{v} \cdot (A\vec{w}).$$

($\Leftarrow$) Suppose $(A\vec{v}) \cdot \vec{w} = \vec{v} \cdot (A\vec{w})$ holds for all vectors in $\mathbb{R}^n$. Write the i, j-element of A as $(A\vec{e}_j) \cdot e_i$. Then

$$a_{ij} = (A\vec{e}_j) \cdot \vec{e}_i = \vec{e}_j \cdot (A\vec{e}_i) = a_{ji},$$

so A is symmetric.

Definition 60 *A matrix P is orthogonal if and only if $P^T P = I$.*

This means that the inverse of an orthogonal matrix is its transpose. This is especially useful when a change of basis matrix is orthogonal. If we have a symmetric matrix A and we change basis with an orthogonal matrix P, then

$$(P^{-1}AP)^T = P^T A^T (P^T)^T = P^T AP = P^{-1}AP,$$

so the matrix with respect to the new basis is symmetric also.

For the next result, we will need to recall something about complex numbers. If $a, b \in \mathbb{R}$, then the conjugate of the complex number $z = a + bi$ is $\bar{z} = a - bi$. In particular, note that $z\bar{z} = a^2 + b^2$ is a real number. For a vector $\vec{v}$ containing complex entries, $\overline{\vec{v}}$ will be the vector obtained by conjugating each entry.

Theorem 67 *The characteristic polynomial of a $n \times n$ symmetric matrix has n real roots.*

Proof : Suppose A is an $n \times n$ symmetric matrix. We know that the characteristic polynomial of A has degree n, so the fundamental theorem of algebra tells us that it must have n complex roots, counting multiplicity. We have to show that these roots are all real. Suppose that λ is a root of the characteristic polynomial, meaning that λ is an eigenvalue of A. Suppose that $\vec{v}$ is the corresponding eigenvector (which may have nonreal entries if λ is not real). Recall that A has real entries, so we have

$$\lambda\overline{\vec{v}}^T \vec{v} = \overline{\vec{v}}^T (\lambda v) = \overline{\vec{v}}^T (A\vec{v}) = (A\overline{\vec{v}})^T \vec{v} = (\overline{A\vec{v}})^T \vec{v} = (\overline{\lambda\vec{v}})^T \vec{v} = \bar{\lambda}\overline{\vec{v}}^T \vec{v}.$$

Since $\overline{\vec{v}}^T \vec{v}$ is not zero, we must have that $\bar{\lambda} = \lambda$, which implies that the imaginary part of λ is zero.

Theorem 68 *Suppose A symmetric matrix A has eigenvectors $\vec{v}_1$ and $\vec{v}_2$ associated with distinct eigenvalues λ_1 and λ_2, respectively. Then $\vec{v}_1$ and $\vec{v}_2$ are orthogonal.*

Proof : We have that

$$\lambda_1 \vec{v}_1^T \vec{v}_2 = (\lambda_1 \vec{v}_1)^T \vec{v}_2 = (A\vec{v}_1)^T \vec{v}_2 = v_1^T (A\vec{v}_2) = \vec{v}_1^T (\lambda_2 \vec{v}_2) = \lambda_2 \vec{v}_1^T \vec{v}_2.$$

Since $\lambda_1 \neq \lambda_2$, we must conclude that $\vec{v}_1^T \vec{v}_2 = 0$.

Theorem 69 *If A is a symmetric $n \times n$ matrix, then there exists and orthogonal matrix P such that $P^{-1}AP = D$, and diagonal matrix. The eigenvalues of A lie on the main diagonal of D.*

Proof : Omitted.

Section 8.5: Systems of Differential Equations

You have probably had some experience already with differential equations. In general, first order ordinary differential equation has the form

$$y' = f(x, y),$$

where $y = y(x)$ is an unknown function that we wish to solve for. A simple example is the equation

$$y' = ay.$$

What this says is that we are looking for a function whose derivative is a constant times the original function. It is easy to see that the function $y(x) = Ce^{ax}$ is a solution for any choice of the constant C, since

$$y' = aCe^{ax} = ay.$$

In a constant coefficient system of linear ordinary differential equations, instead of one unknown function, we have unknown functions $y_1, \ldots, y_n$, and the derivative of the unknown functions is a linear combination of itself and the other unknown functions. In other words, each unknown function y_i satisfies and equation of the form

$$y_i' = a_{i1}y_1 + \cdots + a_{in}y_n.$$

If we think of each of the unknown functions as being an entry of a vector $\vec{y}(x)$, then we can write our equations as

$$\frac{d}{dx}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} a_{11}y_1 + \cdots + a_{1n}y_n \\ \vdots \\ a_{n1}y_1 + \cdots + a_{nn}y_n \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Therefore, if we consider the coefficients in the equations as the entries of a matrix A, then we can write our system of differential equations as

$$\vec{y}' = A\vec{y},$$

where

$$\vec{y}' = \begin{pmatrix} y_1' \\ \vdots \\ y_n' \end{pmatrix}.$$

Notice that this looks a lot like our simple example $y' = ay$, except that now we have vectors of functions and matrices instead of functions and scalars. In that case, we were able to solve the equation by guessing that an exponential solution would work because of the form. Let's try the same thing here. We'll assume that our solution has the form

$$\vec{y}(x) = e^{rx}\vec{v},$$

where $\vec{v} \in \mathbb{R}^n$ is a constant vector. Then $\vec{y}'(x) = re^{rx}\vec{v}$, and if this guess is indeed a solution, we must have

$$re^{rx}\vec{v} = Ae^{rx}\vec{v}.$$

Canceling the term e^{rx} from both sides, we find that

$$A\vec{v} = r\vec{v}.$$

Of course we recognize this as the equation for $\vec{v} \neq \vec{0}$ to be an eigenvector of A corresponding to the eigenvalue r. Therefore, the vector function $\vec{y}(x) = e^{rx}\vec{v}$ will be a solution to the equation $\vec{y}(x) = A\vec{y}$ provided that r is an eigenvalue of A with eigenvector $\vec{v}$. But notice that r could be any eigenvalue. of A. Also, any constant multiple of a solution is a solution. In fact, if $\vec{y}_1, \ldots, \vec{y}_n$ satisfy $\vec{y}_j = A\vec{y}_j$ and $c_1, \ldots, c_n$ are constants, then

$$\frac{d}{dx}(c_1\vec{y}_1 + \cdots + c_n\vec{y}_n) = c_1\vec{y}_1 + \cdots + c_n\vec{y}_n = c_1A\vec{y}_1 + \cdots + c_nA\vec{y}_n = A(c_1\vec{y}_1 + \cdots c_n\vec{y}_n),$$

so a linear combination of solutions to our equation is again a solution.

As it is, the problem $\vec{y} = A\vec{y}$ doesn't have a unique solution, but we can make the problem have a unique solution if we add another requirement. If we know the value of all the unknown $\vec{y}(x)$ at a certain point x_0, say $\vec{y}(x_0) = \vec{y}_0$, and if we have found solutions $\vec{y}_1(x), \ldots, \vec{y}_n(x)$, then we can find a solution that satisfies this **initial value** by taking a linear combination of these solutions and solving the problem

$$c_1\vec{y}_1(x_0) + \cdots + c_n\vec{y}_n(x_0) = \vec{y}_0.$$

By now, we easily recognize this as a system of linear equations whose coefficient matrix has the columns $\vec{y}_1(x_0), \ldots, \vec{y}_n(x_0)$, and we know that this problem will have a unique solution provided that these columns are linearly independent.

Since the solutions we are using in this linear combination were found from eigenvalues and eigenvectors of the matrix $n \times n$ matrix A, we need to be able to find n linearly independent eigenvectors to be able to find our unique solution to the initial value problem

$$\vec{y}'(x) = A\vec{y}(x), \quad \vec{y}(x_0) = \vec{y}_0.$$

(Actually, this condition can be relaxed, but we'll leave the details of that for an actual class on ODE.)

Example: Solve the initial value problem

$$\vec{y}' = \begin{pmatrix} -3 & 4 \\ 6 & -5 \end{pmatrix} \vec{y}, \quad \vec{y}(0) = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

First, we must find the eigenvalues and corresponding eigenvectors of the matrix. Then we can use them to construct solutions to the equation. The characteristic polynomial is

$$0 = \det(\lambda I - A) = \det \begin{pmatrix} \lambda+3 & -4 \\ -6 & \lambda+5 \end{pmatrix} = (\lambda+3)(\lambda+5) - 24 = \lambda^2 + 8\lambda - 9.$$

Completing the square gives us

$$\lambda^2 + 8\lambda + 16 = 9 + 16 \;\Rightarrow\; (\lambda+4)^2 = 25 \;\Rightarrow\; \lambda + 4 = \pm 5 \;\Rightarrow\; \lambda = -4 \pm 5 = 1 \text{ or } -9.$$

To find an eigenvector corresponding to $\lambda = 1$:

$$\begin{pmatrix} 4 & -4 \\ -6 & 6 \end{pmatrix} \sim \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \;\Rightarrow\; v_1 = v_2 \;\Rightarrow\; \vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

To find an eigenvector corresponding to $\lambda = -9$:

$$\begin{pmatrix} -6 & -4 \\ -6 & -4 \end{pmatrix} \sim \begin{pmatrix} 1 & \frac{2}{3} \\ 0 & 0 \end{pmatrix} \;\Rightarrow\; v_1 = -\frac{2}{3} v_2 \;\Rightarrow\; \vec{v}_2 = \begin{pmatrix} -2 \\ 3 \end{pmatrix}.$$

Therefore, two solutions to our equation are

$$\vec{y}_1(x) = e^x \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \vec{y}_2(x) = e^{-9x} \begin{pmatrix} -2 \\ 3 \end{pmatrix}.$$

In order to find a solution that satisfies our initial condition, we solve

$$c_1\vec{y}_1(0)+c_2\vec{y}_2(0)=\begin{pmatrix}2\\4\end{pmatrix} \Rightarrow \begin{pmatrix}1 & -2\\1 & 3\end{pmatrix}\begin{pmatrix}c_1\\c_2\end{pmatrix}=\begin{pmatrix}2\\4\end{pmatrix}$$

$$\left(\begin{array}{cc|c}1 & -2 & 2\\1 & 3 & 4\end{array}\right)\sim\left(\begin{array}{cc|c}1 & -2 & 2\\0 & 5 & 2\end{array}\right)\sim\left(\begin{array}{cc|c}1 & 0 & \frac{12}{5}\\0 & 1 & \frac{2}{5}\end{array}\right).$$

Therefore, the solution to our initial value problem is

$$\vec{y}(x)=\frac{12}{5}e^{x}\begin{pmatrix}1\\1\end{pmatrix}+\frac{2}{5}e^{-9x}\begin{pmatrix}-2\\3\end{pmatrix}.$$

Example: Solve the initial value problem

$$\vec{y}'=\begin{pmatrix}9 & 5\\-6 & -2\end{pmatrix}\vec{y},\quad \vec{y}(0)=\begin{pmatrix}1\\0\end{pmatrix}.$$

The characteristic polynomial of the matrix is

$$0=\det\begin{pmatrix}\lambda-9 & -5\\6 & \lambda+2\end{pmatrix}=(\lambda-9)(\lambda+2)+30=\lambda^2-7\lambda+12.$$

Completing the square, we have

$$\lambda^2-7\lambda+\frac{49}{4}=-12+\frac{49}{4} \Rightarrow \left(\lambda-\frac{7}{2}\right)^2=\frac{1}{4} \Rightarrow \lambda-\frac{7}{2}=\pm\frac{1}{2} \Rightarrow \lambda=\frac{7}{2}\pm\frac{1}{2}=4 \text{ or } 3.$$

We find an eigenvector corresponding to $\lambda=4$

$$\begin{pmatrix}-5 & -5\\6 & 6\end{pmatrix}\sim\begin{pmatrix}1 & 1\\0 & 0\end{pmatrix} \Rightarrow v_1=-v_2 \Rightarrow \vec{v}_1=\begin{pmatrix}-1\\1\end{pmatrix}$$

and $\lambda=3$

$$\begin{pmatrix}-6 & -5\\6 & 5\end{pmatrix}\sim\begin{pmatrix}1 & \frac{5}{6}\\0 & 0\end{pmatrix} \Rightarrow v_1=-\frac{5}{6}v_2 \Rightarrow \vec{v}_2=\begin{pmatrix}-5\\6\end{pmatrix}.$$

Therefore,

$$\vec{y}(x)=c_1e^{4x}\begin{pmatrix}-1\\1\end{pmatrix}+c_2e^{3x}\begin{pmatrix}-5\\6\end{pmatrix}.$$

The initial condition implies a system of equations with augmented matrix

$$\left(\begin{array}{cc|c} -1 & -5 & 1 \\ 1 & 6 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 5 & -1 \\ 0 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & -6 \\ 0 & 1 & 1 \end{array}\right),$$

so the solution to the initial value problem is

$$\vec{y}(x) = -6e^{4x}\begin{pmatrix} -1 \\ 1 \end{pmatrix} + e^{3x}\begin{pmatrix} -5 \\ 6 \end{pmatrix}.$$