A GUIDE TO PROOFS IN LINEAR ALGEBRA

by Curtis Paul

WHAT IS A PROOF?

What is a proof and why do we care? Previously in your mathematical life you have mostly focused on computation. You were not concerned with whether or not things were true but with whether you got the “right” answer. It is not uncommon for students to say, “I know the answer. Why do I need to understand the steps?” There are a number of responses to this. How do you know your answer is correct? How can you convince someone else your answer is correct? If you are given a harder problem to which you don’t know the answer, how are you going to approach it? Proof addresses these concerns.

A proof is a sequence of statements justified by axioms, theorems, definitions, and logical deductions, which lead to a conclusion. Your first introduction to proof was probably in geometry, where proofs were done in two column form. This forced you to make a series of statements, justifying each as it was made. This is a bit clunky. We want to have the same content, but without noting every trivial detail. A proof should read naturally. How much detail should we give? it depends on the mathematical background of the writer and the reader. A professional mathematician will need fewer details to understand the argument than will a beginning student. in both cases, though, the proof should be as clear and concise as possible.

The process of writing a proof forces us to clearly understand the ideas involved. In formulating our thoughts in such a way as to explain the ideas to others, we clarify the ideas to ourselves. This is akin to what happens when we teach someone else. This process shows us where we have gaps in our understanding. This helps us to see the structure behind what we’re trying to show. Mathematics is very much about structure. When we see a new statement, when we’re given a new problem, when we try to write a new proof, we try to see how everything fits into this structure. What do the words mean? What kinds of objects are we talking about? What do we know about such kinds of objects? How do they relate?

Developing the ability to write good proofs takes time and practice. Here’s an example from an actual assignment of how wrong it can go.

Exercise: If **u**, **v**, and **w** are vectors such that **u** + **v** + **w** = 0, prove that span{**u**,**v**} = span{**v**,**w**}

Here was the student’s argument:

The only way this would not be true is if (**u**,**v**) and (**v**,**w**) spanned different things. Since **u** + **v** = -**w**, there can be no dimension W_n in W that does not have a corresponding component in either -U_n or -V_n or some combination of the two that can span the same dimension the same way.

There are a number of things wrong with this. The use of the word dimension is nonsensical. What does n refer to? What does W_n mean? What is a corresponding component? What is $-U_n$ or $-V_n$ or some combination of the two? What does the phrase "span the same dimension the same way" mean? Virtually nothing in this statement is understandable. Does the student at least have the right idea behind the argument? I don't know, but this is a long way from clearly showing the reader why the statement is true.

This guide is not a textbook. It will contain some definitions and some theorems, but it focuses on how to think about linear algebra and how to put these thoughts into clear statements. No book, no teacher, no class can magically give you abilities. Skills are earned through practice and hard work. The assisted proofs in this guide will help you develop your skills, but it is imperative that you write many proofs and rewrite those proofs and rewrite those proofs. Read proofs. Share proofs. Discuss them. Argue them. Don't be afraid to be wrong. Be open to criticism. Critique yourself. If your peers don't understand your arguments, figure out why. If you don't understand your peers' arguments, figure out why.

The theme, which I will restate time and again, is that writing a strong proof is a matter of clear thought and hard work. From the beginning, invest time and energy in understanding the ideas and learning to express them well. There is no substitute for hard work.

Above, we noted that a proof is a sequence of statements justified by axioms, theorems, definitions, and logical deductions, which lead to a conclusion. Let's look at these.

AXIOMS, DEFINITIONS, and THEOREMS

Axioms are the statements in mathematics which we accept without proof. Every proof ultimately falls back to these beginning statements. There are very interesting questions about which statements we should start with in mathematics and what the consequences of so doing are. Mathematicians in the late 19th and early 20th centuries expended a great deal of thought and effort into this. Looking back to these statements, deciding how to choose them, and studying the consequences of doing so is beyond the scope of the linear algebra course, but you are encouraged to look into the subject.

Mathematics is the structure which results from the consequences of our axioms. Our job is to understand parts of this structure. To grapple with mathematics, we have to be able to refer to pieces of it and ask how it all fits together. Definitions focus our questions on certain objects or sets or relationships in mathematics. In general, the broader a definition, the more parts of mathematics it applies to, and the less sharp the theorems about it can be. The more restrictive a definition, the fewer parts of mathematics it applies to, and the sharper the theorems can be. Possibly the most important aspect of writing proofs is to understand the definitions of the words we are using. Often in beginning linear algebra, writing out the definitions involved in our statement is half the battle. When we talk about vector spaces, dimensions, bases, and so on, we have to be absolutely clear about what we are referring to.

Theorems are the statements in mathematics which we know to be true. Typically, we reserve the word for statements which are not immediately obvious, but there’s no hard and fast rule for this. In proving theorems we often try to break our arguments up into digestible pieces so that the organization of the proof is clear. A lemma is a subtheorem that we prove separately so that it can be used in our proof without breaking the flow of the argument. This is akin to developing a subroutine in programming.

SOME BASIC LOGIC

Logical deduction was the fourth element in our list of ingredients for writing proofs. Much of our logical structure is buried in the development of axiomatic structure and set theory. From this we get the theorems we've previously developed in mathematics such as Euclidean geometry, algebra, trigonometry, and calculus. We are fortunate to have this structure to work from, so that we already have a solid box of tools when we start studying linear algebra. We do need some more discussion of the basics of logic, though. We'll look at some symbolic logic now.

Let's start with the following symbols: $P \Rightarrow Q$. Here P stands for a given statement, $\Rightarrow$ means "implies", and Q stands for a conclusion statement.

For example, if P is the statement, "It is raining", and Q is the statement, "Water is falling on the ground", then $P \Rightarrow Q$ says, "It is raining implies water is falling on the ground". Another way that we say this is, "If it is raining, then water is falling on the ground." This is called the if-then form of the statement. This is a logical statement that we would deem to be true, in general.

Another example, if P is the statement, "You didn't answer your phone", and Q is the statement, "You are cheating on me", then $P \Rightarrow Q$ says, "You didn't answer your phone, so you are cheating on me". This is a logical statement that may or may not be true.

Another example, if P is the statement, "The sky is blue", and Q is the statement, "Pigs can fly", then $P \Rightarrow Q$ says, "If the sky is blue, then pigs can fly". This is a logical statement which is not true.

It is clear from the preceding examples that if we want to know if Q is true, then we need to know that P is true and that the implication is valid. When writing proofs, we must check these two things. We must start with statements we know to be true and show the implication is forced, so that Q must be true.

If $P \Rightarrow Q$, we say that P is SUFFICIENT for Q to be true and we say that Q is NECESSARY for P to be true.

THE CONVERSE

The converse of $P \Rightarrow Q$ is $P \Leftarrow Q$ (equivalently, $Q \Rightarrow P$). In general, we don't expect the converse to be true. For example, "If it is raining, then water is falling on the ground" seems reasonable, but "If water is falling on the ground, then it is raining" seems less reasonable to those who own sprinklers. As for a math example, “$x = 3$, so $x^2 = 9$” cannot be reversed to “$x^2 = 9$, so $x = 3$”.

THE DOUBLE IMPLICATION

Sometimes the implication does work both ways. If $P \Rightarrow Q$ and $Q \Rightarrow P$, then we write $P \Leftrightarrow Q$ and we say that P is true if and only if Q is true. An example from mathematics is “$x = 3$ if and only if $x + 2 = 5$”. To prove a double implication, we often have to show each of the implications separately.

THE NEGATION

The negation of the statement P, not P, is written $\sim P$ and it means the opposite of P. For example, if P is “$x = 3$”, then $\sim P$ is “$x \neq 3$”.

THE CONTRAPOSITIVE

If $P \Rightarrow Q$, then the contrapositive is $\sim Q \Rightarrow \sim P$. These are considered logically equivalent. For example, “$x = 3$” implies “$x^2 = 9$” has as the contrapositive “$x^2 \neq 9$” implies “$x \neq 3$”. In proofs this shows up as the technique Proof by Contradiction. If one needs to show $P \Rightarrow Q$, then it suffices to show $\sim Q \Rightarrow \sim P$. This technique will be examined in its own section.

EXAMPLE

This example is from Lewis Carroll:

Babies are illogical.

Nobody is despised who can manage a crocodile.

Illogical persons are despised.

We can write this symbolically by doing the following:

Define P = { a person is a baby },

Q = { a person is illogical },

S = { a person is despised }, and

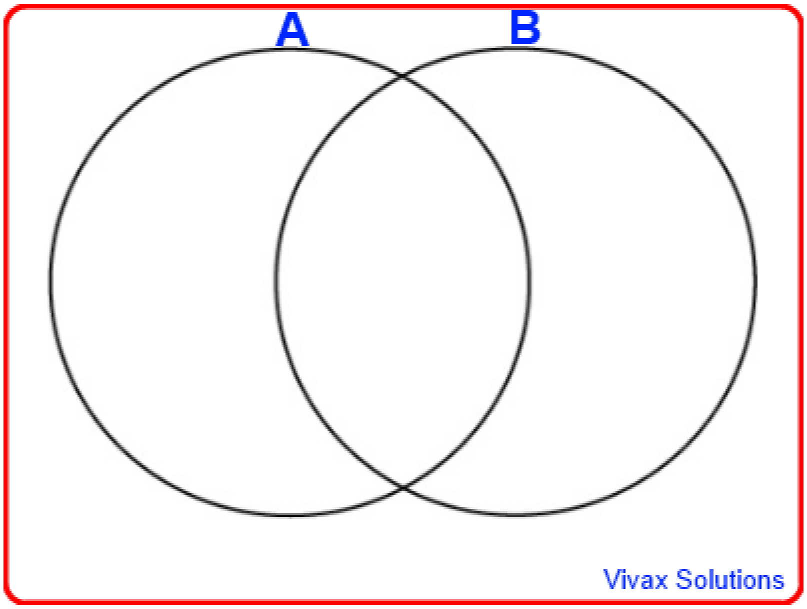
T = { a person can manage a crocodile }.

The statements then become P ⇒ Q, T ⇒ ~S, and Q ⇒ S. The contrapositive of the second statement is S ⇒ ~T. Stringing these together gives us P ⇒ Q ⇒ S ⇒ ~T, which gives us P ⇒ ~T. Our conclusion, then, is the statement, "Babies cannot manage crocodiles". This process of stringing together logical statements is called SYLLOGISM.

Another type of syllogism involves inclusions or conditions. These can often be aided by Venn diagrams. Consider the following example:

If some decisions are careful reflections and all uses of free will are decisions, do we know that some uses of free will are careful reflections?

Let's set A as the set of decisions, B the set of careful reflections, and C the set of uses of free will. The Venn diagram for "some decisions are careful reflections" is



"All uses of free will are decisions" means that we put a circle for C inside of A. Does that circle have to intersect B? No, it can be drawn in the left hand side of A, so we don't know that some uses of free will are careful reflections. They may be, but it is not a logical consequence of the given statements.

EXERCISES

Now try to determine whether the following statements follow logically from the given statements.

Statement 1: If Dilbert is to finish the project, he will have to move out of his cubicle. But Dilbert will get to move out of his cubicle only if he strokes Catbert. So for Dilbert to finish the project, he is going to have to stroke Catbert.

Statement 2: Some directors of human resources are cats, and some cats purr when stroked. So some directors of human resources purr when stroked.

Statement 3: All men are mortal. Socrates is a man. Therefore, Socrates is mortal.

Statement 4: All barbiturates are drugs. Marijuana is not a barbiturate. So marijuana is not a drug.

Statement 5: No free choices are caused occurrences. Some natural processes are not caused occurrences. So some natural processes are not free choices.

Now try a couple of questions.

Question 1: from *The Merchant of Venice:*

Portia was a woman desired by many men. It was arranged she would marry the man who could correctly guess which of three caskets contained her portrait. One casket was inscribed with, "Who chooseth me shall gain what many men desire." One man concluded that, since many men desired Portia, her portrait must be in that casket. Was this logically sound?

Question 2: What is the conclusion to the following statement?

Anyone who speaks in tongues does not speak to men but to God. But everyone who prophesies speaks to men for their strengthening, encouragement, and comfort. (Paul, I Corinthians, 14:2-3)

Putting together the definitions and theorems with logical connectors to prove our statements is called DEDUCTIVE REASONING. Let’s look at a technique which is a good place to start in our understanding, but which is NOT proof: INDUCTIVE REASONING.

INDUCTIVE REASONING

When trying to understand a new concept or when trying to decide whether or not a statement is true, it is generally helpful to look at examples. Understanding how something works for specific examples often leads to how something works in general. Coming up with good examples is a vital skill in learning mathematics and in developing proofs. Looking at a number of examples, coming up with a statement that is true about those examples, and then proposing that the statement is true more generally is called INDUCTIVE REASONING. We use this a great deal and to great benefit in our everyday life. If my girlfriend has gotten mad every time I've talked with my mouth full, then she's probably going to get mad the next time I do it. If the sun has come up every morning, it's probably going to come up tomorrow morning. In mathematics, however, inductive reasoning never constitutes proof unless we can look at every single possible example. Inductive reasoning is good for setting up hypotheses. It's good for looking for ideas. But it's not good for proof.

An extreme example of this is a common error when students first start writing proofs. Asked to prove something in general, the student will often give a single example and be done. This would be like saying, "my mom's name is Susan so all moms are named Susan." Or when asked a question about matrices in general, the student will randomly assume that the matrix is 2X2. This is not adequate for proof and inductive reasoning does not always lead to correct statements.

In the 17th century, the mathematician Fermat realized that for a number of the form $2^k + 1$ to be a prime, k would have to have the form 2^n. So he started looking at numbers of the form $2^{2^n} + 1$. These are called Fermat numbers. For n = 0, 1, 2, 3, 4, the Fermat numbers are 3, 5, 17, 257, and 65537. These Fermat numbers are prime. But their size grows very fast and for n = 5, the Fermat number is already 4294967297. This number was too big for Fermat to determine whether it was prime are not and the numbers after that were way, way too big. But, reasoning inductively from a small sample, he hypothesized that all Fermat numbers are prime. This was wildly wrong. Not a single other Fermat number studied has proven to be prime and mathematicians now strongly suspect that the number of Fermat numbers which are prime is finite.

Even a very large number of examples can lead us astray. There are functions in number theory called $\pi(x)$ and $li(x)$. In the early 20th century all evidence pointed to $li(x)$ being bigger than $\pi(x)$. It's true for x up to some enormously large number, a number so big that we're only confident that it is smaller than something like $10^{10^{10^{34}}}$. But it turns out that which one is bigger, $\pi(x)$ or $li(x)$, switches back and forth an infinite number of times.

The point is that you cannot prove statements by example unless you show every possible example is true. If you want to show something is true about symmetric matrices, it is not possible to list out every symmetric matrix. If you want to show that the composition of linear transformations is a linear transformation, it is not possible to list out every linear transformation. If you want to show that a property is independent from a change of basis, it is not possible to consider every change of basis individually.

HOW TO APPROACH A PROOF

The approach to writing a proof is much like the approach to solving word problems. The first thing is to read through the problem and make sure you understand exactly what the problem is asking. What do the words mean? What information is given? What do you know about the topic under discussion? Are there any diagrams or pictures that will help you? Throughout this guide, whenever a proof is being discussed, you'll be asked leading questions. Pay attention to the questions. These are the questions you should be asking yourself. Learning to ask yourself appropriate questions is a very important skill in learning to write proofs. Write down your questions. Write down the definitions of the terms involved. Write in complete sentences. Make sure these sentences are meaningful. Make sure the statements are correct.

The mathematician, George Polya, developed the following four-step approach to problem solving:

Step 1: Understand the Problem

Step 2: Devise a Plan

Step 3: Carry Out the Plan

Step 4: Look Back

This is simplistic, but it does give a place to start. He also suggested the following heuristics for when you are struggling with a problem or proof:

Analogy: Can you find a problem analogous to your problem and solve that?

Generalization: Can you find a problem more general than your problem?

Induction: Can you solve your problem by deriving a generalization from some examples?

Variation: Can you vary or change your problem to create a new problem (or set of problems) whose solution(s) will help you solve your original problem?

Auxiliary: Can you find a subproblem or side problem whose solution will help you solve your problem?

Relation: Can you find a problem related to yours that has already been solved and use that to solve your problem?

Specialization: Can you find a problem more specialized?

Decomposition: Can you decompose the problem and "recombine its elements in some new manner"?

Reversal: Can you start with the goal and work backwards to something you already know?

Drawing: Can you draw a picture of the problem?

Extension: Can you add some new element to your problem to get closer to a solution?

Working with others can be extremely helpful. Having someone to bounce ideas off of, having someone to read your work, having someone whose work you can critique, are all beneficial. But remember that your peers cannot do your learning for you. Your parents cannot learn for you. Your teacher cannot learn for you. This is a struggle that you must go through yourself. Mathematics can be hard and confusing. You are going to get stuck. You have to give yourself somewhere to go. You have to have the tenacity to get yourself unstuck.

A GEOMETRIC EXAMPLE

Here we are looking at how the proof process should work and we'll write out the completed final proof. Often this guide will leave questions open ended or fill-in-the-blank, but for now we'll just walk through the process.

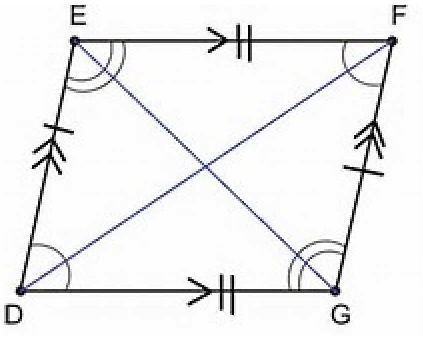
Let's consider a geometric example which requires knowing the basics of vectors.

Prove that the diagonals of a parallelogram bisect one another.

How should we approach this problem? First we need to know the meanings of the words. What is a parallelogram? What are the diagonals? What does it mean to bisect? This is a geometric problem so it is natural for us to draw a picture. Once we do, we need to translate the words and images into mathematical symbols.

Draw the picture.

What things can we add to this picture? We can draw the diagonals. We can label the vertices. We can draw arrows for vectors.



This is a two-dimensional picture so we should be able to choose two vectors in this picture somehow and then reference all the other vectors here by those two. It doesn't really matter which two, as long as they're not parallel. For example, we could choose the vector from D to E and the vector from D to G but there is no one particular RIGHT answer for which to choose. We could also choose, say, E to G and F to D.

Let's say that **u** equals the vector from D to E and **v** equals the vector from D to G. Add this information to your picture. Also label the point of intersection of the diagonals. Let's call it I. Using **u** and **v**, how do we represent the vector from D to F? the vector from D to G? the vector from E to G? the vector from E to D? etc. Play around with these.

You'll notice that we don't yet know how to write the vectors involving I in terms of **u** and **v** since we haven't yet proven our proposition. Now, what do we want to show?

We want to show that the vector from D to I is half of the vector from D to F and we want to show that the vector from E to I is half the vector from E to G. How do we represent half the vector from D to F? How do we represent half the vector from E to G? Add these vectors to your list. Now try representing various vectors in the picture. In terms of **u** and **v**, what happens when we add the vector from D to E and half the vector from E to G?

Now let's try to put the pieces together. When we add the vector from D to E and half the vector from E to G, we should land at I. In terms of **u** and **v**, this should be **u** + ½(**v** – **u**). Simplify this. Is it in the direction of D to F? Does it give us I?

Finally, let's write our proof.

Proof: Given the parallelogram DEFG, let **u** be the vector from D to E and **v** be the vector from D to G. Then **v** – **u** is the vector from E to G and **v** + **u** is the vector from D to F. **u** + ½(**v** – **u**) = ½(**v** + **u**) is then a vector from D to a point on the diagonal from E to G. Since **v** + **u** is the vector from D to F, this point is also on the diagonal from D to F, making this the point of intersection of the diagonals. The vector from E to this point is ½ the vector from E to G and the vector from D to this point is ½ the vector from D to F. This shows that the diagonals of the parallelogram bisect one another.

A LINEAR INDEPENDENCE EXAMPLE

Now let's consider an example from early in the course involving linear independence.

Question: If **u**, **v**, **w** ∈ $\mathbf{R}^n$ are linearly independent, does it follow that (**u** + **v**), (**v** + **w**), and (**u** + **w**) are also linearly independent? If the implication holds, prove it. If the implication is false, provide a counterexample.

It is certainly reasonable to start by looking at some examples in $\mathbf{R}^3$ or $\mathbf{R}^4$ or $\mathbf{R}^5$ (why can't we look at examples in $\mathbf{R}^2$?). Let's ask the following questions and make sure that we know the answers before moving on. How do we test for a linear independence? Along the same lines, what do we know about vectors which are linearly independent? To test for the independence of (**u** + **v**), (**v** + **w**), and (**u** + **w**), what equation do we need to set up? Can this equation be transformed into an equation about **u**, **v**, and **w**? Why would we do this? What system of equations does this lead to? How many techniques do we have to test for the number of solutions to this system? How can we put all of this together? With these pieces, try writing a proof or finding a counterexample.

Once you've attempted your proof or found your counterexample, compare it to the following:

Answer: Consider the equation $c_1(\mathbf{u} + \mathbf{v}) + c_2(\mathbf{v} + \mathbf{w}) + c_3(\mathbf{u} + \mathbf{w}) = 0$. This is equivalent to $(c_1 + c_3)\mathbf{u} + (c_1 + c_2)\mathbf{v} + (c_2 + c_3)\mathbf{w} = 0$. Since **u**, **v**, and **w** are linearly independent, $(c_1 + c_3)$, $(c_1 + c_2)$, and $(c_2 + c_3)$ are each equal to zero. This is a linear system of three equations in three unknowns:

$$\begin{aligned} c_1 + c_3 &= 0 \\ c_1 + c_2 &= 0 \\ c_2 + c_3 &= 0 \end{aligned}$$

Using either Gaussian elimination or noting that the determinant of the associated matrix is not zero, we get $c_1 = c_2 = c_3 = 0$. This implies that (**u** + **v**), (**v** + **w**), and (**u** + **w**) are linearly independent.

ANOTHER LINEAR INDEPENDENCE EXAMPLE

Now let's try to do a guided proof. Again, make sure to answer all of the questions leading us to the proof. Then fill in the blanks in the proof and pay attention to how the proof itself is constructed.

Prove: Every subset of a finite linearly independent set is linearly independent

Let's start by giving names to the set and the elements of the set. I'm going to choose to call the set U and the elements of the set u_1, u_2, …, u_n. To prove something for every subset it suffices to prove it for a generic subset. So let V be any subset of U and call the elements of the subset v_1, v_2, …, v_m. V^c (V complement) is the set of elements in U which are not in V. Let's call the elements of V^c v_{m+1}, …, v_n. Note that the sets $\{u_i\}$ and $\{v_j\}$ are the same but probably in different orders.

Now, what does it mean for the set U to be linearly independent?

____

What does it mean for the set V to be linearly independent?

____

Think about how we can connect these two.

We're going to use the word "complement" and the phrase "proper subset." What do they mean?

PROOF

The empty subset and the subset of all elements trivially satisfy the theorem. Let U be a finite linearly independent set with elements ________________________ and let V be a proper nonempty subset of U with elements ______________________. Let V^c be the complement of V in U. It has elements __________________. The set $\{u_1, u_2, \ldots,$ ____$\}$ is the same as the set $\{v_1, v_2, \ldots,$ ___$\}$, so $v_1, v_2, \ldots, v_n$ are linearly ________________. Let $c_1v_1 + c_2v_2 + \ldots + c_mv_m =$ ______. Then $c_1v_1 + c_2v_2 + \ldots + c_mv_m + 0\ v_{m+1} + \ldots +$ _______ = ______. $v_1, v_2, \ldots, v_n$ are _____________________, so $c_1 = c_2 = c_m = 0 = 0 = \ldots = 0$. But this means that ____, ____, …, ____ are linearly independent. Since the empty set, U, and any nonempty subset are linearly independent, every subset of a finite linearly independent set is linearly independent.

PROOF BY CONTRADICTION

The proofs we have seen to this point have been constructive in nature. We started from given information and deduced our conclusion. There are a couple of other approaches to writing proofs, one of which is proof by contradiction.

As noted above, proof by contradiction is the logical contrapositive. When might we use this? Mathematics often distinguishes between something and its opposite. For example, rationals and the irrationals, something happens or it does not, true or false. Sometimes it is easier to work on one side than it is on the other. The rationals have a lot of structure, but the irrationals do not. If you roll a pair of dice five times, it is easier to computer the probability of never rolling a 12 than of rolling at least one 12.

Let's look at a classical example, not from linear algebra.

THEOREM: There are an infinite number of primes.

Which is easier to work with, an infinite number or a finite number? The answer is a finite number because then we can just list them out. What is the logical implication that we're trying to prove? If the mathematics we have developed to this point is correct then there are an infinite number of primes. What is the contrapositive? Under the format $P \Longrightarrow Q$, P here is "the mathematics we have developed to this point is correct" and Q is "there are an infinite number of primes". Remember that the contrapositive is $\sim Q \Longrightarrow \sim P$. $\sim Q$ is "there are a finite number of primes" and $\sim P$ is "the mathematics we have developed to this point is not correct". Note that $\sim Q$ is now included in the mathematics we have developed to this point because it's the given in our logical implication. This phrase, "the mathematics we have developed to this point is not correct" has two possibilities. The first possibility is that there's something fundamentally wrong with mathematics. The second possibility is that our assumption $\sim Q$ is not correct. We work from the presumption that the first possibility has been addressed and disposed of. This leaves the second possibility which would be a contradiction of our assumption. Hence the phrase, "proof by contradiction". The idea is to assume $\sim Q$ and show that this leads to a contradiction.

IDEA OF PROOF OF THEOREM: Assume that there are a finite number of primes. List them out. We need to give them names. A standard way to create a list of unknown values is to choose a variable and then use subscripts. Standard letters for primes are p and q. So let's refer to our primes as $p_1, p_2, \ldots, p_n$, where n is the finite number of primes which exist. Remember that this was the point of choosing proof by contradiction, so that we can work with a finite list of primes. We obviously don't know exactly how many there are, but it is some number, so we give it a name. The letter n is common for enumeration. Now we have to figure out what we're going to do with these primes. What contradiction might we be looking for?

We know that ultimately we think that there are an infinite number of primes but we only have a finite list. A bunch of them must be missing. So let's find a missing prime. The idea that we're going to use is that two consecutive numbers cannot both be multiples of three or multiples of five or multiples of any other prime. Is there a number that is a multiple of all the primes in our list? Sure, multiply them all together. This product is a multiple of each of the primes. But that means that the next number is not a multiple of any of the primes. That's a problem because every number not on the list has to factor into a product of primes which are on the list. This means that it has to be a multiple of some prime on the list. There's our contradiction. Now let's put all of this together into a nice clean proof.

PROOF OF THEOREM: Assume that there are only a finite number of primes. Call them p_1, p_2, …, p_n. Consider the product of all the primes $p_1p_{2\ldots}p_n$ and the number following it, $p_1p_{2\ldots}p_n + 1$. Each prime is a factor of the product, so no prime is a factor of $p_1p_{2\ldots}p_n + 1$. This is a contradiction as every integer has a prime factorization. Therefore, there are an infinite number of primes.

PROOF BY INDUCTION

Another method of proof is proof by induction. This only works when we have a list of statements we wish to prove and each statement depends on the one before it.

Proof by induction is not the same thing as inductive reasoning. Inductive reasoning is considered in a previous section where we discuss how proof can never be done by inductive reasoning unless we can examine every single example. Proof by induction, however, is lining up a sequence of statements, showing that the truth of a statement in the sequence implies the truth of the next statement in the sequence, and showing that the first statement is true. An analogy is to think of the statements as dominoes. We line up the dominoes and push the first one to knock them all down.

Let's say that we have a sequence of statements, P_1, P_2, …. Lining up the dominoes is to show the inductive step that $P_n \Longrightarrow P_{n+1}$, i.e., if statement P_n is true then statement P_{n+1} is true. Pushing the first domino is showing that statement P_1 is true.

Here's an example:

Prove: If A is an invertible matrix, then A^n is an invertible matrix with $(A^n)^{-1} = (A^{-1})^n$

for all positive integers n.

What tips us off that this might be appropriate for proof by induction? We already know that this is true for n = 1 because that is just the statement that A is invertible. And this is really a sequence of statements, P_2 is $(A^2)^{-1} = (A^{-1})^2$, P_3 is $(A^3)^{-1} = (A^{-1})^3$, etc. So what do we need? We have the truth of P_1, which is $A^{-1} = A^{-1}$. Now we need $P_n \Longrightarrow P_{n+1}$. In other words, we need to show that if $(A^n)^{-1} = (A^{-1})^n$ then $(A^{n+1})^{-1} = (A^{-1})^{n+1}$. How do we do that? Think about it and then let's write our proof.

Proof: 1) A is invertible. $(A^1)^{-1} =(A)^{-1} = (A^{-1}) = (A^{-1})^1$

2) Given $(A^n)^{-1} = (A^{-1})^n$, $(A^{n+1})^{-1} = (A^nA)^{-1} = A^{-1}(A^n)^{-1} = A^{-1}(A^{-1})^n = (A^{-1})^{n+1}$

By induction, our statement is proven.

Here's another example where we need to be familiar with matrix multiplication:

Show that $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}, n = 1,2,3, \ldots$

As above, this looks to be appropriate for proof by induction. What is our initial statement? Is it true? What is the inductive step? Let's write our proof.

Proof: 1) $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}^{1} =$

____

2) Given $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}^{n} = \begin{pmatrix}1 & n\\ 0 & 1\end{pmatrix}$, $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}^{n+1} = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}^{\_\_}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}^{\_\_}$

$$= \left(\begin{matrix}\_\_ & \_\_\\ \_\_ & \_\_\end{matrix}\right)\left(\begin{matrix}\_\_ & \_\_\\ \_\_ & \_\_\end{matrix}\right) = \left(\begin{matrix}\_\_ & \_\_\\ \_\_ & \_\_\end{matrix}\right)$$

By induction, our statement is proven.

VECTOR SPACES WITH NO ADDITIONAL STRUCTURE

Something to keep in mind is that at the vector space level every n-dimensional vector space over the reals is exactly the same as every other n-dimensional vector space over the reals. You may be thinking that the space of 2 X 2 matrices, $M_{2,2}$, is different from R^4. Both are 4-dimensional spaces, but you can multiply matrices, whereas you cannot multiply vectors in R^4. That's true, but that multiplication is not part of the vector space structure. The vector space structure is only about linear combinations of the vectors. Additional structure, such as an inner product or vector multiplication, is very important, but it is not part of the vector space structure itself. This can be confusing because we do a lot of matrix manipulation in dealing with vector spaces. For example, the standard basis for R^4 is e_1 = (1,0,0,0), e_2 = (0,1,0,0), e_3 = (0,0,1,0), and e_4 = (0,0,0,1). So when we want to represent (1,2,3,4) in vector form relative to this basis, it's easy. It's $\begin{bmatrix}1\\2\\3\\4\end{bmatrix}$. The standard basis for $M_{2,2}$ is $e_{11} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$, $e_2 = \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}$, $e_3 = \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}$, and $e_4 = \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$. Now when we want to represent $\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$ in vector form relative to this basis, we get the same thing, $\begin{bmatrix}1\\2\\3\\4\end{bmatrix}$. In both cases, if we want to use matrices to manipulate the vectors in the vector spaces, it's this form, $\begin{bmatrix}1\\2\\3\\4\end{bmatrix}$, that we need to work with.

Once we start adding more structure, things get more complicated. In general, the more structure you impose, the mathematical places to which it is applicable become fewer but the theorems become sharper and more abundant.

MATRIX MULTIPLICATION

Part of the skill in developing and writing proofs is in how one thinks of the concepts involved. Here's a discussion on how to think about matrix multiplication.

Given matrices A and B, the product AB is defined to be the matrix C = [c_{ij}], where c_{ij} is the dot product of the i^{th} row of A and the j^{th} column of B. This definition is sometimes the way to think about the product but there are two other productive ways to see it. You can think of the matrix A acting on the rows of the matrix B or you can think of the matrix B acting on the columns of the matrix A. The first row of the matrix C is a linear combination of the rows of the matrix B. The coefficients of the linear combination are the elements of the first row of the matrix A. Similarly, the first column of the matrix C is a linear combination of the columns of the matrix A. The coefficients of the linear combination are the elements of the first column of the matrix B.

For example, if A = $\begin{bmatrix} 1 & 3 & -2 \\ 2 & 1 & 0 \\ -1 & 0 & 4 \end{bmatrix}$ and B = $\begin{bmatrix} -1 & 2 & 4 \\ 0 & 1 & 2 \\ 3 & -5 & 3 \end{bmatrix}$, then AB = $\begin{bmatrix} -7 & 15 & 4 \\ -2 & 5 & 10 \\ 13 & -22 & 8 \end{bmatrix}$.

Notice that $[-7 \quad 15 \quad 4]$ = $1[-1 \quad 2 \quad 4] + 3[0 \quad 1 \quad 2] - 2[3 \quad -5 \quad 3]$ and that

$$\begin{bmatrix} -7 \\ -2 \\ 13 \end{bmatrix} = -1\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + 0\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} -2 \\ 0 \\ 4 \end{bmatrix}.$$

Of course, there's nothing special about the first row or first column. The i^{th} row of C is still a linear combination of the rows of B, but now the coefficients are the elements of the i^{th} row of A. The j^{th} column of C is still a linear combination of the columns of A, but now the coefficients are the elements of the j^{th} column of B.

Thinking of matrix multiplication in these various ways can sometimes help in constructing proofs. For example, in the linear system $\begin{cases} a_{11}x + a_{12}y = c_1 \\ a_{21}x + a_{22}y = c_2 \end{cases}$, we may want to know which values of c_1 and c_2 make the system consistent. In matrix multiplication form this is $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$. Instead of thinking of this product as a collection of dot products or of the matrix A =$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ acting on B =$\begin{bmatrix} x \\ y \end{bmatrix}$, let's think of B acting on the columns of A. This says that $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = x\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + y\begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}$. It is immediate from this that $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ is a linear combination of the columns of A and hence is in the column space of A. If we consider all the possible values of x and y, we see that the values of c_1 and c_2 which make the system consistent are exactly those for which $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ Is in the column space of A.

Here's another example.

Prove: If A and B are matrices such that B has a column of zeroes and the product AB is defined, then AB also has a column of zeroes.

Do we want to think of A acting on the rows of B or B acting on the columns of A? Either will work, but if the ith column of B is the column of zeroes, how does B act on the columns of A to generate the ith column of AB? Let's put this together to generate our proof.

Proof: Let u_1, …, u_n be the columns of A and let v_i be a column of zeroes of B. Then v_i = $\begin{bmatrix} b_{1i} \\ \dots \\ b_{ni} \end{bmatrix} = \begin{bmatrix} 0 \\ \dots \\ 0 \end{bmatrix}$. The ith column of AB is $b_{1i}u_1 + \dots + b_{ni}u_n = 0u_1 + \dots + 0u_n = \begin{bmatrix} 0 \\ \dots \\ 0 \end{bmatrix}$.

On the other hand, if A and B are matrices such that B has a row of zeroes and the product AB is defined, then do we get a row of zeroes in AB? Remember that the columns of B act on the columns of A and the rows of A act on the rows of B, so this doesn't sound too promising. To show that a hypothesis fails, all we have to do is find one example where the hypothesis is not true. The opposite of a statement always being true is not that it is always false, it's that it is false at least once. Can you find such an example here? Keep it simple and look at small matrices.

COMMUTATIVE DIAGRAMS

Visual aids are often helpful in mathematics. When we talk about mappings or functions we need to know what space we're mapping from, what space we're mapping to, and what the mapping does to each element. Often, multiple mappings are involved in our work. Diagrams can help us keep track of these mappings. To designate a mapping from A to B, we write

$$A \to B$$

We may position these arrows pointing left, right, up, down, or diagonally. In general, we draw our diagrams in whatever way seems most clear to us. Sometimes we may have different paths to get from one place in the diagram to another. Typically, we want the diagram to be independent of path. If all directed paths in the diagram with the same beginning points and endpoints give the same result then we say that the diagram is COMMUTATIVE.

As an example, consider the following diagram where U and V are vector spaces, T is a linear transformation from U to V, and $T \times T$ is the mapping from $U \times U$ to $V \times V$ defined by $(T \times T)\ (u_1, u_2) = (T(u_1), T(u_2))$:

$$\begin{tikzcd}
U \times U \arrow[r, "+"] \arrow[d, "T \times T"'] & U \arrow[d, "T"] \\
V \times V \arrow[r, "+"] & V
\end{tikzcd}$$

The only starting point and endpoint with multiple paths in this diagram is from $U \times U$ to V. The first path is

$$U \times U \xrightarrow{+} U \xrightarrow{T} V$$

which we get by going right and then down. The second path is

$$U \times U \xrightarrow{T \times T} V \times V \xrightarrow{+} V$$

which we get by going down and then right. Saying that a linear transformation preserves addition is equivalent to saying that these two paths give the same results, which is equivalent to saying that this diagram commutes.

An example where commutative diagrams are helpful is in change of bases. Let's say for example that T is a linear transformation from U to V, that $\mathcal{B}$ and $\mathcal{B}'$ are bases of U, and that $\mathcal{C}$ and $\mathcal{C}'$ are bases of V. Let $U_{\mathcal{B}}$ be the set of vectors of U represented as column vectors under the basis $\mathcal{B}$, $U_{\mathcal{B}'}$, the set of vectors of U represented as column vectors under the basis $\mathcal{B}'$, $V_{\mathcal{C}}$

the set of vectors of V represented as column vectors under the basis $\mathcal{C}$, $V_{\mathcal{C}'}$ the set of vectors of V represented as column vectors under the basis $\mathcal{C}'$, and let P be the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}'$, and Q be the change of basis matrix from $\mathcal{C}$ to $\mathcal{C}'$. Finally, let $T_{\mathcal{B}\mathcal{C}}$ be the matrix representation of T from $U_{\mathcal{B}}$ to $V_{\mathcal{C}}$ and $T_{\mathcal{B}'\mathcal{C}'}$ be the matrix representation of T from $U_{\mathcal{B}'}$ to $V_{\mathcal{C}'}$.

Then we have the following commutative diagram which lets us see everything that is going on:

$$\begin{tikzcd}
U_{\mathcal{B}} \arrow[r, "T_{\mathcal{B},\mathcal{C}}"] \arrow[d, shift right, "P"'] & V_{\mathcal{C}} \arrow[d, shift right, "Q"'] \\
U_{\mathcal{B}'} \arrow[r, "T_{\mathcal{B}',\mathcal{C}'}"] \arrow[u, shift right, "P^{-1}"'] & V_{\mathcal{C}'} \arrow[u, shift right, "Q^{-1}"']
\end{tikzcd}$$

LINEAR TRANSFORMATIONS

It is common in mathematics to look at maps between spaces. Generally, the maps which are most interesting are those which preserve structure. If the map is from a vector space to a vector space then we call the map a linear transformation. What do we mean by preserving structure and what structure are we preserving? Vector spaces have two operations: addition and scalar multiplication. Preserving addition means that we get the same result whether we add first and then map or map first and then add. Similarly, preserving scalar multiplication means that we get the same result whether we multiply by a scalar first and then map or map and then multiply by a scalar.

Let's call our vector spaces V and W and our map T. We write this as T : V → W. If T maps $\mathbf{v}$ to $\mathbf{w}$, we write T($\mathbf{v}$) = $\mathbf{w}$. Preserving addition means that if $\mathbf{v_1} + \mathbf{v_2} = \mathbf{v}$ and if $\mathbf{w_1} + \mathbf{w_2} = \mathbf{w}$, where T($\mathbf{v_1}$) = $\mathbf{w_1}$ and T($\mathbf{v_2}$) = $\mathbf{w_2}$, then T($\mathbf{v}$) = $\mathbf{w}$. Graphically this can be seen as the following commutative diagram.

$$\begin{tikzcd}
U \times U \arrow[r, "+"] & U \\
(u_1,u_2) \arrow[r] \arrow[d, "T \times T"'] & u_1 + u_2 \arrow[d, "T"] \\
V \times V \arrow[r, "+"] & V \\
(T(u_1),T(u_2)) \arrow[r] & T(u_1)+T(u_2) \\
 & = T(u_1 + u_2)
\end{tikzcd}$$

Similarly, preserving scalar multiplication means that if $c\mathbf{v_1} = \mathbf{v}$ and if $c\mathbf{w_1} = \mathbf{w}$ where T($\mathbf{v_1}$) = $\mathbf{w_1}$, then T($\mathbf{v}$) = $\mathbf{w}$. Graphically this can be seen as the following commutative diagram.

$$\begin{tikzcd}
R \times U \arrow[r, "\times"] & U \\
(c,v_1) \arrow[r] \arrow[d, "I \times T"'] & cv_1 \arrow[d, "T"] \\
R \times V \arrow[r, "\times"] & V \\
(c,T(v_1)) \arrow[r] & cT(v_1) = T(cv_1)
\end{tikzcd}$$

Prove: If A = $\begin{bmatrix} \quad a_{ij} \quad \end{bmatrix}$ and B = $\begin{bmatrix} \quad b_{ij} \quad \end{bmatrix}$ are n x n upper triangular matrices, then AB is upper triangular.

What does it mean to be upper triangular? Think about this both graphically and algebraically. What is the algebraic definition? For what values of *i* and *j* is a_{ij} = 0? b_{ij}? How does one compute the *ij*th entry of AB?

Proof: Let c_{ij} be the *ij*th entry of AB. For $i > j$, c_{ij} = $\sum_{k=1}^{n}$ ________ = $\sum_{k=1}^{i-1}$ ________ + $\sum_{k=i}^{n}$ ________. But for $1 \leq k < i$, a_{ik} = ____ and for $i \leq k \leq n$, b_{kj} = _____. So $\sum_{k=1}^{i-1}$ ________ = _____ and $\sum_{k=i}^{n}$ ________ = _____, and so $\sum_{k=1}^{n}$ ________ = c_{ij} = ____. By definition, AB is upper triangular.

BILINEAR AND MULTILINEAR MAPS

A linear transformation, as we saw above, is a linear map from a vector space to a vector space. If we want to combine two vectors to get an output, we can think of this as a map from the direct product of two vector spaces to a target space. Again, we're going to want the map to preserve linear structure. This means that it needs to preserve the structure in the first space and in the second. The dot product is a good example of this. Distribution of the dot product across sums is the preservation of addition. Factoring a constant from either of the factors is the preservation of scalar multiplication. Let's see how this looks in function notation. I will use D as the symbol for the mapping representing the dot product in R^n. What should be the input for D? What should be the output for D? When we take the dot product of two vectors we get a scalar, so the input for D should be two vectors and the output for D should be a scalar. So we have

$$D: R^n \text{ x } R^n \rightarrow R, \ D(\mathbf{u},\mathbf{v}) = \mathbf{u}\cdot\mathbf{v}.$$

The distributive laws are then

$$D((\mathbf{u_1} + \mathbf{u_2}), \mathbf{v}) = D(\mathbf{u_1},\mathbf{v}) + D(\mathbf{u_2},\mathbf{v}) \qquad [(\mathbf{u_1} + \mathbf{u_2}) \cdot \mathbf{v} = \mathbf{u_1} \cdot \mathbf{v} + \mathbf{u_2} \cdot \mathbf{v}]$$

$$D(\mathbf{u}, (\mathbf{v_1} + \mathbf{v_2})) = D(\mathbf{u}, \mathbf{v_1}) + D(\mathbf{u},\mathbf{v_2}) \qquad [\mathbf{u}\cdot (\mathbf{v_1} + \mathbf{v_2}) = \mathbf{u} \cdot \mathbf{v_1} + \mathbf{u} \cdot \mathbf{v_2}]$$

And factoring through a scalar is

$$c\ D(\mathbf{u},\mathbf{v}) = D(c\mathbf{u},\mathbf{v}) = D(\mathbf{u},c\mathbf{v}) \qquad [c(\mathbf{u}\cdot\mathbf{v}) = (c\mathbf{u})\cdot\mathbf{v} = \mathbf{u}\cdot(c\mathbf{v})]$$

Any map from the direct product of two spaces satisfying the above rules is called BILINEAR. A map from the direct product of an arbitrary number of spaces which is linear in each component, similar to the above, is called MULTILINEAR. The most common multilinear map in beginning linear algebra is the determinant, where the input vectors are the columns and the output is a scalar. You can use the multilinearity of the determinant to develop Cramer's Rule.

CRAMER'S RULE

Cramer's Rule says that given the system of equations

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$
$$\vdots$$
$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n,$$

if the system has a unique solution, then the solution is

$$x_i = \frac{D_i}{D}, \text{ where}$$

$$D = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix},$$

and D_i is the same as D, but with the i^{th} column $\begin{matrix} a_{1i} \\ \vdots \\ a_{ni} \end{matrix}$ replaced by $\begin{matrix} b_1 \\ \vdots \\ b_n \end{matrix}$.

This is rather wordy and not always useful. For larger systems being done by hand, computing the determinants is more cumbersome than doing Gaussian elimination. But for systems of three equations and three unknowns it may be useful. For systems of two equations and two unknowns it is very quick.

How might we prove this? The fact that it involves determinants certainly implies that we should write the system in matrix form. Since we're manipulating columns, we should be thinking of the determinant as a function of the columns. We know that the determinant is a multilinear function of the columns. Can we use that?

In matrix form, our system is

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Again, we are focusing on columns. We have seen that we can think of matrix multiplication in three different ways. What are they? Which one involves columns? What does this tell us about the column vector $\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ as a linear combination of the column vectors of

$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$? How can we use this in the definition of D_i? Let's put all of this together for a proof.

PROOF OF CRAMER'S RULE

Given the system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n, \end{aligned}$$

the matrix form is ____.

This means that $\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ = ___ $\begin{bmatrix} \_\_ \\ \vdots \\ \_\_ \end{bmatrix} +$ ___ $\begin{bmatrix} \_\_ \\ \vdots \\ \_\_ \end{bmatrix} + \cdots +$ ___ $\begin{bmatrix} \_\_ \\ \vdots \\ \_\_ \end{bmatrix}$.

$D_i = \begin{bmatrix} a_{11} \cdots & a_{1(i-1)} & b_1 & a_{1(i+1)} & \cdots a_{1n} \\ a_{21} \cdots & a_{2(i-1)} & b_2 & a_{2(i+1)} & \cdots a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} \cdots & a_{n(i-1)} & b_n & a_{n(i+1)} & \cdots a_{nn} \end{bmatrix}$. Substituting for the$i^{th}$column gives

$D_{i=}$

____.

Multilinearity of the columns gives

$D_{i=}$ + … + + _____ D +

____________________________ ____________________________

+ + … +

___________________________ ______________________________

But we know that if two columns in a determinant are equal, then the determinant is ____. So the above long equation simplifies to _____________. We know that D is not zero because ____. Solving for x_i then gives us our result.

INJECTIVITY

Injectivity is essentially the notion of the horizontal line test. In linear algebra, we don't have x- and y-axes to represent our inputs and outputs for mappings, but we still do have inputs and outputs. The horizontal line test in college algebra is a visual way to check that no output can be associated with more than one input. In linear algebra we have to write this condition algebraically. If T is a mapping from U to V, and v is any element of V which is in the range of T, we need to check that there can only be one element u in U mapping to v. We do this by assuming that two outputs match; $T(u_1) = T(u_2)$. If these outputs are our v, then both u_1 and u_2 must be our u. In other words, if $T(u_1) = T(u_2)$, then $u_1 = u_2$. This is the definition we use for injectivity.

T: U → V is INJECTIVE (is an INJECTION) if $T(u_1) = T(u_2)$ implies $u_1 = u_2$.

As in college algebra, when a map as injective (one-to-one) it doesn't destroy information. For example, the function $f(x) = x^2$ is not one-to-one because you get more than one input for a particular output. If you know the square of a number is 9, you don't know if the number is 3 or -3. The mapping destroyed information about whether the number was positive or negative. If our mapping T is a linear transformation from a finite dimensional vector space then being injective means that the range of T has the same dimension as the domain of T. The linearity of T gives us an easy way to check for injectivity. Let's try to prove the following theorem:

A linear transformation T: U → V is injective if T(u) = 0 implies u = 0.

What is the difference between this statement and our original definition? In our original definition we had to consider any possible output. Now we're only concerned with an output of zero. Why is that sufficient? Think about it and try writing your own proof before reading the following.

Proof: If T(u) = 0 implies u = 0 and if $T(u_1) = T(u_2)$, then $T(u_1) - T(u_2) = 0$. Linearity of T gives $T(u_1 - u_2) = 0$. Our given implication gives us $u_1 - u_2 = 0$, which implies $u_1 = u_2$. By definition, T is injective.

Let’s try another proof. This time we’ll talk through the ideas and I’ll provide a template for the proof.

Let S: V → W and T: U → V be linear transformations.

Prove that if S◦T is injective, then so is T.

First note that we can write our composition of transformations in the following way:

$$U \overset{T}{\rightarrow} V \overset{S}{\rightarrow} W$$

Using the idea that injective mappings do not destroy information, what does our theorem say? It says that if we go from U to W along these arrows that we don’t destroy information, so we couldn’t have destroyed information with the first arrow. This seems intuitively clear but, of course, we have to translate this into mathematical statements. Using the definition of injectivity, what do we need to show that T is injective? This gives us “two” elements of V, but we need to relate them to having “two” elements of W. How can we take elements of V and push them forward into elements of W? What does the injectivity of S◦T tell us? First try writing your proof without referring to the following.

Proof: Given $T(u_1) = T(u_2)$, then _____(u_1) = _____(u_2). By the injectivity of _______,

_____ = _____. Therefore, T is ___________________.

In this particular case, the proof was a lot shorter than the discussion.

SURJECTIVITY

Injectivity is about not losing information in a mapping. Surjectivity is about having every element of the image space actually being in the range of the map. The definition of surjectivity is exactly as it seems it should be.

A map T: U → V is SURJECTIVE (is a SURJECTION) if for each v in V, there exists u in U such that T(u) = v.

The image space of an injective linear transformation has to have dimension at least as big as the domain space and the rank of the transformation has to match the dimension of the domain space.

The image space of a surjective linear transformation has to have dimension at least as small as the domain space and the rank of the transformation has to match the dimension of the range space.

If a linear transformation is both injective and surjective, we say that it is BIJECTIVE. This means that the rank of the transformation matches the dimensions of the domain and range spaces and so each element of the range space is associated to exactly one element of the domain space and the transformation is invertible.

Let's try a proof involving surjectivity.

Prove: If linear transformations T: U → V and S: V → W are both surjective,
then so is S ∘ T.

Let's start by drawing the diagram: $\text{U} \xrightarrow{T} \text{V} \xrightarrow{S} \text{W}$.

What are we trying to show? ____

What does it mean that S is surjective? ______________________________________

What does it mean that T is surjective? ______________________________________

How can we put these together? ____

Now let's write our proof.

Proof: Let w be an element of W. S is surjective means that there exists an element v of V such that ____________________. T is surjective means that there exists an element u of U such that _____________________. (S ∘ T)(u) = __________. By definition, ______________________________________.

EIGENVALUES AND EIGENVECTORS

Eigen is German for own or inherent. So eigenvalues and eigenvectors are things inherent to a matrix as a linear transformation. As a linear transformation, a square matrix A sends a vector to another vector. If there is a direction that A stretches, then a vector in that direction is an eigenvector and the amount of stretch is the associated eigenvalue. In mathematical terms, λ is an eigenvalue associated to the eigenvector v if $Av = \lambda v$. This leads to an appropriate way to find the eigenvalues. If $Av = \lambda v$, then $Av - \lambda v = 0$, so $Av - \lambda Iv = 0$, and $(A - \lambda I)v = 0$. This means that v is in the nullspace of the matrix $(A - \lambda I)$, which means that $(A - \lambda I)$ is not invertible and so $\det(A - \lambda I) = 0$.

What kind of structure is attached to eigenvalues and eigenspaces?

Prove: If A is a matrix representing a linear transformation from V to W, and λ is an eigenvalue of A, then the eigenspace, V_λ, associated to λ is a subspace of V.

How do we show subsets of vector spaces to be subspaces? What two things about V_λ do we need to show?

Proof: Let $v_1, v_2 \in$ ______. A(___ + ___) = _____ + ______ = _____ + _____ = λ (______)

so $v_1 + v_2$ _____________. Let c be a scalar. A(______) = cA(_____) = c_____ = λ______ so cv_1 ___________.

Prove: If A is a matrix representing a linear transformation from V to W, and λ_1 and λ_2 are distinct eigenvalues of A with corresponding eigenspaces $V_{\lambda 1}$ and $V_{\lambda 2}$, then $V_{\lambda 1} \cap V_{\lambda 2} = 0$.

How do we show that the intersection of the eigenspaces is trivial? What should be true about an element in the intersection as an element of $V_{\lambda 1}$? What should be true about an element in the intersection as an element of $V_{\lambda 2}$?

Proof: Let v ∈ ______. Since v ∈ ____, Av = ____. Since v ∈ ____, Av = ____. This means Av = ____ = ____ and so $(\lambda_1 - \lambda_2)$___ = ____. $\lambda_1 - \lambda_2 \neq 0$, so v = ___. This implies $V_{\lambda 1} \cap V_{\lambda 2} = 0$.

SIMILAR MATRICES

Matrices A and B, each n x n, are called SIMILAR if there exists an invertible matrix P such that $B = P^{-1}AP$. Why is this a concept of interest? If we consider A and B as linear transformations and P as a change of basis, then A and B represent the same linear transformation, just as representations under different bases. The commutative diagram for this relationship is

$$\begin{tikzcd} V \arrow[r, "B"] \arrow[d, shift left, "P"] & V \arrow[d, shift left, "P"] \\ V \arrow[r, "A"] \arrow[u, shift left, "P^{-1}"] & V \arrow[u, shift left, "P^{-1}"'] \end{tikzcd}$$

Similarity is an equivalence relation, which means that we have the option to treat all of the matrices which are similar to a matrix A as a single algebraic object or we can exchange A for some other representative of this “object”. For example, if A is diagonalizable, then there is a diagonal matrix in this object to which A is similar. This diagonal matrix is generally much easier to work with than A. For example, if A is similar to the diagonal matrix D, then there exists P such that $A = PDP^{-1}$. So what is A^k? Multiplying matrices tends to be difficult and raising matrices to powers consequently more so. But how about D? What is PDP^{-1} PDP^{-1}? What can you then say about A^k?

It is also easy to see that similar matrices have the same characteristic polynomials and hence the same eigenvalues. Moreover, we can show the following.

Prove: If A and B are similar and λ is an eigenvalue of A and B, then the geometric multiplicity of λ is the same for both matrices.

How can we start? What is the geometric multiplicity? Is there a relationship between the eigenvectors of A and B? What is the relationship between A and B?

Proof: Let v be an eigenvector of A associated to λ and let P be a change of basis matrix such that $A = PBP^{-1}$. Av = ______, so $PBP^{-1}v$ = ______. Multiplying both sides on the left by ____ gives $BP^{-1}v$ = _____ = ______. By definition, ______ is an eigenvector of B associated to λ. For any set of independent eigenvectors, $v_1, \ldots, v_k$, of A, associated to λ, this gives eigenvectors ___, …, ____, of B, associated to λ. ____ is invertible, so these vectors are also independent.

This means that the geometric multiplicity of λ for B is greater than or equal to the geometric multiplicity of λ for A. The same argument works for starting with eigenvectors of B, so the geometric multiplicity of λ for A is greater than or equal to the geometric multiplicity of λ for B. Together, these give our result.

ABSTRACT VECTOR SPACES

We want to extend our ideas about vector spaces beyond Euclidean space. To that end, we define abstract vector spaces. Ultimately, we want a set of vectors and a set of scalars so that linear combinations are well behaved. The scalars need to be a field. For the beginning linear algebra course, that field is generally the real numbers or the complex numbers. First, let's consider vector addition. We need that set with that operation to be an abelian group. That means we need the following. Given vectors **u**, **v**, **w** in V,

1.	**u** + **v** is in V	Closure under addition
2.	**u** + **v** = **v** + **u** `	Commutativity
3.	(**u** + **v**) + **w** = **u** + (**v** + **w**)	Associativity
4.	There exists **0**, such that **0** + **u** = **u**	Identity
5.	There exists **–u**, such that **–u** + **u** = **0**	Inverse

We also need associativity of the scalars and distribution. Let c and d be scalars.

6.	c**u** is in V	Closure under scalar multiplication
7.	c(d**u**) = (cd)**u**	Scalar associativity
8.	c(**u** + **v**) = c**u** + c**v**	Vector distribution
9.	(c + d)**u** = c**u** + d**u**	Scalar distribution

Finally, we need to make sure that scalar multiplication works the way it is supposed to.

10. 1**u** = **u**

Your textbook should have plenty of examples of abstract vector spaces. Function spaces are common vector spaces. One of the more important ones is the **dual space**.

Let V be a finite dimensional real vector space. The dual space, V^*, is the set of all linear maps of V into **R**, the set of real numbers.

Prove: V^* is a vector space with the same dimension as V.

Proving that V is a vector space is simply a matter of checking that the above rules hold. To find the dimension of V^* we can find a basis related to a basis for V. This dual basis should depend on the choice of basis for V. What is the simplest linear map which isolates the first basis element? The second, etc.? Use these. You need to show that these linear maps forma basis for the space.

THE ADJOINT

The word “adjoint” is used in two different ways in linear algebra. The classical adjoint, also called the “adjugate”, is the transpose of the cofactor matrix. The adjugate can be used to compute the inverse of a square matrix. The other adjoint, however, is about the behavior of matrices in inner products. If we refer to the adjoint of A as A*, then for an inner product $<,>$, A* is defined to be the matrix so that for all vectors **u** and **v** for which the inner product is defined, $<A\mathbf{u},\mathbf{v}> = <\mathbf{u},A^*\mathbf{v}>$. In other words, we get the same product whether we act on the left vector by A or on the right vector by its adjoint. If the inner product is the usual dot, what is the relationship between A and A*?

Prove: If $<,>$ is the usual dot product over a finite dimensional real vector space, then $A^* = A^T$.

The idea here is pretty straight forward. Translate the dot product into a matrix product, reassociate the matrix with the second vector, and translate back.

An isometry is a map which preserves lengths. In the context of linear algebra, A is an isometry under the inner product $<,>$ if $<A\mathbf{u},A\mathbf{v}> = <\mathbf{u},\mathbf{v}>$.

For the next proof, we need to know the following:

Fact: If $<\mathbf{u}, M\mathbf{v}> = 0$ for all **u**, **v**, then $M = 0$.

Proof: (by contradiction). Assume $M \neq 0$. Then $\text{rank}(M) \neq 0$ implies that there is a vector **v** such that $M\mathbf{v} \neq 0$. Let $\mathbf{u} = M\mathbf{v}$. $<\mathbf{u},M\mathbf{v}> = <M\mathbf{v},M\mathbf{v}> = 0$. But the only vector for which $<\mathbf{u},\mathbf{u}> = 0$ is the zero vector, which contradicts our statement that $M\mathbf{v} \neq 0$. Therefore, $M = 0$.

Prove: If $<,>$ is the usual dot product over a finite dimensional real vector space, then A is an isometry if and only if A is an orthogonal matrix.

Again, the idea is straight forward. Use the adjoint to move the A from the first vector in the inner product to the second. What does it say? How can we take this statement and translate it to the fact we proved above?

ADDITIONAL

EXERCISES

EXERCISE 1

Prove that if A is an orthogonally diagonalizable invertible matrix, then A^{-1} is orthogonally diagonalizable

The first thing that we should notice about this problem is that we have a lot of information about A. As usual, we want to make sure that we know what the words mean.

What does it mean for A to be invertible?

What does it mean for A to be diagonalizable?

How is orthogonally diagonalizable different from just diagonalizable?

This is one of those problems where if we just do what the words tell us to do, everything falls into place.

PROOF

A is orthogonally diagonalizable implies that there exists matrices Q and D such that

_______________ = A where Q is invertible and Q^{-1} = ______ and where D is ________________

and invertible with diagonal entries $d_{ii.}$ D is invertible because ____________________________

____ which implies $d_{ii} \neq$ ______ for all i. D^{-1} is also

___________________ with diagonal entries _________. As above, _______________ = A with all

matrices being invertible. So A^{-1} = ____________ = ____________. By definition, A^{-1} is

orthogonally diagonalizable.

EXERCISE 2

Prove that if A is nilpotent and diagonalizable, then A must be the zero matrix.

As usual, let's start by thinking about what the statement says. What does it mean for a matrix to be nilpotent? What does it mean for a matrix to be diagonalizable? We're interested in both the ideas and the definitions. Diagonalizable means that under an appropriate change of basis the matrix becomes diagonal, i.e. all the entries not on the diagonal are zero. What are the entries on the diagonal? Do they have any significance? We know that under a change of basis the characteristic polynomial stays the same. But for a diagonal matrix the characteristic polynomial is very easy. What is it? This means that the diagonal entries are the eigenvalues. Remember that the eigenvalues tell how much the matrix, as a linear transformation, stretches special directions, the eigenvectors. But what does it mean for a matrix to be nilpotent? As a linear transformation, if we apply it enough times it sends everything to zero. So it eventually kills off all vectors, including the eigenvectors. But each time we apply the matrix we just stretch an eigenvector by its eigenvalue. What does this tell us about the eigenvalue?

After reasoning through the above, you may ask yourself why we need A to be diagonalizable. As long as we have a basis of eigenvectors then that should be enough. Is there a relationship between having a basis of eigenvectors and being diagonalizable?

This should cover the ideas behind the proof. Now we need the definitions.

How do we translate "A is nilpotent" into mathematics? _________________________________

How do we translate "A is diagonalizable" into mathematics? ____________________________

PROOF

A is nilpotent implies there exists n such that ______________________________.

A is diagonalizable implies there exists an invertible matrix, P, and a diagonal matrix, D, such that ______________________________. Substituting gives us (______)^n = 0, which implies _________ = 0. Multiplying both sides of this equation on the left by _______and on the right by _________ gives us the equation _______ = 0. Let d be any diagonal entry of D. Then d^n = ____, which implies d = ____. This in turn implies that D is the ___________ matrix.

Substituting this into the equation given by the definition of the A being diagonalizable, gives us

A = ___________ = ____, which was to be proven.

IDEMPOTENT MATRICES

The word, idempotent, comes from roots meaning same power. So idempotent matrices are those which when raised to powers stay the same. In other words, a square matrix A is idempotent if $A^2 = A$.

Let's try a short proof involving this idea:

Prove that the only invertible idempotent n x n matrix is the identity matrix.

What does it mean for A to be invertible? This leads to what equation? Since $A^2 = A$, whenever we see A we can write A^2 instead. This is substitution. Put this together and write a proof. Once you have finished a proof compare it to the following:

A invertible means there exists A^{-1} such that $AA^{-1} = I$. A idempotent means $A^2 = A$. Substituting A^2 for A gives us $A^2A^{-1} = I$, which implies $A(AA^{-1}) = I$, giving $AI = I$, and finally, $A = I$.

INNER PRODUCT SPACES

An inner product is a generalization of the dot product. We want to add a relationship between the vectors in a vector space which gives us some geometry. What did the dot product give us? Lengths and angles. Recall that $\mathbf{u}\cdot\mathbf{u}$ was the square of the length of $\mathbf{u}$, so we needed that $\mathbf{u}\cdot\mathbf{u}$ be a nonnegative real number which is only 0 when $\mathbf{u} = \mathbf{0}$. We also needed the dot product to act like a product in relation to addition, so that the dot product is bilinear. And finally, the angles and lengths didn't change when we reversed the order of the vectors, so $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$.

In summary then, we define an inner product on V as a map $<\cdot,\cdot> : V \times V \rightarrow \mathbf{R}$ such that

1) $<\mathbf{v},\mathbf{v}> \geq 0$ with equality if and only if $\mathbf{v} = \mathbf{0}$
2) $<\mathbf{v},\mathbf{w}> = <\mathbf{w},\mathbf{v}>$
3) $<\mathbf{v}_1 + \mathbf{v}_2,\mathbf{w}> = <\mathbf{v}_1,\mathbf{w}> + <\mathbf{v}_2,\mathbf{w}>$
4) $<c\mathbf{v},\mathbf{w}> = c<\mathbf{v},\mathbf{w}>$

Let's check a couple of spaces with inner products.

Prove: If $f(x), g(x)$ are functions (i.e., vectors) in the space of real valued continuous functions on the interval [0,1], C[0,1], show that $< f, g >= \int_0^1 fg \ dx$ is an inner product on C[0,1].

Prove: If V and W are real vector spaces where W has an inner product $<\cdot,\cdot>$ and if T: V → W is a linear transformation, show that $< \mathbf{v}_1,\mathbf{v}_2>' = <T(\mathbf{v}_1), T(\mathbf{v}_2)>$ defines an inner product on V if and only if T is one-to-one.