Writing proofs

Tim Hsu, San José State University

January 11, 2003

Contents

Introduction

The goal of this handout is to help you learn to write proofs. The handout assumes that you have either taken or are currently taking linear algebra. The only reason for this assumption is that to talk about proofs, we need something to prove, and linear algebra is something that many people in your situation are familiar with.

There are four parts to this handout. Part I describes what a proof is and what it does; Part II describes the fundamental structure a proof, featuring the *if-then method* for writing proofs; Part III describes how to apply the if-then method; and Part IV describes a few special cases where the if-then method doesn't apply directly.

Part I

Fundamentals

1 Definitions and theorems

The theoretical structure of mathematics can be broken down into *definitions* and *theorems*, and the first step in understanding proofs is to understand the difference between them. The idea is that definitions describe the objects we choose to study, and theorems are logical consequences that we subsequently deduce about those objects.

Much of the power of theoretical mathematics lies in the fact that, if we choose our definitions well, then:

1. The definitions will be natural and simple enough that no reasonable person will disagree with them.

2. Nevertheless, we can deduce interesting theorems about them.

The result is to obtain mathematical conclusions that are based on only a small set of reasonable assumptions, but nevertheless have a wide variety of applications.

Now, if you don't have much experience thinking about definition-theorem mathematics, one natural tendency is to lump definitions and theorems together as a list of facts that are all "true." However, to understand what's really going on in a math class where theorems and proofs play an important role, it's important that you understand which facts are true by definition (i.e., because we said so), and which facts are true by theorem (i.e., because we deduced them by logic).

For example, in linear algebra, it's true by definition that:

Definition 1. Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ be vectors in $\mathbf{R}^3$. If every vector of $\mathbf{R}^3$ is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, then the vectors $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ span $\mathbf{R}^3$.

That's exactly the definition of span. On the other hand, it's a theorem (i.e., something that follows logically from the definitions, without making additional assumptions) of linear algebra that:

Theorem 2. *Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ be vectors in $\mathbf{R}^3$, and let A be the 3×3 matrix whose columns are $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$. If the rank of A is 3, then the vectors $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ span $\mathbf{R}^3$.*

In the rest of this handout, we'll discuss the process by which theorems are deduced from definitions: namely, the proceess of *proof*.

2 What is a proof?

A *proof* is just a logical explanation of a theorem. For example, consider:

Theorem 3. *If an integer is (exactly) divisible by 6, it must be even.*

Suppose you understand why Theorem 3 is true. How would you write down a proof of it? Well, one good place to start is to imagine that you're talking to a friend of yours, and just write down how you would explain it to them. (Even better: Don't imagine it, actually do it.)

For example, you might say:

Proof. If a number is divisible by 6, that means 6 goes into it without a remainder. Since 2 goes into 6 without a remainder, that means that 2 goes into the original number without a remainder. □

Or:

Proof. If a number is divisible by 6, it has to be equal to 6 times some other number. But since 6 is even, if you take any number and multiply it by 6, you get an even number, so the original number must be even. □

Note that one confusing aspect of the second proof is that the phrase "the number" refers to several different numbers. Reading such a proof can be like reading a story where none of the characters have names. This is the reason algebraic symbols were invented: to let you name quantities that you will use repeatedly.

For example, rewriting the second explanation using symbols, we get:

Proof. If n is divisible by 6, then there's some number d such that $n = 6d$. But since 6 is even, $6d$ is even, so n is even. □

Clearer, and shorter as well. (We'll see another way to get this proof in Section 4.) In any case, the main point is, a proof is just an explanation, so when you are asked to prove something, you're just supposed to explain why it's true.

Part II

The structure of proofs

3 Assumptions and conclusions

To understand the structure of proofs, let's first look at the structure of theorems. Broadly speaking, a mathematical theorem states that certain assumptions lead to certain conclusions. For example, in Theorem 3 from Section 2, the assumption is that we have an integer n divisible by 6, and the conclusion is that n is also even.

It is important to keep in mind that the conclusion of a theorem always depends on certain assumptions; if nothing is assumed, nothing can be concluded. One practical consequence of this fact is that when you're trying to prove a theorem, and it seems like you have no place to start, you can always start with the assumptions, since the conclusions of a theorem must rely on the assumptions of the theorem.

Now, in many circumstances, it may be enough just to think about proof as explaining how the conclusions of a theorem follow from the assumptions of a theorem. On the other hand, it is sometimes helpful to have a more formal way of looking at this process, especially when working with more abstract material. Therefore, in the rest of Part II, we'll discuss two fundamental technical aspects of proofs: the *if-then method* (Section 4) and working with *sets* (Section 5).

4 The if-then method

To describe the assumptions/conclusions structure from Section 3 in a more formal way, we can use the idea of an *if-then* statement: "If (we assume that) A is true, then C follows." For example, let's consider Theorem 3 again, slightly restated:

Theorem 4. *If an integer n is divisible by 6, then n must be even.*

If you don't have much experience with proofs, you may find it useful at first to separate such a statement into *background assumptions*, the *if* assumptions, and the *then* conclusions. Specifically:

- **Background assumptions:** n is an integer.
- **If:** n is divisible by 6;
- **Then:** n is even.

Note that the background assumptions are as important as the "if" assumptions. In fact, the theorem could easily be restated to include the background assumptions as part of the "if":

> *If n is an integer, and n is divisible by 6, then n must be even.*

Therefore, in the sequel, we will ignore the distinction between background assumptions and "if" assumptions, and just lump them together as assumptions.

In any case, once you have a theorem divided into assumptions and conclusion, you can prove it using the following method.

The if-then method

1. Carefully write out all assumptions (the "if" part) at the beginning of the proof. Usually this involves expanding what's written in the assumptions using the definitions of the terms that appear there.
2. Write out the conclusion of the theorem (the "then") at the end of the proof, and expand it using definitions as well. This is what we want to show follows from our assumptions.
3. The point of the proof is now to show that given the assumptions, logical deduction leads to the conclusion. One of the best ways to do this is to work forward logically from the assumptions (think: what follows from the "if"?) and backwards from the conclusion (think: what would imply the "then"?) until you meet in the middle.

To paraphrase a well-known cartoon character, that last step can be a doozy. However, especially in a class situation, doing the first two steps can really make the interesting part (step 3) easier.

For example, applying the if-then method to Theorem 4:

1. The assumptions of the theorem are: "n is an integer divisible by 6." By the definition of divisibility of integers, this means that $n = 6d$ for some integer d.
2. The conclusion of the theorem says: "n is even." By the definition of even, that is the same as saying that n is divisible by 2. By the definition of divisibility, this means that we want to show that $n = 2r$ for some integer r.
3. So now, we want to **assume** that $n = 6d$ for some integer d, and then somehow **deduce** that $n = 2r$ for some integer r. However, if we know that $n = 6d$, then after a while, we might see that $n = 2(3d)$, which means that $n = 2r$ holds for $r = 3d$.

We therefore obtain the following proof:

Proof. Assume that $n = 6d$. Therefore, $n = 2(3d)$. So, if we let $r = 3d$, we see that $n = 2r$ for some integer r, which means that n is even. □

If you find this approach to be too mechanical, you don't need to follow it strictly. As long as you can logically explain how the conclusions of the theorem follow from the assumptions, you'll have a valid proof. The point is, if you don't immediately see how to

get from assumptions to conclusions, the if-then method gives you an initial direction in which to proceed.

For a more complicated example, consider the following theorem.

Theorem 5. *Let* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *be vectors in* $\mathbf{R}^n$*, and let* $\mathbf{v}$ *and* $\mathbf{w}$ *be linear combinations of* $\mathbf{u}_1$ *and* $\mathbf{u}_2$*. If* $\mathbf{x}$ *is a linear combination of* $\mathbf{v}$ *and* $\mathbf{w}$*, then* $\mathbf{x}$ *is also a linear combination of* $\mathbf{u}_1$ *and* $\mathbf{u}_2$*.*

Again separating the parts of this statement into assumptions and conclusions, we get:

- **Assumptions:** $\mathbf{u}_1$ and $\mathbf{u}_2$ are vectors in $\mathbf{R}^n$, $\mathbf{v}$ and $\mathbf{w}$ are linear combinations of $\mathbf{u}_1$ and $\mathbf{u}_2$, and $\mathbf{x}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$.
- **Conclusion:** $\mathbf{x}$ is also a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$.

Next, let's rewrite everything using the definition of linear combination: for example, a linear combination of $\mathbf{v}$ and $\mathbf{w}$ is precisely some vector of the form $c_1\mathbf{v} + c_2\mathbf{w}$ for some numbers c_1, c_2. (You have to know your definitions if you want to do proofs!) We then get:

- **Assumptions:** $\mathbf{u}_1, \mathbf{u}_2 \in \mathbf{R}^n$, $\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ for some numbers a_1, a_2, $\mathbf{w} = b_1\mathbf{u}_1 + b_2\mathbf{u}_2$ for some numbers b_1, b_2, and $\mathbf{x} = c_1\mathbf{v} + c_2\mathbf{w}$ for some numbers c_1, c_2.
- **Conclusion:** $\mathbf{x} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2$ for some numbers d_1, d_2.

Applying if-then, we first write:

Beginning and end of proof, no middle yet. Assume that $\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ for some numbers a_1, a_2, $\mathbf{w} = b_1\mathbf{u}_1 + b_2\mathbf{u}_2$ for some numbers b_1, b_2, and $\mathbf{x} = c_1\mathbf{v} + c_2\mathbf{w}$.

$$\vdots$$
(the middle part to be filled in)
$$\vdots$$

Therefore, $\mathbf{x} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2$ for some numbers d_1, d_2. □

After writing that out, you might eventually think of filling in the middle by substituting for $\mathbf{v}$ and $\mathbf{w}$ in $\mathbf{x} = c_1\mathbf{v} + c_2\mathbf{w}$. (There's not really much else you can do.) This gives the following proof:

Proof. We know that $\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ for some numbers a_1, a_2, and $\mathbf{w} = b_1\mathbf{u}_1 + b_2\mathbf{u}_2$ for some numbers b_1, b_2. Assume that $\mathbf{x} = c_1\mathbf{v} + c_2\mathbf{w}$. Substituting for $\mathbf{v}$ and $\mathbf{w}$, we see that:

$$\begin{aligned}
\mathbf{x} &= c_1\mathbf{v} + c_2\mathbf{w} \\
&= c_1(a_1\mathbf{u}_1 + a_2\mathbf{u}_2) + c_2(b_1\mathbf{u}_1 + b_2\mathbf{u}_2) \\
&= c_1a_1\mathbf{u}_1 + c_1a_2\mathbf{u}_2 + c_2b_1\mathbf{u}_1 + c_2b_2\mathbf{u}_2 \\
&= (c_1a_1 + c_2b_1)\mathbf{u}_1 + (c_1a_2 + c_2b_2)\mathbf{u}_2.
\end{aligned}$$

Therefore, $\mathbf{x} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2$ for $d_1 = (c_1a_1 + c_2b_1)$ and $d_2 = (c_1a_2 + c_2b_2)$. The theorem follows. □

(Using "The theorem follows" here is a slightly awkward but effective way to let the reader know that the proof is done.)

We hope you agree that applying the if-then method here really gives you a big hint on how to finish the proof.

5 Sets, elements, and the if-then method

A *set* S is a bunch of objects, and those objects are called the *elements* of S. A finite set can be described by listing its elements inside { }. For example, the elements of the set $S = \{2, 3, 5, 7, 11\}$ are the numbers 2, 3, 5, 7, and 11. We also write $2 \in S$, $3 \in S$, and so on, to mean that 2 is an element of S, 3 is an element of S, and so on.

Often, it is convenient to describe a set S not by listing the elements of S, but by giving a precise condition for being an element of S. In notation, this looks something like

$$S = \{x \mid (\text{defining condition on } x)\},$$

which says: "S is the set of all x such that x satisfies the condition (defining condition)."

For example, for vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n, the span of $\{\mathbf{u}, \mathbf{v}\}$ is defined to be:

$$\text{Span}\{\mathbf{u}, \mathbf{v}\} = \{\mathbf{x} \in \mathbf{R}^n \mid \mathbf{x} = a\mathbf{u} + b\mathbf{v} \text{ for some } a, b \in \mathbf{R}\}.$$

In words: The span of $\{\mathbf{u}, \mathbf{v}\}$ is the set of all elements $\mathbf{x}$ of $\mathbf{R}^n$ such that $\mathbf{x} = a\mathbf{u} + b\mathbf{v}$ for some real numbers a and b.

The following principle describes how to work with a set given by a defining condition.

> **The Defining Condition Principle:** If a set S is given by a defining condition, then saying that x is an element of S is the same thing as saying that x satisfies the defining condition of S.

For example, from the definition of span given above, we see that the statement "$\mathbf{x} \in \text{Span}\{\mathbf{v}, \mathbf{w}\}$" is the same thing as saying that "$\mathbf{x} = a\mathbf{v} + b\mathbf{w}$ for some real numbers a and b."

Because many sets are described by defining conditions, we often need the defining condition principle to turn an if-then statement into something we can use in a proof. For example, consider the following theorem:

Theorem 6. *Let* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *be vectors in* $\mathbf{R}^n$. *If* $\mathbf{v} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, $\mathbf{w} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, *and* $\mathbf{x} \in \text{Span}\{\mathbf{v}, \mathbf{w}\}$, *then* $\mathbf{x} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

Applying if-then and defining condition, we have:

- **Assumptions:** Our first assumption is that $\mathbf{u}_1, \mathbf{u}_2 \in \mathbf{R}^n$. Next, the first statement in the "if" is $\mathbf{v} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Since the defining condition of $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ is "$= a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ for some numbers a_1, a_2", the statement "$\mathbf{v} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$" is equivalent to "$\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ for some numbers a_1, a_2". (Important: Note that we

have changed a, b to a_1, a_2. We can do that because the a and b in the definition of $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ are just dummy variables representing arbitrary numbers.)

By exactly the same reasoning, the other parts of the "if" become "$\mathbf{w} = b_1\mathbf{u}_1 + b_2\mathbf{u}_2$ for some numbers b_1, b_2" and "$\mathbf{x} = c_1\mathbf{v} + c_2\mathbf{w}$ for some numbers c_1, c_2".

- **Conclusion:** Similarly, we want to show that $\mathbf{x} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2$ for some numbers d_1, d_2.

We then proceed as before. (In fact, can you now see that Theorem 6 is exactly the same as Theorem 5?)

Finally, we note that the defining condition principle often applies when a theorem refers to "an element", an "arbitrary element", and so on. The way to handle this situation is to give that arbitrary element a name, and then proceed as before. For example, consider yet another version of Theorems 5 and 6:

Theorem 7. *Let* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *be vectors in* $\mathbf{R}^n$*, and suppose that* $\mathbf{v} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ *and* $\mathbf{w} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$*. Any vector in* $\text{Span}\{\mathbf{v}, \mathbf{w}\}$ *is also in* $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

This version also hides the "if-then" part of the theorem, so it's probably a good idea to figure that out first. Specifically, the last sentence of Theorem 7 can be written as:

> If you take a vector in $\text{Span}\{\mathbf{v}, \mathbf{w}\}$, then that vector is also in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

Once the statement is in "if-then" form, it's easier to see that the theorem is about one particular vector, which we assume is an element of $\text{Span}\{\mathbf{v}, \mathbf{w}\}$, and then deduce is an element of $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. We may as well call that vector $\mathbf{x}$, which changes the last sentence of Theorem 7 to:

> If $\mathbf{x} \in \text{Span}\{\mathbf{v}, \mathbf{w}\}$, then $\mathbf{x} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

In other words, Theorem 7 is equivalent to Theorem 6.

Part III

Applying if-then

6 Converting theorems to if-then statements

Theorems are often stated in a way that doesn't immediately make it clear how they can be expressed as if-then statements. Here are a few common examples of how such statements can be converted to if-then statements.

Proving "for every" statements. One variation on if-then statements is the "for every" statement, i.e., a statement like:

Theorem 8. *For every integer $n > 1$, there is a prime number p such that p divides n.*

To prove a "for every" statement, you turn it into an if-then statement as follows:

Theorem 9. *If n is an integer and $n > 1$, then there is a prime number p such that p divides n.*

Proving "if and only if" statements. An if and only if statement is one like the following.

Theorem 10. *Let $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{x}$ be vectors in $\mathbf{R}^n$. The vector $\mathbf{x}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$ if and only if $\mathbf{x}$ is a linear combination of $\mathbf{v}+\mathbf{w}$ and $\mathbf{v}-\mathbf{w}$.*

This statement is precisely the sum of two if-then statements:

1. If $\mathbf{x}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$, then $\mathbf{x}$ is a linear combination of $\mathbf{v}+\mathbf{w}$ and $\mathbf{v}-\mathbf{w}$.

2. If $\mathbf{x}$ is a linear combination of $\mathbf{v}+\mathbf{w}$ and $\mathbf{v}-\mathbf{w}$, then $\mathbf{x}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$.

So, to prove the combined "if and only if" statment, you prove both if-then statements separately.

The Following Are Equivalent. More generally, we might have a statement like the following.

Theorem 11. *Let $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{x}$ be vectors in $\mathbf{R}^n$. The following are equivalent:*

1. *The vector $\mathbf{x}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$.*

2. *The vector $\mathbf{x}$ is a linear combination of $\mathbf{v}+\mathbf{w}$ and $\mathbf{v}-\mathbf{w}$.*

3. *The vector $\mathbf{x}$ is a linear combination of $\mathbf{v}$ and $\mathbf{v}+2\mathbf{w}$.*

In other words, any one of these three statements implies the other two.

Often the most efficient way to prove such a TFAE statement is to prove: If (1), then (2); if (2), then (3); and if (3), then (1). A similar "circle of implications" can be used to prove a TFAE statement with 4 or more parts.

Proving "or" statements. Occasionally, it's necessary to prove an "if-then" statement where the "then" part of the statement has an "or" in it. One famous example is:

Theorem 12. *Let a and b be integers. If ab is even, then either a is even or b is even.*

This is actually not an easy theorem to prove, so for our purposes here, we'll just try to understand how the proof starts. Specifically, how do you reach the conclusion that "Either a is even or b is even"?

Well, one standard method is to realize that the statement "Either a is even or b is even" is equivalent to the statement

> If a is not even, then b must be even.

(Think of it this way: If you want to be assured that either A or B must be true, then you only have to worry about what happens when A is false, in which case you just have to make sure that B is true.)

After you realize this equivalence, the main statement of Theorem 12 becomes:

> If ab is even, then if a is not even, then b is even.

This kind of "nested if" statement may look a little alarming at first, but makes more sense once the two "if" parts are combined (as they can be, logically):

> If ab is even and a is not even, then b is even.

So Theorem 12 can be broken down in our usual format as follows.

- **Assumptions:** a and b are integers. ab is even and a is not even.
- **Conclusion:** b is even.

The beginning and the end of the proof of Theorem 12 therefore look like:

Proof. Assume that a and b are integers, that ab is even, and a is odd.

$$\vdots$$
(substantive part)
$$\vdots$$

Therfore, b is even. □

7 For every ... there exists

One common variation on if-then statements that needs special consideration is the "for every ... there exists" statement. A typical example of such a statement is:

Theorem 13. *For every vector $\mathbf{v}$ in $\mathbf{R}^n$, there exists a vector $\mathbf{w} \in \mathbf{R}^n$ such that $\mathbf{v}+\mathbf{w} = \mathbf{0}$.*

Following Section 6, this translates into:

> If $\mathbf{v}$ is a vector in $\mathbf{R}^n$, then there exists a vector $\mathbf{w} \in \mathbf{R}^n$ such that $\mathbf{v}+\mathbf{w} = \mathbf{0}$.

The "then" part of this statement is typical of a "for every ... there exists" proof, in that to complete the proof, we have to *make something up* (the vector $\mathbf{w}$) that satisifies a given condition ($\mathbf{v} + \mathbf{w} = \mathbf{0}$).

Now, making things up can be quite difficult; in fact, you could say that making things up is a big part of what makes theoretical mathematics a creative subject. Nevertheless, if you understand the format of a "for every ... there exists" proof, you'll at least have a framework in which you can apply your creativity.

Returning to our example, one standard method of proving our statement follows the following format:

Proof. Assume that $\mathbf{v}$ is a vector in $\mathbf{R}^n$.

Let $\mathbf{w} =$ (this part to be filled in).

$$\vdots$$
$$\text{(check that } \mathbf{w} \text{ satisfies the condition } \mathbf{v} + \mathbf{w} = \mathbf{0})$$
$$\vdots$$

Therefore, $\mathbf{v} + \mathbf{w} = \mathbf{0}$, which means that $\mathbf{w}$ satisfies the condition that we wanted it to satisfy. □

To finish this proof, you next need to figure out what $\mathbf{w}$ should be. There are many ways to do this, but the basic is idea is trial and error: Take a guess as to what $\mathbf{w}$ is, try it and see if it works, and if not, try another value. In our example, after a while, you might guess that $\mathbf{w} = (-1)\mathbf{v}$ works, and it does. We may therefore complete our proof to:

Proof. Assume that $\mathbf{v}$ is a vector in $\mathbf{R}^n$.

Let $\mathbf{w} = -\mathbf{v}$. We then see that

$$\mathbf{v} + \mathbf{w} = \mathbf{v} + (-1)\mathbf{v} = (1 - 1)\mathbf{v} = 0\mathbf{v} = \mathbf{0}.$$

Therefore, $\mathbf{v} + \mathbf{w} = \mathbf{0}$, which means that $\mathbf{w}$ satisfies the condition that we wanted it to satisfy. □

Note that something that you might expect, namely, an explanation of how you came up with $\mathbf{w}$, need not be a part of the proof. This is mostly because such an explanation is not logically necessary; as long as there is some $\mathbf{w}$ that satisfies $\mathbf{v} + \mathbf{w} = \mathbf{0}$, you've shown that the "then" part of the theorem holds, given the "if", so there's not necessarily any reason to explain that you solved for $\mathbf{w}$, or found it by a process of trial-and-error, and so on. A secondary reason is that it's sometimes the case that even the author doesn't completely understand how he or she came up with the "there exists" object, other than by inspired guessing. In any case, for the purposes of proof, it doesn't matter how you come up with the "there exists" object; all that matters is that it works as you claim it does.

8 Containment and equality of sets

To say that a set A is *contained* in another set B, or alternately, that A is a *subset* of B (written $A \subseteq B$), means that every element of A is an element of B. In other words, $A \subseteq B$ means that if x is an element of A, then x is also an element of B. This last formulation of $A \subseteq B$ is often useful in proofs, as this turns $A \subseteq B$ into an if-then statement.

For example, consider the following statement.

Theorem 14. *For* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^n$, $\text{Span}\{\mathbf{u}, \mathbf{v}\} \subseteq \text{Span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$.

To prove this statement, we first turn it into an if-then statement:

Theorem 15. *For* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^n$, *if* $\mathbf{x} \in \text{Span}\{\mathbf{u}, \mathbf{v}\}$, *then* $\mathbf{x} \in \text{Span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$.

Better yet, since the span of a set of vectors is given by a defining condition (see Section 5), we can change this statement into:

Theorem 16. *For* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^n$, *if* $\mathbf{x} = a\mathbf{u} + b\mathbf{v}$ *for some* $a, b \in \mathbf{R}$, *then* $\mathbf{x} = c\mathbf{u} + d\mathbf{v} + e\mathbf{w}$ *for some* $c, d, e \in \mathbf{R}$.

Finally, to prove two sets A and B are equal (identical), we show that $A \subseteq B$ and $B \subseteq A$. Compare the proof of an "if and only if" statement in Section 6.

9 Closure of a set under an operation

First, a binary operation on a set X is an operation (say $*$) that has a value $x * y$ defined for all pairs of elements $x, y \in X$. For example, $+$ and $\times$ are binary operations on the real numbers.

Definition 17. Suppose X is a set and $*$ is a binary operation defined on X. We say that X is *closed under the operation* $*$ if, for all $x, y \in X$, $x * y \in X$.

For example, the integers are closed under addition: if x and y are integers, $x + y$ is also an integer. Similarly, the integers are closed under multiplication: if x and y are integers, xy is an integer. On the other hand, the set of all positive integers is *not* closed under subtraction, since 1 and 2 are positive integers, but $1 - 2 = -1$ is not.

To prove that a given set X is closed under an operation $*$, as usual, we convert the definition of closure into an if-then statement. As mentioned above, X is closed if:

> For all $x, y \in X$, $x * y \in X$.

As an if-then statement, this becomes:

> If x and y are elements of X, then $x * y$ is an element of X.

Therefore, to show that X is closed under the operation $*$, we use the following if-then format:

1. **Assumptions:** x and y are elements of X.

2. **Conclusion:** $x * y$ is also an element of X.

For example, define

$$X = \left\{(x, y) \in \mathbf{R}^2 \mid x = 2y\right\}.$$

Using the above ideas, we'll outline the proof of the following theorem.

Theorem 18. *X is closed under vector addition.*

First, using the definition of closure, we restate Theorem 18 as an if-then statement:

> If $\mathbf{v}$ and $\mathbf{w}$ are elements of X, then $\mathbf{v} + \mathbf{w}$ is an element of X.

Following Section 5, we rewrite the property of being an element of X using the defining condition for X:

> If $\mathbf{v} = (x_1, y_1)$ is an element of $\mathbf{R}^2$ such that $x_1 = 2y_1$, and $\mathbf{w} = (x_2, y_2)$ is an element of $\mathbf{R}^2$ such that $x_2 = 2y_2$, then $\mathbf{v} + \mathbf{w} = (x_3, y_3)$ is an element of $\mathbf{R}^2$ such that $x_3 = 2y_3$.

Note that it's convenient to make up names for the coordinates of $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{v} + \mathbf{w}$, so we can use the defining condition for X.

So now, applying the if-then method, we get the following outline:

Proof. Assume that $\mathbf{v} = (x_1, y_1)$ is an element of $\mathbf{R}^2$ such that $x_1 = 2y_1$, and $\mathbf{w} = (x_2, y_2)$ is an element of $\mathbf{R}^2$ such that $x_2 = 2y_2$.

Let (x_3, y_3) be the coordinates of $\mathbf{v} + \mathbf{w}$.

$$\vdots$$
$$\text{(stuff to be filled in)}$$
$$\vdots$$

Therefore, $\mathbf{v} + \mathbf{w} = (x_3, y_3)$ is an element of $\mathbf{R}^2$ such that $x_3 = 2y_3$. □

In fact, once you understand the basic structure of the proof, you can see that filling in the middle just requires computing enough about x_3 and y_3 to see that $x_3 = 2y_3$. (Try it!)

10 Uniqueness

Occasionally we want to prove that an object with certain properties is unique, that is, that there is at most one such object. The standard technique for proving that an object with certain properties is unique is to use the following if-then format:

1. **Assumption:** There are two objects $\mathbf{x}$ and $\mathbf{y}$ with the properties in question.

2. **Conclusion:** What appears to be two objects must actually be just one object; that is, $\mathbf{x} = \mathbf{y}$.

This is not the most obvious approach, certainly, but I hope you agree that if any two objects with certain properties must be equal, then there's at most one object with those properties.

For example, suppose we want to show that the solution to the equation $A\mathbf{x} = \mathbf{b}$ is unique. The above format becomes:

1. **Assumption:** $\mathbf{u}$ and $\mathbf{v}$ are solutions to $A\mathbf{x} = \mathbf{b}$; in other words, $A\mathbf{u} = \mathbf{b}$ and $A\mathbf{v} = \mathbf{b}$.

2. **Conclusion:** $\mathbf{u} = \mathbf{v}$.

In other words, we want to show that: "If $\mathbf{u}$ and $\mathbf{v}$ are solutions to $A\mathbf{x} = \mathbf{b}$, then $\mathbf{u} = \mathbf{v}$."

For a slightly different example, suppose we want to show that vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are linearly independent. By definition, we must show that:

> The only solution to the equation $a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 = \mathbf{0}$ is $a_1 = a_2 = a_3 = 0$.

In other words, we want to show that:

> The solution $a_1 = a_2 = a_3 = 0$ to the equation $a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 = \mathbf{0}$ is unique.

This is slightly different than the previous case, since we already know one solution to the equation. Here, the format becomes:

1. **Assumption:** a_1, a_2, a_3 is a solution to the equation $a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 = \mathbf{0}$. (We do not need to assume the existence of another solution, since we already know that the "other" solution is $a_1 = a_2 = a_3 = 0$.)

2. **Conclusion:** Our "other" solution a_1, a_2, a_3 is actually $a_1 = a_2 = a_3 = 0$.

In other words, we want to show that if $a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 = \mathbf{0}$, then $a_1 = a_2 = a_3 = 0$.

11 Logic I: Negations

Logic can sometimes appear in a theoretical math class in ways more complicated than just if-then statements. One common situation in which this happens is when we consider the *negation* of a statement. In other words, how do we show that a given statement is false?

Negating an if-then statement. If you want to show that "If A, then B," is false, you just have to find a particular example where A is true and B is false. For instance, if you want to disprove the statement that "If V is a subspace of $\mathbf{R}^n$, then V has only a finite number of vectors," you just have to note that $V = \mathbf{R}^1$ is a subspace of $\mathbf{R}^1$, but V has an infinite number of vectors. In other words, to disprove this statement, you don't need some kind of "disproof," you just need a *counterexample*, that is, a single example where the statement fails. In fact, to a mathematician, a counterexample is as convincing as any kind of "disproof" could ever be, and is also often much shorter.

Negating "for all" and "there exists." Along the same lines, suppose you want to negate a statement that includes a "for all" or a "there exists." The basic principle to keep

in mind is that the negation of a "for all" statement is a "there exists" statement, and vice versa.

For example, let V be a subset of $\mathbf{R}^2$, and suppose that we are trying to show that V is not a subspace of $\mathbf{R}^2$. By definition, subspaces are closed under $+$ (see Section 9 for an explanation of closure), so it is enough to show that the statement "V is closed under $+$" is false.

To show that "V is closed under $+$" is false, we need to consider the negation of the property

> **Closure under $+$.** For all $\mathbf{v}, \mathbf{w}$ in V, $\mathbf{v} + \mathbf{w}$ is also in V.

This negation is:

> There exist $\mathbf{v}$ and $\mathbf{w}$ in V such that $\mathbf{v} + \mathbf{w}$ is not a vector in V.

That is, to show that V does not have the closure under $+$ property, you just have to come up with a particular choice of $\mathbf{v}$ and $\mathbf{w}$ in V such that $\mathbf{v} + \mathbf{w}$ is not contained in V. You don't need to prove that the $\mathbf{v}$ and $\mathbf{w}$ you choose are particularly interesting, and you don't need to explain where $\mathbf{v}$ and $\mathbf{w}$ came from (maybe you just made a lucky guess); it's just enough to show that they make the closure property fail.

Similarly, the negation of:

> **Closure under inverse.** For all $\mathbf{v} \in V$, there exists a vector $\mathbf{w} \in V$ such that $\mathbf{v} + \mathbf{w} = \mathbf{0}$.

is:

> There exists some $\mathbf{v}$ in V such that there exists no vector $\mathbf{w}$ in V such that $\mathbf{v} + \mathbf{w} = \mathbf{0}$.

or in other words:

> There exists some $\mathbf{v}$ in V such that for all $\mathbf{w}$ in V, $\mathbf{v} + \mathbf{w} \neq \mathbf{0}$.

(Note that the negation of "For all ... there exists ... " has the form "There exists ... for all ... .") This is a little trickier. As before, you only need to find a single $\mathbf{v}$ that makes the axiom fail. However, once you pick that $\mathbf{v}$, you have to show that *any* vector $\mathbf{w}$ cannot be an additive inverse for $\mathbf{v}$.

Summary. Much of the above discussion can be summarized in the following table:

Given the statement:	To prove it:	To disprove it:
"For all x, x is good."	Assume arbitrary x, show x is good	Find a single bad x (counterexample)
"There exists x such that x is good."	Find a single good x (example)	Assume arbitrary x, show x is bad

12 Logic II: Converse and contrapositive

Another important piece of logic in theoretical math is the relationship among an if-then statement, its *converse*, and its *contrapositive*.

The *converse* of the statement "If P, then Q" is "If Q, then P." It is important to realize that the truth of a statement is completely unrelated to the truth of its converse, as confusing a statement with its converse is a common logical error.

For example, consider the statement

(*) If I play center on a professional basketball team, then I am tall.

The converse of (*) is:

If I am tall, then I play center on a professional basketball team.

Note that (*) is true, but its converse is false. (Counterexample: Find a tall person who doesn't play center on a professional basketball team.)

On the other hand, the *contrapositive* of the statement "If P, then Q" is "If (not Q), then (not P)." The contrapositive of a statement is logically equivalent to it, and is occasionally easier to prove.

For example, the contrapositive to (*) is:

If I am short, then I do not play center on a professional basketball team.

Again, note that this statement is logically equivalent to the statement (*).

For another use of the contrapositive, see Section 13.

13 Functions, one-to-one, and onto

The following ideas are useful in many classes.

Definition 19. Let X and Y be sets. A *function* $f : X \to Y$ is a rule that assigns a $y \in Y$ to every $x \in X$ (i.e., an output y for each possible input x). The set X is called the *domain* of f, and the set Y is called the *codomain* of f. (The codomain is sometimes also called the *range* of f.)

Note that the definition of a function f isn't just the formula for f; it also includes the domain and codomain. In fact, it's easy to find two different functions with the same formula; just take your favorite function (e.g., $f : X \to Y$, $X = \mathbf{R}$, $Y = \mathbf{R}$, $f(x) = x^2$) and make its domain smaller (e.g., $f_0 : X_0 \to Y$, $X_0 = \{1, 2, 3, \dots\}$, $Y = \mathbf{R}$, $f(x) = x^2$) to get a different function with the same formula.

Now let X and Y be sets, and let $f : X \to Y$ be a function.

Definition 20. The function $f : X \to Y$ is said to be *one-to-one*, or *injective*, if, for $x_1, x_2 \in X$, $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$. That is, if we think of f as an equation $y = f(x)$, different x values give different y values.

If you want to prove a function is one-to-one, it's usually easier to use the following version of the definition, which is just the contrapositive (see Section 12) of the definition we first gave, and therefore logically equivalent:

Definition 21. The function $f : X \to Y$ is said to be one-to-one if, for $x_1, x_2 \in X$, $f(x_1) = f(x_2)$ implies $x_1 = x_2$. (That is, if $f(x_1) = f(x_2)$, then $x_1 = x_2$.)

Therefore, to prove a function $f : X \to Y$ is one-to-one, we use the following if-then format:

1. **Assumption:** $f(x_1) = f(x_2)$ for some $x_1, x_2 \in X$.
2. **Conclusion:** x_1 is actually equal to x_2.

Note that this process resembles a uniqueness proof, in that we first make a bogus assumption that two objects might be different, and then eventually find out that they're the same.

Next:

Definition 22. The function $f : X \to Y$ is said to be *onto*, or *surjective*, if, for any $y \in Y$, there is some $x \in X$ such that $f(x) = y$.

To prove a function $f : X \to Y$ is onto:

1. Assume that y is an element of Y; and then
2. Find some element x of X such that $f(x) = y$.

In other words, you have to show that the equation $f(x) = y$ can always be solved for x, given y.

Finally:

Definition 23. The function $f : X \to Y$ is said to be *bijective* if f is both one-to-one and onto (i.e., both injective and surjective).

To prove a function is bijective, you do two proofs: a one-to-one proof, and an onto proof.

14 When are two functions equal?

By definition, two functions f and g are equal if:

1. f and g have the same domain and codomain (e.g., $f : X \to Y$ and $g : X \to Y$); and
2. $f(x) = g(x)$ for every $x \in X$.

That is, two functions are equal if they have the same domains and codomains, and every possible input in the domain produces the same output for both functions.

There are two subtleties to equality of functions. One is that it is possible that two functions f and g agree for infinitely many values of x, but are different functions. For example, the functions $f : \mathbf{R} \to \mathbf{R}$ and $g : \mathbf{R} \to \mathbf{R}$ defined by

$$f(x) = \sin x, \qquad\qquad g(x) = 0,$$

agree for all $x = n\pi$, n an integer, but are not equal as functions, since $f(\pi/2) = 1$ and $g(\pi/2) = 0$.

The other subtlety is that it is possible to have two functions whose formulas appear different, but are nevertheless equal. For example, the functions $f : \mathbf{R} \to \mathbf{R}$ and $g : \mathbf{R} \to \mathbf{R}$ defined by

$$f(x) = (\cos x)^2 + (\sin x)^2, \qquad\qquad g(x) = 1,$$

are equal, as you may recall from trigonometry.

The main substance of proving that two functions f and g are equal comes in showing that every possible input in the domain produces the same output for both f and g. In other words, you have to prove:

> For every $x \in X$, $f(x) = g(x)$.

As an if-then statement, this becomes:

> If x is an arbitrary element of X, then $f(x) = g(x)$.

The corresponding if-then format is:

1. **Assumption:** x is an arbitrary element of X.
2. **Conclusion:** $f(x) = g(x)$.

Part IV

Special techniques

15 Induction

Mathematical induction is based on the following axiom:

> **Principle of induction.** Given a logical statement $P(n)$ that depends on a positive integer n, if we can show that:
>
> 1. **Base case:** $P(1)$ is true, and
> 2. **Induction step:** If $P(k)$ is true, then $P(k+1)$ is true;
>
> Then $P(n)$ is true for all positive integers n.

For example, consider the following theorem:

Theorem 24. *For any integer $n > 0$,*

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

To prove this theorem by induction, we let

$$P(n) = \text{“}1 + 2 + \cdots + n \text{ is equal to } \frac{n(n+1)}{2}.\text{”}$$

The base case can be shown by a direct calculation, as is often the case with base cases, so we'll concentrate on the induction step. There, since we're trying to show that “If $P(k)$ is true, then $P(k+1)$ is true,” we use the following if-then format:

1. **Assumption:** $P(k)$ is true.
2. **Conclusion:** $P(k+1)$ is true.

Restating the if-then format using the definition of $P(n)$, we get:

1. **Assumption:** $1 + 2 + \cdots + k = \dfrac{k(k+1)}{2}$.
2. **Conclusion:** $1 + 2 + \cdots + k + (k+1) = \dfrac{(k+1)((k+1)+1)}{2}$.

16 Epsilon-deltonics

Finally, one intimidating type of statement that students must often prove in analysis courses is an “ϵ-δ” statement like the following:

Theorem 25. *For every real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that if $0 < |x - 3| < \delta$, then $\left|x^2 - 9\right| < \epsilon$.*

If you have experience with a subject called *analysis*, you may recognize that this theorem is equivalent to the statement $\lim_{x \to 3} x^2 = 9$.

Let's break this statement down. First, following Sections 6 and 7, we see that Theorem 25 is equivalent to:

> If we have a real number $\epsilon > 0$, then there exists a real number $\delta > 0$ such that if $0 < |x - 3| < \delta$, then $\left|x^2 - 9\right| < \epsilon$.

Broken down, this becomes:

- **If:** ϵ is a real number such that $\epsilon > 0$;
- **Then:** There exists a real number $\delta > 0$ such that if $0 < |x - 3| < \delta$, then $\left|x^2 - 9\right| < \epsilon$.

So in outline form, the proof becomes:

Proof. Assume ϵ is a real number such that $\epsilon > 0$.
Choose $\delta =$???.

$$\vdots$$
(stuff in the middle)
$$\vdots$$

Therefore, if $0 < |x - 3| < \delta$, then $\left|x^2 - 9\right| < \epsilon$. The theorem follows. □

The new wrinkle here is that the "then" part of the proof itself contains another if-then statement that relies on our choice of δ. However, this isn't so bad, because once we figure out what δ should be, the inner if-then statement can be proven just like any other if-then statement. In fact, expanding out "If $0 < |x - 3| < \delta$, then $\left|x^2 - 9\right| < \epsilon$" using our usual techniques, we see that the proof becomes (still in outline form):

Proof. Assume ϵ is a real number such that $\epsilon > 0$.
Choose $\delta =$???. Now assume that $0 < |x - 3| < \delta$.

$$\vdots$$
(stuff in the middle)
$$\vdots$$

So $\left|x^2 - 9\right| < \epsilon$. Therefore, we have shown that for our choice of δ, if $0 < |x - 3| < \delta$, then $\left|x^2 - 9\right| < \epsilon$. The theorem follows. □

What remains now is to choose an appropriate δ, and then fill in the rest of the verification that, assuming $0 < |x - 3| < \delta$, we have $\left|x^2 - 9\right| < \epsilon$. However, since the art of choosing δ is one of the main hurdles of the beginning of analysis, we'll stop here and leave the rest for an analysis class. Our main point here, besides providing a naturally complicated example of an if-then proof, is that if you understand the basic format of an epsilon-delta proof, you can at least have a framework in which to consider the truly tricky part (choosing δ).