Writing proofs

Tim Hsu, San José State University

Revised February 2016

Contents

Introduction

The goal of these notes is to help you learn to write proofs and begin to study proof-intensive mathematics. We assume 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. We will refer to some other subjects occasionally (number theory, analysis), but we won't assume any knowledge of them.

There are six parts to these notes. The first four parts discuss what a proof is and how to write one; specifically, 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. The final parts of these notes discuss miscellaneous proof-related topics; specifically, Part V describes how to give math presentations of various types; and Part VI gives a "Moore method" introduction to various areas of theoretical mathematics. (The course numbers in Part VI are from San José State, but if you are using these notes elsewhere, there are almost certainly analogous courses at your institution.)

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 1.1. 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 1.2. *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 2.1. *If an integer is (exactly) divisible by 6, it must be even.*

Suppose you understand why Theorem 2.1 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 5.) 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.

3 A word about definitions

When you read a definition of a word in a dictionary, usually you expect it to make some kind of intuitive sense right away. ("Proof: Evidence establishing that a statement is true.") Your intuition is then confirmed by seeing the word used in a sentence. ("She had seen conclusive *proof*, so she was certain that he was innocent.")

One problem with mathematical definitions is that they have often evolved over a period of time to become shorter, more precise, and more logically useful. In that process, these definitions often get farther and farther away from ordinary intuition, making them hard to understand at first glance. Therefore, the usual process of understanding a definition (get an idea, confirm it with a few examples) may not work, or may produce a misleading impression.

Here's a better approach to understanding a mathematical definition:

1. Read the definition, and try to think of it as a set of rules, without doing too much guessing as to what it's actually supposed to mean.

2. **Try to make up your own examples of the thing being defined.**

3. Look at any examples given in the book.

4. While repeating steps 2 and 3, gradually form an intuitive impression of what the definition means.

For example, consider the following definition.

Definition 3.1. A *sequence* is a function $X : \mathbf{N} \to \mathbf{R}$, where $X(n)$ is usually written x_n.

Brief and precise, but if you've never seen this before, your first reaction may be, "Huh? What th'?"

Turning to step 2 of the suggested process, the definition begins:

> A *sequence* is a function $X : \mathbf{N} \to \mathbf{R}$...

This means that a sequence is a function that takes a natural number as input and gives a real number as output. Well, you should be familiar with lots of functions from calculus; let's take everyone favorite, $f(x) = x^2$. Certainly if you plug a natural number into that, you get a real number as output, so that seems to work.

So let's continue with the full definition:

> A *sequence* is a function $X : \mathbf{N} \to \mathbf{R}$, where $X(n)$ is usually written x_n.

For whatever reason, the author of this definition seems to prefer X as the function name, and n as the name of the independent variable, which means that our example should really be written $X(n) = n^2$. In fact, following the last suggestion of the definition, we should really write:

> $x_n = n^2$ is an example of a sequence.

Now turning to step 3, a reasonable text might continue:

> Informally, instead of writing down a formula for x_n, we often list the elements $(x_1, x_2, x_3, \ldots)$ to suggest a pattern. For example, to suggest the sequence $x_n = 2n$, we might list $(2, 4, 6, 8, \ldots)$.

At this point, you should think, "Aha! Let me stop reading and apply this new thought to my own example!" If you do so, after some thought, you might eventually deduce:

> Another way of writing the sequence $x_n = n^2$ is to write $(x_1, x_2, x_3, \ldots) = (1, 4, 9, \ldots)$.

In fact, after a few more examples like this, you might realize that all this nonsense about $X : \mathbf{N} \to \mathbf{R}$ and so on is just a way to make the idea of $1, 4, 9, \ldots$ "going on forever" precise and brief.

Exercise: Try to write your own definition of sequence. Can you find a definition that is as short as the one given above, while still preserving the intuitive idea of sequence?

Part II

The structure of proofs

4 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 2.1 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 5) and working with *sets* (Section 6).

5 The if-then method

To describe the assumptions/conclusions structure from Section 4 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 2.1 again, slightly restated:

Theorem 5.1. *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 5.1:

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 5.2. *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.

6 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 6.1. *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 6.1 is exactly the same as Theorem 5.2?)

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 5.2 and 6.1:

Theorem 6.2. *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 6.2 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 6.2 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 6.2 is equivalent to Theorem 6.1.

Part III

Applying if-then

7 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 7.1. *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 7.2. *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 7.3. *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 7.4. *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.

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 8.1. *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 8.2. *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 6), we can change this statement into:

Theorem 8.3. *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* $c, d, e \in \mathbf{R}$.

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 7.

Finally, we sometimes need to prove that a set A is empty. One natural way to do that is by *contradiction*, as follows:

- **Assumption:** x is an element of A. So the defining properties of objects in A hold for x.

  (stuff)

- **Conclusion:** (some kind of false statement or logical contradiction)

For more general uses of proof by contradiction, see Section 21.

9 Nested if-then statements

Another situation commonly found in proofs is where both the assumption and the conclusion of what you want to prove are themselves if-then statements. For example, consider the following definition and theorem.

Definition 9.1. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be vectors in $\mathbf{R}^n$. To say that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is **linearly independent** means: If, for some $a_i \in \mathbf{R}$, $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$, then $a_1 = 0$, $\ldots$, $a_k = 0$.

Theorem 9.2. *If* $\mathbf{v}, \mathbf{w}, \mathbf{x} \in \mathbf{R}^n$ *and* $\{\mathbf{v}, \mathbf{w}, \mathbf{x}\}$ *is linearly independent, then* $\{\mathbf{v}, \mathbf{w}\}$ *is linearly independent.*

In our usual format, applying the defininition of linear independence, the proof of this theorem starts as:

- **Assumptions:** $\mathbf{v}, \mathbf{w}, \mathbf{x} \in \mathbf{R}^n$. If $a\mathbf{v} + b\mathbf{w} + c\mathbf{x} = \mathbf{0}$ for some $a, b, c \in \mathbf{R}$, then $a = 0$, $b = 0$, and $c = 0$.

  (stuff)

- **Conclusion:** If $d\mathbf{v} + e\mathbf{w} = \mathbf{0}$ for some $d, e \in \mathbf{R}$, then $d = 0$ and $e = 0$.

It is important to realize that so far, our assumptions are fairly useless on their own. The reason is, assuming "If P, then Q" says nothing about the truth of either P or Q, so our only concrete assumption so far is that $\mathbf{v}, \mathbf{w}, \mathbf{x} \in \mathbf{R}^n$. On the other hand, our conclusion is more useful than it might first appear, because to prove the if-then statement "If $d\mathbf{v} + e\mathbf{w} = \mathbf{0}$ for some $d, e \in \mathbf{R}$, then $d = 0$ and $e = 0$," we just need to show that the assumption "$d\mathbf{v} + e\mathbf{w} = \mathbf{0}$ for some $d, e \in \mathbf{R}$" leads to the conclusion "$d = 0$ and $e = 0$." We therefore create another, nested, level of assumptions and conclusions:

- **Assumptions:** $\mathbf{v}, \mathbf{w}, \mathbf{x} \in \mathbf{R}^n$. If $a\mathbf{v} + b\mathbf{w} + c\mathbf{x} = \mathbf{0}$ for some $a, b, c \in \mathbf{R}$, then $a = 0$, $b = 0$, and $c = 0$.
  - **Assumption:** $d\mathbf{v} + e\mathbf{w} = \mathbf{0}$ for some $d, e \in \mathbf{R}$.
    (stuff)
  - **Conclusion:** $d = 0$ and $e = 0$.
- **Conclusion:** If $d\mathbf{v} + e\mathbf{w} = \mathbf{0}$ for some $d, e \in \mathbf{R}$, then $d = 0$ and $e = 0$.

The missing stuff then boils down to showing that the assumption $d\mathbf{v} + e\mathbf{w} = \mathbf{0}$ for some $d, e \in \mathbf{R}$ implies a statement of the form $a\mathbf{v} + b\mathbf{w} + c\mathbf{x} = \mathbf{0}$ for some $a, b, c \in \mathbf{R}$, whereupon we may then conclude that $a = 0$, $b = 0$, and $c = 0$. (Exercise!)

For another situation involving nested if-then statements, suppose W, X, Y, and Z are sets, and suppose we want to prove:

Theorem 9.3. *If $W \subseteq X$ then $Y \subseteq Z$.*

Applying our standard approach to set inclusions, we have:

- **Assumptions:** If $a \in W$, then $a \in X$.

  (stuff)

- **Conclusion:** If $b \in Y$, then $b \in Z$.

Again, the structure of our proof is dictated by the conclusion we wish to draw, and we get the following nested if-then structure:

- **Assumptions:** If $a \in W$, then $a \in X$.
  - **Assumption:** $b \in Y$.
    (stuff)
  - **Conclusion:** $b \in Z$.
- **Conclusion:** If $b \in Y$, then $b \in Z$.

10 Or statements

Occasionally, it's necessary to prove an "if-then" statement where either the assumption or the conclusion has an "or" in it. If the assumption of the statement involves an "or", then essentially, you have to prove two theorems. For example, to prove "If P or Q, then R,", you prove "If P, then R" and "If Q, then R."

If the conclusion of a theorem has an "or" in it, things are a bit more complicated. One notable example is:

Theorem 10.1. *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. 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 10.1 becomes:

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

Or in other words:

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

Applying our "nested if-then" techniques from Section 9, this becomes:

- **Assumptions:** a and b are integers and ab is even.
  - **Assumption:** a is not even.
    (stuff)
  - **Conclusion:** b is even.

11 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 11.1. *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 7, 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 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.

12 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 $a_1 = a_2 = a_3 = 0$.

13 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 22 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

14 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 15.

15 Functions; ill-defined and well-defined

The following ideas are useful in many classes.

Definition 15.1. 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.

Occasionally, when you try to define a function with certain properties, you end up with a function whose formula/definition is ambiguous, incomplete, or self-contradictory. Such a function is called *ill-defined*. For example, suppose we want to define a function $f : \mathbf{R} \to \mathbf{R}$ by the following formula:

$$f(x) = \begin{cases} 0 & \text{if } x > 0, \\ 1 & \text{if } x \text{ is a rational number.} \end{cases}$$

The definition of f has two kinds of problems. On the one hand, there are certain elements x of the domain of f such that $f(x)$ has no definition at all. For example, for $-\pi \in \mathbf{R}$, $f(-\pi)$ is not covered by the above formula, since $-\pi$ is not positive and not rational. On the other hand, there are certain elements of the domain of f such that $f(x)$ has more than one value. For example, for $3 \in \mathbf{R}$, the above formula says that $f(3) = 0$, since $3 > 0$, but it also says that $f(3) = 1$, since 3 is a rational number.

Therefore, we say that a function $f : X \to Y$ is *well-defined* by a given formula or other rule, if, for every $x \in X$, that formula/rule produces *at least* one value $f(x) \in Y$ and also *at most* one value $f(x) \in Y$. In shorthand form, a well-defined function is therefore one whose formula produces *exactly* one value in its codomain for every input from its domain. However, in practice, it is often helpful to prove that a function is well-defined in two steps: (1) Show that every input produces at *least* one output, and (2) Show that every input produces at *most* one output.

For example, suppose we want to define a function $g : \mathbf{R} \to \mathbf{R}$ by the following formula:

$$g(x) = \begin{cases} 0 & \text{if } x \text{ is rational,} \\ 1 & \text{if } x \text{ is irrational.} \end{cases}$$

Then g is a well-defined function because: (1) For every $x \in \mathbf{R}$, x is either rational or irrational, which means that at least one of the options in the above formula applies, and (2) For every $x \in \mathbf{R}$, x cannot be both rational and irrational, which means that at most one of the options in the above formula applies.

16 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

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

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:** $f(x) = g(x)$.

17 One-to-one and onto

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

Definition 17.1. 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 14) of the definition we first gave, and therefore logically equivalent:

Definition 17.2. 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 17.3. 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.

Putting the ideas of one-to-one and onto together, we get:

Definition 17.4. 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.

To conclude, the following comparison may help to avoid confusing the ideas of being well-defined, one-to-one, and onto. Let $f : X \to Y$ be a function (possibly not well-defined).

- To say that f is *well-defined* means that for every *input* $x \in X$, there exists *exactly one* output $f(x) \in Y$.
- To say that f is *one-to-one* means that for every *possible output* $y \in Y$, there exists *at most one* $x \in X$ such that $f(x) = y$.
- To say that f is *onto* means that for every *possible output* $y \in Y$, there exists *at least one* $x \in X$ such that $f(x) = y$.

Note that this version of the definition of one-to-one ("at most one input") is not very useful for proving f is one-to-one, but it may provide some intuition.

18 Inverses of functions

While you are probably familiar with the idea of the inverse f^{-1} of a function f from calculus, the following precise version of that definition becomes important in certain more advanced classes, especially advanced linear algebra.

Definition 18.1. The *identity function* on a set X is the function $\mathrm{id}_X : X \to X$ defined by $\mathrm{id}_X(x) = x$ for all $x \in X$.

Definition 18.2. Let $f : X \to Y$ and $g : Y \to Z$ be functions. We define the *composite function* $g \circ f : X \to Z$ by the formula $(g \circ f)(x) = g(f(x))$ for all $x \in X$.

Definition 18.3. We say that functions $f : X \to Y$ and $g : Y \to X$ are *inverses* if $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. (See Section 16 for the definition of equality of functions.) Given $f : X \to Y$, we say that f is *invertible* if there exists some $g : Y \to X$ such that f and g are inverses, and we say that g is an *inverse* of f.

One can then show that inverses are unique if they exist (exercise), which means that we can refer to *the* inverse f^{-1} of a function f. The key point is then the following theorem, which characterizes when a function is invertible.

Theorem 18.4 (The inverse theorem)**.** *Let $f : X \to Y$ be a function. Then the following are equivalent:*

1. *f is bijective.*
2. *The function $g : Y \to X$ given by the formula*

$$g(y) = \text{the unique } x \in X \text{ such that } f(x) = y.$$

   is well-defined.

3. *f is invertible.*

Moreover, if any (and therefore, all) of these conditions hold, then f^{-1} is the function g defined in condition (2).

Exercise: Prove the inverse theorem. Note that the definition of f^{-1} coming from the inverse theorem is essentially the same as the definition of f^{-1} given in calculus. (Compare the way the exponential function is used to define the natural log function.)

19 Restrictions

Since the domain and codomain of a function f are part of the definition of f, and not just something derived from the formula of f, it is sometimes useful to have a way to describe the relationship between two functions with the same formula, but different domains or codomains. We therefore have the following.

Definition 19.1. Let $f : X \to Y$ be a function, let A be a subset of X, and let B be a subset of Y. We define the *restriction of f to A* to be the function $f|_A : A \to Y$ defined by $f|_A(x) = f(x)$ for all $x \in A$. (I.e., we use the same formula, but have fewer possible inputs.) Similarly, if it happens to be the case that for all $x \in X$, we have $f(x) \in B$, we define the *co-restriction of f to B* to be the function $f|^B : X \to B$ given by $f|^B(x) = f(x)$ for all $x \in X$. Finally, if it happens to be the case that for all $x \in A$, we have $f(x) \in B$, we define the *bi-restriction of f to A, B* to be the function $f|_A^B : A \to B$ given by $f|_A^B(x) = f(x)$ for all $x \in A$.

Note that the terms co-restriction and bi-restriction are not standard; in fact, there do not seem to be standard terms for these ideas. They are nevertheless useful; just keep an eye out for different names for them (or sometimes, no name at all).

Part IV

Special techniques

20 Induction

The goal of induction is to prove an infinite sequence of theorems using a special kind of "infinite proof". The point is that while all proofs must be finite in length, accepting the principle of mathematical induction (really an axiom) effectively allows the use of one very particular kind of infinite proof.

More precisely, suppose we want to prove an infinite sequence of theorems Thm(1), Thm(2), Thm(3), ... indexed by the positive integers. The induction axiom says:

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

Note that the assumption in the "if" part of the induction step is that Thm(k) is true for **one** fixed but arbitrary value of k, not for **all** values of k. (If we knew that Thm(k) were true for all values of k, we would be done!) In practice, this means that when you are proving an induction step, you cannot make any assumptions about the value of k, other than $k \geq 1$, and you cannot choose or otherwise change the value of k within your proof.

For example, consider the following theorem:

Theorem 20.1. *For any integer $n > 0$,*

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

To prove this theorem by induction, we let

$$\text{Thm}(n) = \text{"}1 + 2 + \cdots + n \text{ is equal to } \frac{n(n+1)}{2}.\text{"}$$

The base case can be shown by direct calculation, so we'll concentrate on the induction step. There, since we're trying to show that "If Thm(k) is true for some positive integer k, then Thm($k+1$) is true," we use the following if-then format:

1. **Assumption:** Thm(k) is true for some fixed but arbitrary positive integer k.
2. **Conclusion:** Thm($k+1$) is true.

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

1. **Assumption:** $1 + 2 + \cdots + k = \dfrac{k(k+1)}{2}$ for some fixed but arbitrary positive $k \in \mathbb{Z}$.

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

Variation: Base case $\neq 1$. One small variation comes if we want to prove that, for example, "Thm(n) is true for all integers $n \geq 47$". To prove that statement, we prove:

1. **Base case:** Thm(47) is true, and
2. **Induction step:** If Thm(k) is true for some positive integer $k \geq 47$, then Thm($k+1$) is true.

Variation: Strong induction. Another variation is the idea of *strong induction*. Briefly, it can be shown that the usual axiom of induction is equivalent to the following axiom:

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

The main benefit of strong induction is that the format for the induction step becomes:

1. **Assumption:** n is a fixed but arbitrary positive integer, Thm(k) is true for all positive integers $k < n$.
2. **Conclusion:** Thm(n) is true.

In other words, we are allowed to assume more than we are allowed to assume when using regular induction. This can be quite helpful, especially if the statement Thm(n) depends on n multiplicatively and not additively (e.g., theorems having to do with factorizing integers).

Variation: Proving two-variable theorems. Suppose now we want to prove a logical statement Thm(n, k) for all positive integers n and k. One way to approach this by induction is to define the statement $Q(n) =$ "Thm(n, k) is true for all positive integers k", and set up the induction as follows.

1. **Base case:** $Q(1)$ is true, and
2. **Induction step:** If $Q(k)$ is true for some positive integer k, then $Q(k+1)$ is true.

In other words, it is enough to show that:

1. **Base case:** Thm($1, k$) is true for all positive integers k, and
2. **Induction step:** If Thm(n, k) is true for some positive integer n and all positive integers k, then Thm($n+1, k$) is true for all positive integers k.

To get even fancier, you might then try to prove Thm($1, k$) and Thm($n+1, k$) (for fixed but arbitrary n) by induction on k. This technique is called *multiple induction*.

21 Contradiction

The idea behind *proof by contradiction* is, if you want to show that the assumptions of a theorem lead to its conclusion, you can do the following:

1. Assume that the conclusion of the theorem is false.
2. Deduce logically, from this assumption, either that:
   (a) The hypotheses of the theorem are false, contradicting the fact that they have been assumed, or
   (b) The assumption made in step 1 (that the conclusion of the theorem is false) is itself false.

The theorem then follows.

Proof by contradiction is often used to show that something does not exist, or that "it is impossible to find…". For example, consider the following example, which is due to Euclid, and is probably the most famous example of a proof by contradiction. First, a definition:

Definition 21.1. A *prime number* is an integer $p > 1$ such that the only positive integers that divide p are 1 and p itself.

To prove Euclid's theorem, we'll use the following fact (without proof, for brevity):

Theorem 21.2. *Every integer $n > 1$ is divisible by some prime number.*

Euclid's theorem says:

Theorem 21.3. *It is impossible to find the "largest" prime number p.*

Proof. Assume that p is the largest possible prime number. We will show that this leads to a logical contradiction.

Let $N = 1 \times 2 \times \cdots \times p + 1$. By Theorem 21.2, N must be evenly divisible by some prime number q. On the one hand, we know by assumption that p is the largest possible prime number, so we must have $2 \leq q \leq p$. On the other hand, no positive number between 2 and p can divide N evenly, because when you divide N by any number between 2 and p, you get a remainder of 1. Contradiction. Therefore, our original assumption that there exists a largest possible prime number p must be incorrect. □

22 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 22.1. 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 22.2. *X is closed under vector addition.*

First, using the definition of closure, we restate Theorem 22.2 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 6, 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!)

23 Epsilonics

If you're taking analysis, you know (or you'll soon discover) that proving any statement involving an ϵ can seem intimidating. However, with a little work, such proofs can be analyzed in our "if-then" framework, just like every other proof.

23.1 The limit of a sequence

In analysis, the heart of the idea of the limit of a sequence is what we might call an "ϵ-N" statement. For example, by the definition of the limit of a sequence, to prove that the limit of the sequence $a_n = \dfrac{2^n}{2^n+1}$ is 1, we need to prove:

Theorem 23.1. *For every real number $\epsilon > 0$, there exists a natural number N such that if $n > N$, then $\left|\dfrac{2^n}{2^n+1} - 1\right| < \epsilon$.*

Let's break this statement down. First, following Sections 7 and 11, we see that Theorem 23.1 is equivalent to:

> If we have a real number $\epsilon > 0$, then there exists a natural number $N > 0$ such that if $n > N$, then $\left|\dfrac{2^n}{2^n+1} - 1\right| < \epsilon$.

Broken down, this becomes:

- **If:** ϵ is a real number such that $\epsilon > 0$;
- **Then:** There exists a natural number $N > 0$ such that if $n > N$, then $\left|\dfrac{2^n}{2^n+1} - 1\right| < \epsilon$.

So in outline form, the proof becomes:

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

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

Therefore, if $n > N$, then $\left|\dfrac{2^n}{2^n+1} - 1\right| < \epsilon$. 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 N. However, this isn't so bad, because once we figure out what N should be, the inner if-then statement can be proven just like any other if-then statement. In fact, expanding out "If $n > N$, then $\left|\dfrac{2^n}{2^n+1} - 1\right| < \epsilon$" 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 $N =???$. Now assume that $n > N$.

$$\vdots$$

So $\left|\dfrac{2^n}{2^n+1} - 1\right| < \epsilon$. Therefore, we have shown that for our choice of N, if $n > N$, then $\left|\dfrac{2^n}{2^n+1} - 1\right| < \epsilon$. The theorem follows. □

What remains now is to choose an appropriate N, probably depending on ϵ in some fashion, and then fill in the rest of the verification. However, since the art of choosing N can be tricky, we'll stop here. The main point, besides providing a naturally complicated example of an if-then proof, is that if you understand the basic format of an ϵ-N proof, you can at least have a framework in which to consider the truly tricky part (choosing N).

One variation on the above outline analysis comes when a sequence has a limit of "$+\infty$" or "$-\infty$". For example, to prove that the limit of the sequence $n^2 + 1$ is $+\infty$, we need to prove:

Theorem 23.2. *For every real number $K > 0$, there exists a natural number N such that if $n > N$, then $n^2 + 1 > K$.*

Proceeding as in the finite case, Theorem 23.2 is equivalent to:

> If K is a real number such that $K > 0$, then there exists a natural number N such that if $n > N$, then $n^2 + 1 > K$.

Broken down, this becomes:

- **If:** K is a real number such that $K > 0$;
- **Then:** There exists a natural number N such that if $n > N$, then $n^2 + 1 > K$.

And in outline form, the proof becomes:

Proof. Assume K is a real number such that $K > 0$.

Choose $N =???$. Now assume that $n > N$.

$$\vdots$$

So $n^2 + 1 > K$. Therefore, we have shown that for our choice of N, if $n > N$, then $n^2 + 1 > K$. The theorem follows. □

23.2 Limits and continuity of functions

Similarly, when you study the limit of a function $f(x)$ as the input variable x approaches some given value, you often need to prove an "ϵ-δ" statement. For example, to prove that $\lim_{x\to 3} x^2 = 9$, you need to prove:

Theorem 23.3. *For every real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that if $0 < |x-3| < \delta$, then $\left|x^2 - 9\right| < \epsilon$.*

Similarly, to prove that $f(x) = x^2$ is continuous at $x = 3$, you need to prove:

Theorem 23.4. *For every real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that if $|x-3| < \delta$, then $\left|x^2 - 9\right| < \epsilon$.*

Since the two statements are so similar, we'll stick with outlining the proof of Theorem 23.3. Again, following Sections 7 and 11, if we break down the statement of Theorem 23.3, we see that it 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 $\delta =$???.

$\vdots$

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

To go one step further, expanding out "If $0 < |x-3| < \delta$, then $\left|x^2 - 9\right| < \epsilon$" using our usual techniques, we see that the proof becomes:

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

$\vdots$

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. □

Again, we are left only with the truly tricky part of choosing δ.

As with the limit of a sequence, we also have a slightly different outline decomposition with infinite limits. For example, to prove that $\lim_{x\to 3} \dfrac{1}{(x-3)^2} = +\infty$, by definition of an infinite-valued limit, we must show:

Theorem 23.5. *For every real number $K > 0$, there exists a real number $\delta > 0$ such that if $0 < |x - 3| < \delta$, then $\dfrac{1}{(x-3)^2} > K$.*

Analyzing Theorem 23.5, we eventually arrive at the following outline:

Proof. Assume K is a real number such that $K > 0$.

Choose $\delta =$???. Now assume that $0 < |x - 3| < \delta$.

$$\vdots$$

So $\dfrac{1}{(x-3)^2} > K$. Therefore, we have shown that for our choice of δ, if $0 < |x - 3| < \delta$, then $\dfrac{1}{(x-3)^2} > K$. The theorem follows. □

As a final variation, we can also consider "limits at infinity". For example, to prove that $\lim\limits_{x \to +\infty} \dfrac{2x-3}{x} = 2$, by definition, we must show:

Theorem 23.6. *For every real number $\epsilon > 0$, there exists a real number $N > 0$ such that if $x > N$, then $\left|\dfrac{2x-3}{x} - 2\right| < \epsilon$.*

Again, the proof of Theorem 23.6 has the following outline:

Proof. Assume ϵ is a real number such that $\epsilon > 0$.

Choose $N =$???. Now assume that $x > N$.

$$\vdots$$

So $\left|\dfrac{2x-3}{x} - 2\right| < \epsilon$. Therefore, we have shown that for our choice of N, if $x > N$, then $\left|\dfrac{2x-3}{x} - 2\right| < \epsilon$. The theorem follows. □

As the reader may have noticed, this last structure is essentially the same as the outline for proving the limit of a sequence, except that instead of having an integer independent variable n, we have a real-valued independent variable x.

23.3 Sequential definition of continuity

One last type of (hidden!) ϵ proof comes from what is sometimes known as the sequential definition of continuity. For example:

Theorem 23.7. *Let $f(x) = x^2$. If x_n is a sequence such that $\lim\limits_{n \to \infty} x_n = 3$, then $\lim\limits_{n \to \infty} f(x_n) = f(3) = 9$.*

Interestingly, this theorem is equivalent to the statement that $f(x)$ is continuous at $x = 3$ under the ϵ-δ definition. Similarly, add the condition $x_n \neq 3$ for all n to the "if", and you get a statement equivalent to Theorem 23.3.

The main benefit of this sequence-based approach to continuity is that we can prove facts about continuity without apparently having to resort to epsilons. For example, the outline for Theorem 23.7 starts:

Proof. Assume x_n is a sequence such that $\lim_{n\to\infty} x_n = 3$.

$$\vdots$$

Therefore, $\lim_{n\to\infty} f(x_n) = f(3) = 9$. The theorem follows. □

And in fact, if you already happen to know something about $\lim_{n\to\infty} f(x_n)$, given $\lim_{n\to\infty} x_n$, then you can apply that result here. However, if you don't know how to obtain $\lim_{n\to\infty} f(x_n)$ given $\lim_{n\to\infty} x_n$, then what you need to do is an ϵ-N proof given an ϵ-N assumption:

Proof. Assume x_n is a sequence such that $\lim_{n\to\infty} x_n = 3$. Then by definition, for any $\epsilon_x > 0$, there exists some natural number $N_x(\epsilon_x)$ such that if $n > N_x(\epsilon_x)$, then $|x_n - 3| < \epsilon$.

Assume ϵ is a real number such that $\epsilon > 0$.

Choose $N =$???.

$$\vdots$$

Therefore, if $n > N$, then $|f(x_n) - 9| < \epsilon$. Therefore, $\lim_{n\to\infty} f(x_n) = f(3) = 9$, and the theorem follows. □

Note that we use ϵ_x and N_x when we assume $\lim_{n\to\infty} x_n = 3$, as we will also be dealing with a separate ϵ and N that may have some complicated relationship with ϵ_x and N_x.

Part V

Presentations

24 How to give a math lecture

This section gives a brief guide to giving a math lecture, i.e., a classroom presentation where you are teaching definitions, theorems, proofs, and examples to a class of students.

1. **Write down everything you say.** The main difference between a math lecture and almost any other kind of public speaking you can think of (a literature lecture, a sermon, a political speech) is that **you must write down everything that you say.** This is because, unlike almost any other subject, math generally doesn't have an underlying kind of intuitive sense to it that you can communicate verbally and vaguely without dealing with details. Therefore, anything that you want your audience to understand **must be written down.**

2. **Write it once, say it twice.** As a rule, it often helps your audience if you say what you're going to write as you write it down, and then repeat it once it's already written down. You don't have to do this to a completely mechanical extent, but if you practice this, it should start to become fairly natural.

3. Don't just write down the equations, **write down the words between equations.** This is especially true in a proof class, where the words between the equations and symbols ("for every", "there exists") are almost more important than the equations and symbols themselves.

4. If you are presenting definitions, theorems, and proofs, clearly indicate which is which. It's especially important to separate theorem from proof, and to indicate what you're assuming and what you want to conclude in your proof.

5. Structure your lecture from the top down, like any other kind of oral or written communication. In other words, start with an outline on the highest/broadest level, and then fill in the details.

6. When writing on the board, go from top to bottom, left to right. Don't skip around or proceed in a nonlinear fashion.

7. Avoid large-scale erasing of mistakes if you can, as erasing makes it hard to take notes. Instead, cross things out, or if you do erase, pause for a moment to let people catch up.

8. If you follow all of the above tips, you may start to feel as if you are speaking at an incredibly slow pace and that you are going to grow old and die at the board. If so, you're going at the correct speed to be understood! In fact, slowing yourself down is yet another reason to **write everything down**.

Part VI

Section One

In the "Moore method" of studying theoretical math, at the beginning of the class, instead of a textbook, students are given a list of definitions and theorems. Students then spend the class proving the basic theorems in the subject, essentially writing the textbook they would ordinarily be given.

We won't try anything quite so ambitious here. Instead, our goal is to give you a taste of various topics in theoretical math by giving a Moore method presentation of some of the "Section Ones" (the easy parts!) of these subjects. Each topic is also labelled with the SJSU class(es) in which it is covered; for any non-SJSU users of these notes, it shouldn't be hard to find an analogous class where you are.

25 Abstract algebra (Math 128A): Groups, part I

Definition 25.1. A *group* is a set G and a binary operation $\cdot : G \times G \to G$, written as multiplication (e.g., for $a, b \in G$, $a \cdot b = ab$ is the result of applying the operation to a and b), satisfying the following three axioms:

1. *Associativity.* For $a, b, c \in G$, we have $(ab)c = a(bc)$.
2. *Identity.* There exists an element $e \in G$ such that for all $a \in G$, we have $ea = ae = a$.
3. *Inverse.* For every $a \in G$, there exists an element $b \in G$ such that $ab = ba = e$ (where e is the identity element from the previous axiom).

Theorem 25.2. *Let G be a group. Then the identity element of G is unique.*

Suggestion: See the section on uniqueness (Section 12). If e and e' are identity elements, consider ee'.

Theorem 25.3. *Let G be a group, and let a be an element of G. Then a has a unique inverse element, which we may therefore denote by a^{-1}.*

Suggestion: If b and c are inverses of a, consider bac (why is bac well-defined?).

Theorem 25.4. *Let G be a group, and let a be an element of G. Then $(a^{-1})^{-1} = a$.*

Suggestion: Use the uniqueness of inverses.

26 Abstract algebra (Math 128A): Groups, part II

Theorem 26.1. *Let G be a group. For $a, b, c \in G$, if $ab = ac$, then $b = c$; and if $ac = bc$, then $a = b$.*

Suggestion: Be careful about associativity, and remember that "multiplication" in group need not be commutative.

Theorem 26.2. *Let G be a group. For any $a, b \in G$, there exists a unique $x \in G$ such that $ax = b$, and there exists a unique $y \in G$ such that $ya = b$.*

Suggestion: Again, watch associativity. By the way, it follows from this theorem that the multiplication table (or *Cayley table*) of a group G has the "Sudoku", or *Latin square*, property, i.e., every element of G shows up exactly once in each column and each row of the table.

Theorem 26.3. *Let G be a group, and let v, w, x, y, z be elements of G. Then $v(w(x(yz))) = ((vw)(xy))z$.*

Suggestion: Apply associativity carefully, and specify how you are applying it at each step.

27 Abstract algebra (Math 128A): Group homomorphisms

Definition 27.1. Let G and H be groups. A *homomorphism* from G to H is a function $\varphi : G \to H$ such that $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in G$.

Theorem 27.2. *Let G and H be groups, with identity elements e and e', respectively, and let $\varphi : G \to H$ be a homomorphism. Then $\varphi(e) = e'$.*

Suggestion: Consider $\varphi(ee)$.

Definition 27.3. Let a be an element of a group G with identity element e. We define $a^0 = e$, $a^{n+1} = a^n a$ for any nonnegative integer n, and $a^{-n} = (a^{-1})^n$. Note that expressions like $a^4 = aaaa$ are well-defined, by associativity.

Theorem 27.4. *Let G and H be groups, let $\varphi : G \to H$ be a homomorphism. Then for $g \in G$ and $n \in \mathbf{N}$, $\varphi(g^n) = \varphi(g)^n$.*

Suggestion: Induction.

Theorem 27.5. *Let G, H, and K be groups, and let $\varphi : G \to H$ and $\psi : H \to K$ be homomorphisms. Then $\psi \circ \varphi : G \to K$ is a homomorphism.*

Suggestion: Write out the definition of homomorphism as an if-then statement and then prove the if-then statement for $\psi \circ \varphi$.

28 Abstract algebra (Math 128A/128B): Rings

Definition 28.1. An *additive abelian group* is a group A (see Section 25) with its operation written as the addition symbol $+$ (instead of multiplication), identity written 0, and the inverse of $a \in A$ written $(-a)$, that has the additional property that $a + b = b + a$ for all $a, b \in A$. In an additive abelian group, we define *subtraction* by $a - b = a + (-b)$.

Definition 28.2. A ring is a set R with two binary operations $+ : R \times R \to R$ and $\cdot : R \times R \to R$, satisfying the following axioms:

1. The set R and the operation $+$ form an additive abelian group.
2. *Associativity of multiplication.* For all $a, b, c \in R$, $(ab)c = a(bc)$.
3. *Distributivity.* For all $a, b, c \in R$, $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$.

Theorem 28.3. *Let R be a ring. For all $a \in R$, $a0 = 0a = 0$.*

Suggestion: Use the uniqueness of the identity 0.

Theorem 28.4. *Let R be a ring. For all $a, b \in R$, $a(-b) = (-a)b = -(ab)$.*

Suggestion: Use the uniquesness of inverses.

Theorem 28.5. *Let R be a ring. For all $a, b \in R$, $(-a)(-b) = ab$.*

29 Abstract algebra (Math 128B): Integral domains/fields

Definition 29.1. A *commutative ring* is a ring R such that for $a, b \in R$, $ab = ba$.

Definition 29.2. A *ring with unity* is a ring R that contains an element 1 such that $1a = a1 = a$ for all $a \in R$.

Definition 29.3. An *integral domain* is a commutative ring R with unity such that for $a, b \in R$, if $ab = 0$, then either $a = 0$ or $b = 0$.

Theorem 29.4. *Let R be an integral domain. For $a, b, c \in R$, if $ab = ac$, and $a \neq 0$, then $b = c$.*

Suggestion: Put everything over on one side of the equation.

Definition 29.5. A *field* is a commutative ring F with unity with the property that every $a \in F$ such that $a \neq 0$ has a multiplicative inverse, i.e., that there exists $b \in F$ such that $ab = ba = 1$.

Theorem 29.6. *If F is a field, then F is an integral domain.*

Suggestion: To prove the conclusion "p or q," we prove the if-then statement "If not p, then q."

Theorem 29.7. *The integers $\mathbb{Z}$ are an integral domain, but not a field.*

Suggestion: Give a specific counterexample.

30 Analysis (Math 131A): The limit of a function, part I

Definition 30.1. Let I be a subinterval of $\mathbf{R}$ or a subinterval of $\mathbf{R}$ minus the point a, and let $f : I \to \mathbf{R}$ be a real-valued function. To say that the *limit of f at a is L*, or

$$\lim_{x \to a} f(x) = L,$$

means that for every $\epsilon > 0$, there exists a $\delta > 0$ such that if $|x - a| < \delta$, $x \neq a$, then $|f(x) - L| < \epsilon$.

Theorem 30.2. *Let $f : \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = 7$, and let a be a real number. Then*

$$\lim_{x \to a} f(x) = \lim_{x \to a} 7 = 7.$$

Suggestion: Start with an outline!

Theorem 30.3. *Let $f : \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x$, and let a be a real number. Then*

$$\lim_{x \to a} f(x) = \lim_{x \to a} x = a.$$

Suggestion: Start with an outline; see also the section on epsilonics (Section 23).

31 Analysis (Math 131A): The limit of a function, part II

Theorem 31.1. *Let f be a real-valued function, let c be a real number, and suppose that*

$$\lim_{x \to a} f(x) = L.$$

Then

$$\lim_{x \to a} cf(x) = cL.$$

Suggestion: Outline.

Theorem 31.2. *Let f and g be real-valued functions, and suppose that*

$$\lim_{x \to a} f(x) = L, \qquad \lim_{x \to a} g(x) = M.$$

Then

$$\lim_{x \to a} (f(x) + g(x)) = L + M.$$

Suggestion: Outline.

32 Analysis (Math 131A): Continuous functions

Definition 32.1. Let I be a subinterval of $\mathbf{R}$ containing the point a in its interior, and let $f : I \to \mathbf{R}$ be a real-valued function. To say that f *is continuous at* a means that

$$\lim_{x \to a} f(x) = f(a),$$

or in other words, for every $\epsilon > 0$, there exists a $\delta > 0$ such that if $|x - a| < \delta$, then $|f(x) - f(a)| < \epsilon$.

For the rest of this section, let I be a subinterval of $\mathbf{R}$ containing the point a in its interior, and let $f : I \to \mathbf{R}$ be a real-valued function.

Theorem 32.2. *Suppose f is continuous at a. If (x_n) is a sequence such that $x_n \to a$, then $f(x_n) \to f(a)$.*

Suggestion: This is difficult. Start with an outline, and try to combine the N coming from the definition of limit of a sequence with the δ in the definition of continuity.

Theorem 32.3. *Suppose f is* **not** *continuous at a. There exists a sequence (x_n) such that $x_n \to a$ but $f(x_n) \not\to f(a)$.*

Suggestion: This is really difficult! Start by negating the definition of continuity, and consider $\delta = \dfrac{1}{n}$.

Corollary 32.4. *Let I be a subinterval of $\mathbf{R}$ containing the point a, and let $f : I \to \mathbf{R}$ be a real-valued function. Then f is continuous at a if and only if, for every sequence (x_n) such that $x_n \to a$, we have $f(x_n) \to f(a)$.*

Suggestion: This is just the union of the previous two theorems. That is, if f is continuous at a, then the convergence condition applies to every sequence (x_n), and if f is not continuous at a, then there exists a sequence (x_n) to which the convergence condition does not apply.

33 Analysis (Math 131A): Differentiable functions

Definition 33.1. Let I be an interval in the real line, and let a be a point in I. We say that a function $f : I \to \mathbf{R}$ is *differentiable* at $x = a$ if the limit

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = L$$

exists. If the limit exists, we define $f'(a) = L$.

Theorem 33.2. *Let $f : I \to \mathbf{R}$ and $g : I \to \mathbf{R}$ be functions that are differentiable at $x = a \in I$. Then $h(x) = f(x) + g(x)$ is differentiable at $x = a$, and $h'(a) = f'(a) + g'(a)$.*

Suggestion: Use the limit laws in Section 31.

Theorem 33.3. *Let $f : I \to \mathbf{R}$ be a function that is differentiable at $x = a \in I$, and let $c \in \mathbf{R}$. Then $h(x) = cf(x)$ is differentiable at $x = a$, and $h'(a) = cf'(a)$.*

Suggestion: Limit laws again.

Theorem 33.4. *Let $f : I \to \mathbf{R}$ be a function that is differentiable at $x = a \in I$. Then $f(x)$ is continuous at a.*

Suggestion: Use the ϵ-δ definition of continuity; $f'(a)$ actually gives linear control of ϵ in terms of δ.

34 Complex analysis (Math 138): Holomorphic functions

This section assumes some familiarity with complex numbers.

Definition 34.1. If $x + yi$ is a complex number, then $|x + yi| = \sqrt{x^2 + y^2}$.

Definition 34.2. Let U be a region in $\mathbb{C}$ or a region of $\mathbb{C}$ minus the interior point a, and let $f : U \to \mathbb{C}$ be a complex-valued function. To say that the *limit of f at a is L*, or

$$\lim_{z \to a} f(z) = L,$$

means that for every $\epsilon > 0$, there exists a $\delta > 0$ such that if $|z - a| < \delta$, $z \neq a$, then $|f(z) - L| < \epsilon$.

Definition 34.3. Let U be a region in $\mathbb{C}$, let a be a point in U, and let $f : U \to \mathbb{C}$ be a complex-valued function. We say that a function $f : U \to \mathbf{R}$ is *holomorphic* at $z = a$ if the limit

$$\lim_{h \to 0} \frac{f(z + h) - f(z)}{h} = L$$

exists, where h is allowed to take complex values. If the limit exists, we define $f'(a) = L$.

Theorem 34.4. *Let U be a region in $\mathbb{C}$, let a be a point in U, let $f : U \to \mathbb{C}$ be a complex-valued function that is holomorphic at $z = a$, and let $u(z)$ and $v(z)$ be real-valued functions that are the real and imaginary parts of $f(z)$, respectively (i.e., let $f(z) = u(z) + iv(z)$, where u and v are real-valued). Then at $z = a$,*

$$\frac{\partial f}{\partial x} = -i \frac{\partial f}{\partial y},$$

or in other words,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Suggestion: Note that the limit in Definition 34.3 that defines $f'(a)$ must have the same value, no matter what h is. Therefore, we must get the same result if we let h take real values and approach 0 as if we let $h = ik$ take purely imaginary values and approach 0. This produces the equation in the first part of the theorem, which then yields the second part.

35 Graph theory (Math 142/179): Basic definitions

Definition 35.1. A *graph* is a triple $G = (V, E, \partial)$, where

1. V is a set, called the *vertex set* of G;
2. E is a set, called the *edge set* of G; and
3. ∂ is a function $\partial : E \to \mathcal{P}(V)$ (the power set of V), where $|\partial(e)| = 1$ or 2 for all $e \in E$.

Definition 35.2. Let $G = (V, E, \partial)$ be a graph. A *self-loop* in G is an $e \in E$ such that $|\partial(e)| = 1$. We say that G has *multiple edges* if, for some $x, y \in V$, there exists more than one $e \in E$ such that $\partial(e) = \{x, y\}$. If G has no self-loops or multiple edges, we say that G is a *simple* graph.

Example 35.3. Write down some intuitive examples of graphs (dots connected by lines) and formalize each of your examples in terms of Definition 35.1 (i.e., for each example, write down V, E, and ∂). Make sure you have an example with at least one self-loop and no multiple edges, an example with multiple edges and no self-loops, an example with multiple loops and self-edges, and an example of a simple graph.

Theorem 35.4. *Let V be any set, let V_2 be the set of all subsets of V of size 2, let E be a subset of V_2, and let $\partial : E \to \mathcal{P}(V)$ be defined by $\partial(\{x, y\}) = \{x, y\}$. Then $G = (V, E, \partial)$ is a simple graph.*

Suggestion: Go back to the definitions.

36 Graph theory (Math 142/179): Paths and connectedness

From here onwards, we assume all graphs are simple (no self-loops, no multiple edges).

Definition 36.1. Let G be a graph. We say that two vertices v and w in G are *adjacent* if there exists an edge e in G such that $\partial(e) = \{v, w\}$. (Note that a vertex may be adjacent to itself via a self-loop.)

Definition 36.2. Let G be a graph. For $n \geq 0$, a *path of length n in G* is a finite sequence (or more precisely, an $(n+1)$-tuple) of vertices $(v_0, v_1, \ldots, v_n)$ such that v_i and v_{i+1} are adjacent for $0 \leq i \leq n-1$. We also say that $(v = v_0, v_1, \ldots, v_n = w)$ is a *path from v to w in G*.

Definition 36.3. Let $G = (V, E, \partial)$ be a graph. We define a relation $\sim$ on V by saying that $v \sim w$ if and only if there exists a path (of any length) from v to w in G.

Theorem 36.4. *The relation $\sim$ is an equivalence relation on V.*

Suggestion: Draw pictures of paths.

Definition 36.5. The equivalence classes of $\sim$ are called the *path components* of G. If G has at least one vertex and also has only one path component (i.e., all of the vertices of G are equivalent), we say that G is *path-connected*, or simply *connected*.

37 Graph theory (Math 142/179): The path metric

Definition 37.1. Let G be a connected graph. For vertices v and w of G, we define $d(v, w)$ to the minimum length of any path from v to w.

For the rest of this section, let G be a connected graph, and let v, w, and x be vertices in G.

Theorem 37.2. *The quantity $d(v, w)$ is well-defined.*

Suggestion: What kind of set is guaranteed to have a unique minimum?

Theorem 37.3. *We have that $d(v, w) = 0$ if and only if $v = w$.*

Suggestion: What does a path of length 0 look like?

Theorem 37.4. *We have that $d(v, w) = d(w, v)$.*

Suggestion: How do we turn a path from v to w into a path from w to v?

Theorem 37.5. *We have that $d(v, x) \leq d(v, w) + d(w, x)$.*

Suggestion: How can we take a path from v to w and a path from w to x and make a path from v to x?

38 Graph theory (Math 142/179): Bipartite graphs

Definition 38.1. We say that a graph $G = (V, E, \partial)$ is *bipartite* if V is equal to the disjoint union of sets X and Y such that no vertex in X is adjacent to any other vertex in X, and similarly for Y. (Note that we may think of X and Y as two "colors" for the vertices V, in which case being bipartite means precisely that adjancent vertices have different colors.)

Definition 38.2. A *circuit* in a graph G is a path $(v_0, v_1, \ldots, v_{n-1}, v_n)$ such that $v_0 = v_n$ (i.e., the path starts and ends in the same place).

Theorem 38.3. *If a graph G contains a circuit $(v_0, v_1, \ldots, v_{n-1}, v_n)$ of odd length, then G is not bipartite.*

Suggestion: What are the colors of v_0 and v_1?

Theorem 38.4. *If every circuit in G has even length, then G is bipartite.*

Suggestion: The goal is to pick colors X and Y for all of the vertices of G such that adjancent vertices have different colors. For each path component of G, pick a "home" vertex x and pick a color for x. For any other vertex v in the path component of x, pick a path $(x = v_0, \ldots, v_n = v)$ from x to v, and assign a color to v as forced by the path. Why doesn't this produce an inconsistent color for v (i.e., why is this coloring process well-defined)?

39 Graph theory (Math 142/179): Trees

Definition 39.1. A path $(v_0, v_1, \ldots, v_{n-1}, v_n)$ (Definition 36.2) is said to have a *backtrack* if we can find some k such that $0 \leq k \leq n-2$ and $v_k = v_{k+2}$. A path with no backtracks is said to be *reduced*.

Theorem 39.2. *Let G be a connected graph. For any two vertices v, w in G, there exists a reduced path from v to w.*

Suggestion: At least one path exists from v to w (why?). If this path has backtracks, find a shorter path. Therefore, any shortest path (why must there be a shortest path?) must be reduced.

Definition 39.3. A circuit (Defnition 38.2) of length 0 is said to be *trivial*; a reduced circuit of nonzero length is said to be *nontrivial*.

Definition 39.4. We define a *tree* to be a connected graph with no nontrivial reduced circuits.

Theorem 39.5. *Let G be a connected graph. Then the following are equivalent:*

1. *There is a unique reduced path between any two vertices v, w in G.*
2. *G is a tree (i.e., G has no nontrivial reduced circuits).*

Suggestion for (1) implies (2): Given a nontrivial reduced circuit, find two different reduced paths between two particular vertices.

Suggeston for (2) implies (1): Since G is connected, there must be at least one reduced path from v to w. If there are two different reduced paths, find a reduced circuit in G by looking at the first place the paths are different and the next time they intersect.

40 Linear algebra (Math 129B): Vector spaces

Definition 40.1. A **vector space** is a set V with:

- A binary operation, called vector addition, that defines $\mathbf{v} + \mathbf{w} \in V$ for all $\mathbf{v}, \mathbf{w} \in V$; and
- An operation, called scalar multiplication, that defines $r\mathbf{v} \in V$ for any $r \in \mathbf{R}$ and $\mathbf{v} \in V$;

such that for all $\mathbf{v}, \mathbf{w}, \mathbf{x} \in V$, $r, s \in \mathbf{R}$, the following eight axioms are satisfied:

1. $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.
2. $(\mathbf{v} + \mathbf{w}) + \mathbf{x} = \mathbf{v} + (\mathbf{w} + \mathbf{x})$.
3. There is a fixed vector in V, called $\mathbf{0}$, such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
4. For each $\mathbf{v} \in V$, there exists some $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

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

6. $(r + s)\mathbf{v} = r\mathbf{v} + s\mathbf{v}$.

7. $r(s\mathbf{v}) = (rs)\mathbf{v}$.

8. $1\mathbf{v} = \mathbf{v}$.

In the following, let V be a vector space.

Theorem 40.2. *For any $\mathbf{v} \in V$, $0\mathbf{v} = \mathbf{0}$.*

Suggestion: What is $0\mathbf{v} + 0\mathbf{v}$?

Theorem 40.3. *For any $r \in \mathbf{R}$, $r\mathbf{0} = \mathbf{0}$.*

Suggestion: What is $r\mathbf{0} + r\mathbf{0}$?

Theorem 40.4. *If $r \in \mathbf{R}$, $r \neq 0$, and $r\mathbf{v} = \mathbf{0}$, then $\mathbf{v} = \mathbf{0}$. (In other words, if $r\mathbf{v} = \mathbf{0}$, then either $r = 0$ or $\mathbf{v} = \mathbf{0}$.)*

Suggestion: Divide and use the previous results.

41 Linear algebra (Math 129B): Linear transformations

Definition 41.1. Let V and W be vector spaces (Section 40). We say that a function $T : V \to W$ is a *linear transformation* if, for all $\mathbf{v}, \mathbf{w} \in V$ and $c \in \mathbf{R}$, we have

$$T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}), \qquad T(c\mathbf{v}) = cT(\mathbf{v}).$$

In the rest of this section, let V and W be vector spaces, and let $T : V \to W$ be a linear transformation.

Theorem 41.2. *Let $\mathbf{0}_V$ be the additive identity of V, and let $\mathbf{0}_W$ be the additive identity of W. We have that $T(\mathbf{0}_V) = \mathbf{0}_W$.*

Suggestion: Use the previous section. Alternately, consider $T(\mathbf{0}_V + \mathbf{0}_V)$.

Theorem 41.3. *For any $\mathbf{v} \in V$, we have that $T(-\mathbf{v}) = -T(\mathbf{v})$.*

Suggestion: Since V is an additive abelian group (Section 28), additive inverses are unique (or see Section 25). Show that $T(-\mathbf{v})$ is the additive inverse of $T(\mathbf{v})$.

Theorem 41.4. *For k a positive integer, $c_1, \ldots, c_k \in \mathbf{R}$, and $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$, we have that*

$$T(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + \cdots + c_kT(\mathbf{v}_k).$$

Suggestion: Induction on k.

42 Number theory (Math 126): The Division Algorithm

Theorem 42.1. *If a and b are integers, with $b > 0$, let*

$$S = \{a - bq \mid q \in \mathbb{Z} \text{ and } a - bq \geq 0\}.$$

Then $r = \min S$ exists, and $0 \leq r < b$.

Suggestion: First use the Well-Ordering Principle (why is S nonempty?) to show r exists. Then show that if $r \in S$ and $r \geq b$, then $r \neq \min S$ (i.e., there exists a smaller element of S).

Theorem 42.2 (Division Algorithm)**.** *Let a and b be integers, with $b > 0$. There exist unique integers q and r such that $a = bq + r$ and $0 \leq r < b$.*

Suggestion: The previous theorem shows that at least one pair q, r exists, so it remains to show that q and r are unique. Suppose $a = bq + r = bq' + r'$ with $0 \leq r < b$ and $0 \leq r' < b$.

Note: The Division Algorithm Theorem shows precisely that division with remainder is well-defined; in fact, in elementary terms, q is the quotient and r is the remainder when dividing a by b.

43 Number theory (Math 126): Greatest common divisor

Definition 43.1. If a and n are integers, we say that a *divides* n if $n = ab$ for some integer b.

Definition 43.2. Let a and b be nonzero integers. A *common divisor* of a and b is an integer that divides both a and b. We define $\gcd(a, b)$, or the *greatest common divisor of a and b*, to be the largest positive common divisor of a and b. (Note that 1 is always a common divisor of any two integers, and any common divisor of a and b is no greater than $|a|$, so $\gcd(a, b)$ does exist.)

Theorem 43.3. *Suppose $a, b, x, y, d' \in \mathbb{Z}$. Prove that if $d = ax + by$ and d' divides both a and b, then d' divides d.*

Suggestion: Use the definition of "divides".

Theorem 43.4. *For nonzero integers a and b, there exist integers x and y such that*

$$ax + by = \gcd(a, b).$$

Specifically, if $S = \{ax + by \mid x, y \in \mathbb{Z},\ ax + by > 0\}$, then $\gcd(a, b) = \min S$.

Suggestion: Use the Well-Ordering Principle (why is S nonempty?), and let $d = \min S$. If $a = dq + r$, $0 \leq r < d$ (why is that possible?), then we must have $r = 0$; otherwise, we can find a smaller element of S. A similar argument shows that d also divides b. Finally, by the previous theorem, any common divisor of a and b must divide d, making d the greatest common divisor.

44 Number theory (Math 126): The Euclidean Algorithm

Definition 44.1. Let a and b be positive integers. We choose a sequence of positive integers $r_1, r_2, \ldots, r_k$, by repeatedly applying the Division Algorithm as follows:

$$\begin{aligned}
a &= bq_1 + r_1 && 0 < r_1 < b,\\
b &= r_1 q_2 + r_2 && 0 < r_2 < r_1,\\
r_1 &= r_2 q_3 + r_3 && 0 < r_3 < r_2,\\
&\vdots\\
r_{k-3} &= r_{k-2} q_{k-1} + r_{k-1} && 0 < r_{k-1} < r_{k-2},\\
r_{k-2} &= r_{k-1} q_k + r_k && 0 < r_k < r_{k-1},\\
r_{k-1} &= r_k q_{k+1} + 0.
\end{aligned}$$

Note that the inequality in each step has the form $0 < r_j < r_{j-1}$ and not just $0 \leq r_j < r_{j-1}$, because the process stops as soon as we get $r_k = 0$, which means that the previous remainders are all positive. Note also that decreases by at least 1 each time i increases, which means that the algorithm must stop after a finite number of steps.

Theorem 44.2. *If d divides a, d divides b, and $a = bq + r$, then d divides r. Similarly, if d divides b, d divides r, and $a = bq + r$, then d divides a.*

Suggestion: Use the definition of divides.
For the rest of this section, let a and b be positive integers.

Theorem 44.3. *Let $d = \gcd(a, b)$. Then in the notation of the Euclidean algorithm, d divides r_k.*

Suggestion: Use Theorem 44.2 to explain why d divides r_1, then r_2, and so on.

Theorem 44.4. *Let $d = \gcd(a, b)$. Then in the notation of the Euclidean algorithm, $r_k = d$.*

Suggestion: Use Theorem 44.2 again to explain why r_k divides r_{k-1}, then r_{k-2}, and so on. Thus, d divides r_k and r_k is a common divisor of a and b.

45 Number theory (Math 126): Uniqueness of factorization

Definition 45.1. A *prime* is an integer p such that $p > 1$ and the only positive divisors of p are 1 and p itself.

Theorem 45.2. *Let a and b be positive integers. If p divides ab, then either p divides a or p divides b.*

Suggestion: Suppose p divides ab and p does not divide a. Then $\gcd(a, p) = 1$, so we may apply Theorem 43.4. Multiply both sides by b.

Theorem 45.3 (Uniqueness of factorization)**.** *Suppose n is a positive integer such that*

$$n = p_1 p_2 \dots p_k = q_1 q_2 \dots q_\ell,$$

where all of the p_i and q_j are prime. Then $k = \ell$ (i.e., we have the same number of prime divisors on both sides); moreover, the p_i are precisely the q_j, except possibly numbered in a different order.

Suggestion: Proceed by induction on k. For the induction step, apply the previous theorem with $p = p_k$ to conclude that p_k is the same as one of the primes $q_1, q_2, \dots, q_\ell$. Then divide both sides by p_k and apply the induction hypothesis.

46 Number theory (Math 126): Modular arithmetic

Definition 46.1. Let n be a positive integer. We say that $a \equiv b \pmod{n}$ if n divides $a - b$.

For the rest of this section, fix a positive integer n.

Theorem 46.2. *We have that $a \equiv b \pmod{n}$ if and only if both a and b have the same remainder upon dividing by n (i.e., $a = nq_1 + r$ and $b = nq_2 + r$ for the same r such that $0 \le r < n$).*

Suggestion: Use the Division Algorithm (Section 42).

Theorem 46.3. *If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $a + c \equiv b + d \pmod{n}$.*

Suggestion: Use the $nq + r$ description of being equivalent (mod n). Alternately, use that $a - b = nk$ for some $k \in \mathbb{Z}$.

Theorem 46.4. *If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $ac \equiv bd \pmod{n}$.*

Suggestion: Same ideas as the previous theorem.

47 Number theory (Math 126): Multiplicative functions

Definition 47.1. We say that a function $f : \mathbb{Z}^+ \to \mathbf{R}$ is *multiplicative* if

$$f(mn) = f(m)f(n) \qquad \text{whenever } \gcd(m, n) = 1.$$

Note that when $\gcd(m, n) > 1$, we do not assume that $f(mn) = f(m)f(n)$.

Theorem 47.2. *Let $m = p^a$ and $n = q_1^{b_1} q_2^{b_2} \dots q_k^{b_k}$, where p is prime, the q_i are distinct primes, and $p \neq q_i$ for $1 \le i \le k$. Then $\gcd(m, n) = 1$.*

Suggestion: What are the divisors of m? Can any of them divide n?

Theorem 47.3. *Let $f : \mathbb{Z}^+ \to \mathbf{R}$ be a nonzero multiplicative function. Then $f(1) = 1$.*

Suggestion: What is $\gcd(1, 1)$? Next, prove that if $f(1) = 0$, then $f(n) = 0$ for all $n \in \mathbb{Z}^+$.

Theorem 47.4. *Let $f : \mathbb{Z}^+ \to \mathbf{R}$ be a multiplicative function. If we know the value of $f(p^a)$ for any prime p and any nonnegative integer a, then we can determine the value of $f(n)$ for any positive integer n.*

Suggestion: Factor n as $n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$. Proceed by induction on k.

48 Topology (Math 175): Open and closed sets

Definition 48.1. Let X be a set. A *topology* on X is a collection (set) $\mathcal{T}$ of subsets of X with the following properties:

1. $\emptyset$ and X are in $\mathcal{T}$.
2. If $\{U_\alpha\}$ is a family of elements of $\mathcal{T}$, where α runs over some index set I, then $\bigcup_{\alpha \in I} U_\alpha$ is also an element of $\mathcal{T}$ (i.e., $\mathcal{T}$ is closed under arbitrary union).
3. The intersection of any *finite* subcollection of $\mathcal{T}$ is an element of $\mathcal{T}$ (i.e., $\mathcal{T}$ is closed under finite intersections).

A *topological space* is a set X together with a particular topology on X. If $(X, \mathcal{T})$ (or just X, for short), is a topological space, we call the members of $\mathcal{T}$ the *open subsets* of X. In these terms, the above axioms become:

1. $\emptyset$ and X are open.
2. If $\{U_\alpha\}$ is a family of open sets, where α runs over some index set I, then $\bigcup_{\alpha \in I} U_\alpha$ is also open (i.e., the arbitrary union of open sets is open).
3. The finite intersection of open sets is open.

Definition 48.2. Let X be a topological space. We say that a subset $A \subseteq X$ is *closed* if $X \backslash A$ is open.

Theorem 48.3. *Let X be a topological space. The subsets $\emptyset$ and X are closed.*

Suggestion: Use the definition of closed and open.

Theorem 48.4. *Let X be a topological space. If $\{V_\alpha\}$ is a family of closed sets, where α runs over some index set I, then $\bigcap_{\alpha \in I} V_\alpha$ is also closed.*

Suggestion: Review the set-theoretic laws of complements and intersections.

Theorem 48.5. *Let X be a topological space. If V and W are closed sets, then $V \cup W$ is also closed.*

Suggestion: Same as before.

49 Topology (Math 175): An example

In this section, we fix a set X.

Definition 49.1. The *finite complement topology* on X is defined as follows: We declare that $U \subseteq X$ is open if and only if either $U = X \backslash A$ for some finite subset $A \subseteq X$ or $U = \emptyset$.

We check that this satisfies the three axioms for a topology as follows.

Theorem 49.2. *In the finite complement topology, both $\emptyset$ and X are open.*

Suggestion: By definition.

Theorem 49.3. *Let I be an index set, and for each $\alpha \in I$, let U_α be an open set in the finite complement topology. Then $\bigcup_{\alpha \in I} U_\alpha$ is also open in the finite complement topology.*

Suggestion: What is the union of sets of the form $X \backslash A$? Also, don't forget the empty set.

Theorem 49.4. *Let U and V be open sets in the finite complement topology. Then $U \cup V$ is open in the finite complement topology.*

Suggestion: What is the intersection of $X \backslash A$ and $X \backslash B$?

50 Partially ordered sets: Upsets and downsets

Definition 50.1. A *partially ordered set*, or *poset*, is a set P along with a relation $\leq$ on P that satisfies the following properties:

- **(Reflexive)** For all $x \in P$, $x \leq x$.
- **(Antisymmetric)** For all $x, y \in P$, if $x \leq y$ and $y \leq x$, then $x = y$.
- **(Transitive)** For all $x, y, z \in P$, if $x \leq y$ and $y \leq z$, then $x \leq z$.

For example, any subset of $\mathbf{R}$, along with the usual meaning of $\leq$, is a poset, and the power set of a fixed set X is a poset under the relation that $A \leq B$ if and only if $A \subseteq B$. The first example is *totally ordered*, in that for any $x, y \in \mathbf{R}$, either $x \leq y$ or $y \leq x$, but as for the second example:

Example 50.2. Find a set X and $A, B \subseteq X$ such that $A \not\leq B$ and $B \not\leq A$, under the subset order described above. (This is the reason for the word "partial" in the name "partial order.")

Definition 50.3. Let P be a poset, and let A be a subset of P. If, for every $x \in A$ and every $y \in P$ such that $y \leq x$, we have $y \in A$, we say that A is a *downset* in P. Similarly, if, for every $x \in A$ and every $y \in P$ such that $x \leq y$, we have $y \in A$, we say that A is an *upset* in P. (Downsets are sometimes called *ideals*, and upsets are sometimes called *filters*.)

Theorem 50.4. *Let P be a poset, and let A be a subset of P. If A is a downset of P, then $P \backslash A$ is an upset of P.*

Suggestion: Try contradiction (what if $P \backslash A$ is not an upset?).

51 Preordered sets

Definition 51.1. A *preorder* on a set P is a relation $\preceq$ on P with the following properties:

1. (Reflexive) For $x \in P$, $x \preceq x$.

2. (Transitive) For $x, y, z \in P$, if $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

We call the pair $(P, \preceq)$ a *preordered set.*

Definition 51.2. Let $(P, \preceq)$ be a preordered set. We define a relation $\approx$ on P by saying that, for $x, y \in P$, $x \approx y$ if and only if $x \preceq y$ and $y \preceq x$.

Theorem 51.3. *If $(P, \preceq)$ is a preordered set, then $\approx$ is an equivalence relation on P.*

Definition 51.4. Let $(P, \preceq)$ be a preordered set. For $x \in P$, let $[x]$ (instead of the usual E_x) denote the equivalence class of x under the relation $\approx$, and let $\overline{P} = \{[x] | x \in P\}$, the set of equivalence classes of P under $\approx$. We define a relation $\leq$ on $\overline{P}$ by saying that, for $X, Y \in \overline{P}$, $X \leq Y$ if and only if $x \preceq y$ for some $x \in X$, $y \in Y$.

Recall the definition of poset from Section 50.

Theorem 51.5. *If $(P, \preceq)$ is a preordered set, then $(\overline{P}, \leq)$ is a poset.*

Suggestion: Remember that reflexivity, antisymmetry, and transitivity are to be checked for *equivalence classes* in P. Note that, for example, if $X \leq Y$ and $Y \leq X$, while $x_1 \preceq y_1$ and $y_2 \preceq x_2$ for some $x_1, x_2 \in X$ and $y_1, y_2 \in Y$, we need not have $x_1 = x_2$ or $y_1 = y_2$.

52 Numbers and games: Examples

Definition 52.1. A *game* is a pair of sets $G = (L, R)$, along with a natural number n (called the *day* on which G was created) that satisfy the following (inductive) conditions:

1. The unique game created on day 0 is the game $0 = (\emptyset, \emptyset)$.
2. For any $n > 0$, the games created on day n are precisely the games of the form (L, R), where both L and R are (possibly empty) sets of games created on days 0 through $n - 1$.
3. All games are created on some day $n \in \mathbb{Z}$, $n \geq 0$.

By convention, instead of writing a game $G = (L, R)$ with the usual pair notation, we write $G = \{L|R\}$. Instead of writing L or $R = \emptyset$, we leave the appropriate side of the | blank; for example, we write $0 = \{ \mid \}$.

Example 52.2. For $n \in \mathbb{Z}$, $n \geq 0$, we call the game $\{n| \}$ by the name $n + 1$, and we call the game $\{ |-n\}$ by the name $-(n + 1)$ (where -0 means 0). Prove by induction on n that n and $-n$ actually are games, by the definition given above. (In particular, find the day on which the game n is created.)

Example 52.3. List all games created on days 0 and 1, and list a few games created on day 2. Use the abbreviation $* = \{0|0\}$. (On what day is $*$ created?)

53 Numbers and games: Ordering

Convention 53.1. If $x = \{L|R\}$ is a game, we use x^L to denote an arbitrary element of L and x^R to denote an arbitrary element of R. For example, instead of saying "For all $x^L \in L$ and $x^R \in R$", we simply say "For all x^L and x^R", without having to give names to the sets L and R.

Definition 53.2. We inductively define a relation $\preceq$ on the set of all games by saying that $x \succeq y$ if and only if for all x^R and y^L, $x^R \not\preceq y$ and $x \not\preceq y^L$, and that $x \preceq y$ if and only if $y \succeq x$. (Note that the fact that $\preceq$ is well-defined is a double induction on the creation days of x and y!)

Theorem 53.3. *For all games x, we have that:*

1. *For all x^R, we have $x \not\succeq x^R$;*
2. *For all x^L, we have $x^L \not\succeq x$; and*
3. *$x \preceq x$.*

Suggestion: Prove all three claims simultaneously by induction on the day on which x is created.

Theorem 53.4. *The relation $\preceq$ is transitive, i.e., for games x, y, z, if $x \preceq y$ and $y \preceq z$, then $x \preceq z$.*

Suggestion: Use the previous theorem, and proceed by triple induction on the creation days of x, y, and z. (I.e., the induction assumption is that the theorem is true for x^R, y, and z, as x^R was created on an earlier day than x; for x, y^L, and z, as x^R was created on an earlier day than x; and so on.) Note that the conclusion $x \preceq z$ is naturally proved by contradiction, as by definition, if $x \not\preceq z$, one of two things must occur.

Definition 53.5. We say that two games x and y are *equal* if and only if $x \preceq y$ and $y \preceq x$. We define a relation $\leq$ on equivalence classes of games under this notion of equality, by declaring that, if X and Y are equivalence classes, then $X \leq Y$ if and only if $x \preceq y$ for some $x \in X$, $y \in Y$.

It follows from the results of Section 51 that equvalence classes of games are a poset under $\leq$.

54 Numbers and games: Surreal numbers

Definition 54.1. We define a (surreal) *number* to be a game (see Section 52) x such that for any x^L and x^R (see Section 53 for notation), x^L and x^R are numbers, and $x^L \not\succeq x^R$ (see Section 53 for the definition of $\succeq$).

Theorem 54.2. *The games n ($n \in \mathbb{Z}$) defined in Section 52 are numbers.*

Suggestion: Proceed by induction on n. Apply this both to the game n and the game $-n$.

Definition 54.3. Following Section 53, for games x, y, we say that $x \prec y$ if $x \preceq y$ and $x \not\succeq y$.

Theorem 54.4. *If x is a number, then for all x^L and x^R, $x^L \prec x \prec x^R$.*

Suggestion: Proceed by induction on the creation day of x (see Section 52). Note that by the results of Section 53, it suffices to show that $x^L \preceq x$ and $x \preceq x^R$, each of which can be proven by contradiction.

Theorem 54.5. *If x and y are numbers, then either $x \succeq y$ or $y \succeq x$.*

Suggestion: Assume $x \not\succeq y$; in each of two cases, conclude that $x \prec y$.

Remark: One can go on to define addition, subtraction, and multiplication on the surreal numbers so that, as we have defined them here, the surreal numbers work exactly like the usual *dyadic rationals*, that is, rational numbers whose denominators are nonnegative powers of 2. In fact, if we allow for numbers created not only on finite days, but also the "infinite day" ω, we can recover not only the real numbers, but also more exotic surreal numbers like ω (a surreal number greater than any rational) and its counterpart $\dfrac{1}{\omega}$ (a positive surreal number smaller than any rational).

For more on surreal numbers, see *Surreal Numbers*, by Donald Knuth. To see what this all has to do with games, see *Winning Ways*, by Elwyn Berlekamp, John H. Conway, and Richard Guy. For a high-level (upper-level undergraduate or graduate) account, see *On Numbers and Games*, by John H. Conway.