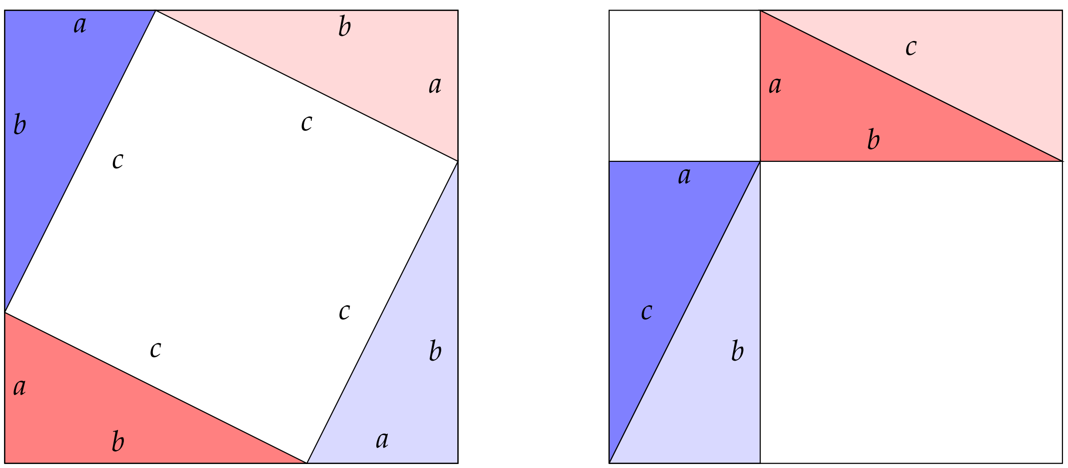
Math 13 — An Introduction to Abstract Mathematics

Neil Donaldson and Alessandra Pantano

With contributions from:
Michael Hehmann, Christopher Davis, Liam Hardiman, and Ari Rosenfield

September 7, 2023



Contents

Useful Texts

- *Book of Proof*, Richard Hammack, 2nd ed 2013. Available free online! Very good on the basics: if you're having trouble with reading set notation or how to construct a proof, this book's for you! These notes are *deliberately* pitched at a high level relative to this textbook to provide contrast.
- *Mathematical Reasoning*, Ted Sundstrom, 2nd ed 2014. Available free online! Excellent resource. If you would like to buy the actual book, you can purchase it on Amazon at a really cheap price.
- *Mathematical Proofs: A Transition to Advanced Mathematics*, Chartrand/Polimeni/Zhang, 3rd Ed 2013, Pearson. The most recent course text. Has many, many exercises; the first half is fairly straightforward while the second half is much more complex and dauntingly detailed.
- *The Elements of Advanced Mathematics*, Steven G. Krantz, 2nd ed 2002, Chapman & Hall and *Foundations of Higher Mathematics*, Peter Fletcher and C. Wayne Patty, 3th ed 2000, Brooks–Cole are old course textbooks for Math 13. Both are readable and concise with good exercises.

Learning Outcomes

1. Developing the skills necessary to read and practice abstract mathematics.
2. Understanding the concept of proof, and becoming acquainted with several proof techniques.
3. Learning what sort of questions mathematicians ask, what excites them, and what they are looking for.
4. Introducing upper-division mathematics by giving a taste of what is covered in several areas of the subject.

Along the way you will learn new techniques and concepts. For example:

Number Theory Five people each take the same number of candies from a jar. Then a group of seven does the same. The, now empty, jar originally contained 239 candies. Can you decide how much candy each person took?

Geometry and Topology How can we visualize and compute with objects like the Mobius strip?

Fractals How to use sequences of sets to produce objects that appear the same at all scales.

To Infinity and Beyond! Why are some infinities greater than others?

1 A Paradigm Shift

1.1 Proof

How can you convince someone that you're telling the truth? Depending on the situation, there are varying amounts of show and tell. If you're a defense attorney, you tell a story to the jury, backed up with evidence and witness testimony, that instills some reasonable doubt that your client committed the crime. If you're a physicist or chemist, you run experiments and collect statistics that are sufficiently close to your predicted values. If you're an investigative journalist, you piece together excerpts from primary sources and interviews with key people to construct a consistent narrative that's as complete as possible.

Critically, the attorney's client doesn't actually need to be innocent – the jury just needs to doubt their guilt. The scientist's results might be consistent with her predictions, but it might not be for the reason she purports. The journalist may have neglected to interview an important, but peripheral person of interest or a condemning document may have been destroyed that would otherwise completely upend their story. Even when all three of these characters act with the best intentions, there's still a gap between what they present to the world and the underlying fact of the matter. Of course, there are reasons these disciplines have differing standards of truth. These could be liberal values concerning justice, the physical limitations of our measurements, or the constraints of time, money and access to information.

Let's compare and contrast these examples with what it means to convince someone in mathematics. Up to this point, the only convincing you've probably had to do is showing that your calculations are correct. You'd start with some given expression like an integral and then jump from equals sign to equals sign using algebraic manipulations and identities. Eventually, you'd get some number that you'd box and say "there's my answer." If asked for justification, you'd point to your chain of equalities and say "there it is."

But sometimes we want to use reasoning that isn't just algebra. You might have gotten a taste of this in your high school geometry class where you had to write *proofs*. Rather than give a technical definition of what a proof is, let's look at an example of what a proof can be.

Theorem 1.1. *If $a \leq b \leq c$ are the side lengths of a right triangle, then $a^2 + b^2 = c^2$.*

Proof. Consider a square with side length $a + b$ with four congruent copies of our triangle arranged as in Figure 2. The total area of the square is $(a + b)^2$, but if we add up the areas of the four triangles and the square in the middle (try to prove that this is indeed a square!), we get $2ab + c^2$.

Now let's slide the lower-left triangle up and to the right so that it forms a rectangle with the upper-right triangle. Then slide the upper-left triangle down and the lower-right triangle to the left so they form a rectangle as well. If we add up the areas of the two rectangles and the two smaller squares, we get $2ab + a^2 + b^2$.

But the area of the whole square hasn't changed, so we must have that

$$2ab + c^2 = 2ab + a^2 + b^2.$$

When we subtract $2ab$ from both sides, we arrive at the desired conclusion.

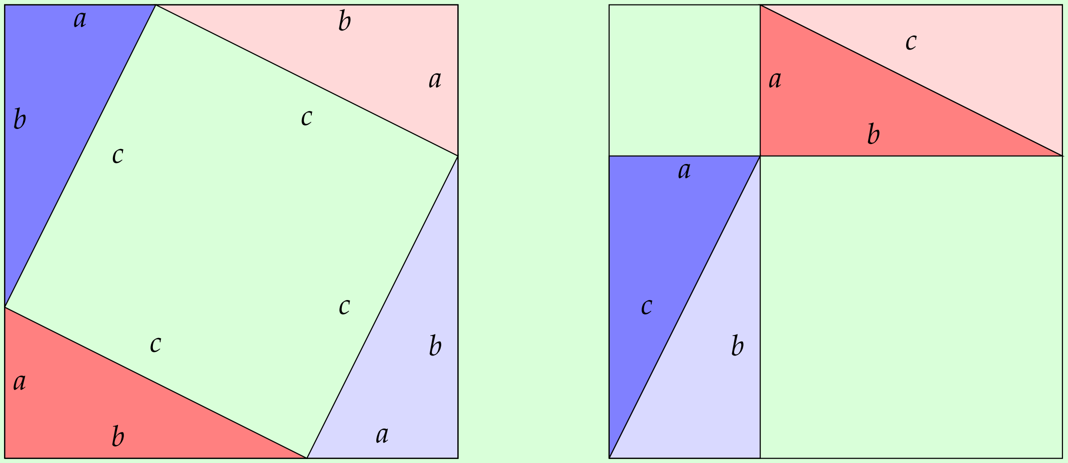


Figure 1: Square of side length $a + b$.

■

Discussion Questions

1. Are you convinced by this proof? If so, why? If not, what feels off?
2. Would this proof work without the picture?

Even though we finished this proof with an equation (the theorem concerns an equation after all), the *essence* of the proof is in the picture. The picture informed which equations we wrote down. That being said, we shouldn't get too attached to the picture. Are all four angles in our square *really* right angles? Are the four triangles *exactly* congruent to each other? If it turned out that one of the line segments we drew wasn't perfectly straight, would we need to scrap the whole argument? What if we want to apply this reasoning to a right triangle with proportions very different from the ones we drew here?

The point here is that the picture isn't a proof in and of itself. Rather, it serves to remind us of our assumptions and their consequences.

1.2 Definition

Mathematicians spend a lot of time proving theorems – stating truths and providing justification. True statements about what? Well Theorem 1.1 seems to be a statement about right triangles. What are those? Whatever it is, it had better satisfy the following *definition*, a set of unambiguous and clear conditions.

Definition 1.2. A right triangle is a triangle, two of whose sides meet at a right angle.

It seems like we've cheated a bit here. This doesn't make any sense until we've also defined "triangle", "right angle", and "side" as well. How deep does the rabbit hole go? Well a triangle is a polygon with three sides and a right angle is exactly half of a straight angle. Now it looks like we have to define "polygon" and "straight angle." We can keep defining objects in terms of other ideas, but at some point we have to stop and say "you know what I mean.[1]"

In a way, definitions serve as the foundation upon which we construct buildings – theorems in this analogy. Let's walk through some other examples, this time from number theory.

Definition 1.3. An integer is even if it is divisible by two.

For now we'll take "the integers" to be the set of positive and negative whole numbers $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$. If we really wanted to, we could come up with definitions for 0, 1, and so on, but we'll take these for granted here.[2] Assuming we know what "integer" and "two" are, it looks like we need to define "divisible." This shouldn't give us too much trouble, assuming that we know how to do basic arithmetic like addition and multiplication of integers.

Definition 1.4. An integer a is divisible by the integer b if $a = bc$ for some integer c.

So for example, 2 itself is even since $2 = 2 \cdot 1$. So is $-50 = 2 \cdot (-25)$.

Only once we have our definitions straight can we start proving things. For example, consider the following theorem:

Theorem 1.5. *The sum of any two even integers is even.*

The proof of this theorem flows straight from the definition.

Proof. Let x and y be *any* two even integers. We want to show that $x + y$ is an even integer.
By definition, an integer is even if it can be written in the form $2k$ for some integer k. Thus there exist integers n, m such that $x = 2m$ and $y = 2n$. We compute:

$$x + y = 2m + 2n = 2(m + n). \tag{$*$}$$

Because $m + n$ is an integer, this shows that $x + y$ is an even integer. ∎

There are several important observations:

- 'Any' in the statement of the theorem means the proof must work *regardless* of what even integers you choose. It is not good enough to simply select, for example, 4 and 16, then write $4 + 16 = 20$. This is an *example*, or test, of the theorem, not a mathematical proof.
- According to the definition, $2m$ and $2n$ together represent *all possible pairs* of even numbers.

[1] The ancient Greek geometer Euclid left some pretty fundamental terms like "point" and "line" undefined by modern standards. Instead of saying "you know what I mean," a point was "that which has no part" and a line was "a length without width."

[2] The use of $\mathbb{Z}$ for the set of integers comes from the German word for "number", "zahlen". The branch of math that would be concerned with defining the integers in terms of more primitive objects would be mathematical logic and set theory.

- The proof makes direct reference to the definition. The vast majority of the proofs in this course are of this type. If you know the definition of every word in the statement of a theorem, you will often discover a proof simply by writing down the definitions.
- The theorem itself did not mention any *variables.* The proof required a calculation for which these were essential. In this case the variables m and n come for free *once you write the definition of evenness!* A great mistake is to think that the proof is nothing more than the calculation $(*)$. This is the easy bit, and it means nothing without the surrounding sentences.

The important link between theorems and definitions is much of what learning higher-level mathematics is about. We prove theorems (and solve homework problems) because they make us use and understand the subtleties of definitions. One does not *know* mathematics, one *does* it. Mathematics is a *practice;* an art as much as it is a science.

Exercises

1.2.1 Consider the following definitions.

Definition. A function from a set A to a set B is called "strictly increasing" if for all $x_1 < x_2$ in the domain A, $f(x_1) < f(x_2)$.

Definition. A function from a set A to a set B is called "strictly decreasing" if for all $x_1 < x_2$ in the domain A, $f(x_1) > f(x_2)$.

Give an example of a function which is neither strictly increasing nor strictly decreasing. Explain your answer.

1.2.2 Consider the following definition.

Definition. A sequence $\{a_n\}$ "goes to ∞" if for all number M, there exist a natural number N, such that $a_n > M$ for all $n > N$.

Test your understanding of the definition by creating some examples and some non-examples. Why do your non-examples fail the definition?

1.2.3 Let's say an integer to be "nearly a square" if it is of the form:

$$2 = 2 \cdot 1, \quad 6 = 3 \cdot 2, \quad 12 = 4 \cdot 3, \quad 20 = 5 \cdot 4, \quad 30 = 6 \cdot 5, \quad 42 = 7 \cdot 6, \ldots$$

Give a (formal) definition for a "nearly a square" integer.

1.2.4 The following are <u>incorrect</u> ways of writing the definition of an **odd integer**. Explain why each fails to properly define an odd integer n.

An integer n is odd if

(a) $n = 2k + 1$ for every integer k.
(b) $n = 2k + 1$ for some number k.
(c) there is some integer k such that $k = 2n + 1$.

1.2.5 Consider the following definition.

Definition: Let n, m be integers. We say n is "divisible by m" if $n = mk$ for some integer k.

Explore using some examples. Then, for each of the following statements, explain why the statement is true or false.

(i) If the last digit of an integer is divisible by 4, then the integer is divisible by 4.

(ii) If the last digit of an integer is divisible by 2, then the integer is divisible by 2.

1.3 Theorem and Conjecture

Theorems are true mathematical statements that we can prove. These include important and widely applicable "named theorems" like the Pythagorean theorem, the fundamental theorem of calculus, and the rank-nullity theorem as well as simpler ones like Theorem 1.5. Essentially, if you can prove it, then it's a theorem.

But what if we're confronted with a statement that we don't know how to prove? It would be a bit arrogant of us to conclude that the statement is false just because we don't know how to prove it. Statements that we believe to be true, but don't yet know how to prove are called *conjectures*. Conjectures push the boundaries of math as we know it. Much of the creativity in mathematics comes from the pursuit of proving or disproving conjectures. When you think about it, we can't lose – we learn something new if we prove the conjecture true *or* false.

A conjecture is the mathematician's equivalent of the experimental scientist's hypothesis: a statement that one would like to be true. The difference lies in what comes next. The mathematician will try to prove that a conjecture is undeniably true by relying on logic, while the scientist will apply the scientific method, conducting experiments attempting, and hopefully failing, to show that a hypothesis is incorrect. To get a taste, consider the following.

Conjecture 1.6. *If n is any odd integer, then $n^2 - 1$ is a multiple of 8.*

Conjecture 1.7. *For every positive integer n, the integer $n^2 + n + 41$ is prime.*

Once a mathematician proves the validity of a conjecture it becomes a *theorem.* The job of a mathematics researcher is thus to formulate conjectures, prove them, and publish the resulting theorems. The creativity lies as much in the formulation as in the proof. As you go through the class, try to formulate conjectures. Like as not, many of your conjectures will be false, but you'll gain a lot from trying to form them.

Let us return to our conjectures: are they true or false? How can we decide? As a first attempt, we may try to test the conjectures by computing with some small integers n. In practice this would be done *before* stating the conjectures.

n	1	3	5	7	9	11	13
$n^2 - 1$	0	8	24	48	80	120	168

n	1	2	3	4	5	6	7
$n^2 + n + 41$	43	47	53	61	71	83	97

Because 0, 8, 24, 48, 80, 120 and 168 are all multiples of 8, and 43, 47, 53, 61, 71, 83 and 97 are all prime, both conjectures appear to be true. Would you bet $100 that this is indeed the case? Is $n^2 - 1$ a multiple of 8 *for every* odd integer n? Is $n^2 + n + 41$ prime *for every* positive integer n? The only way to establish whether a conjecture is true or false is by doing one of the following:

Prove it by showing it must be true in <u>all</u> cases, or,

Disprove it by finding <u>at least</u> one instance in which the conjecture is false.

Let us work with Conjecture 1.6. If n is an odd integer, then, by definition, we can write it as $n = 2k + 1$ for some integer k. Then

$$n^2 - 1 = (2k+1)^2 - 1 = (4k^2 + 1 + 4k) - 1 = 4k^2 + 4k.$$

We need to investigate whether this is *always* a multiple of 8. Since

$$4k^2 + 4k = 4(k^2 + k)$$

is already a multiple of 4, it all comes down to deciding whether or not $k^2 + k$ contains a factor 2 for all possible choices of k; i.e. is $k^2 + k$ even? Do we believe this? We can return to trying out some small values of k:

k	-2	-1	0	1	2	3	4
$k^2 + k$	2	0	0	2	6	12	20

Once again, the claim seems to be true for small values of k, but how do we know it is true for *all* k? Again, the only way is to *prove it* or *disprove it*. How to proceed? The question here is whether or not $k^2 + k$ is *always* even. Factoring out k, we get:

$$k^2 + k = k(k+1).$$

We have therefore expressed $k^2 + k$ as a product of two consecutive integers. This is great, because for any two consecutive integers, one is even and the other is odd, and so their product must be even. We have now proved that the conjecture is true. Conjecture 1.6 is indeed a *theorem!* Everything we've done so far has been investigative, and is laid out in an untidy way. We don't want the reader to have to wade through all of our scratch work, so we formalize the above argument. This is the final result of our deliberations; investigate, spot a pattern, conjecture, prove, and finally present your work in as clean and convincing a manner as you can.

Theorem 1.8. *If n is any odd integer, then $n^2 - 1$ is a multiple of 8.*

Proof. Let n be any odd integer; we want to show that $n^2 - 1$ is a multiple of 8. By the definition of *odd* integer, we may write $n = 2k + 1$ for some integer k. Then

$$n^2 - 1 = (2k+1)^2 - 1 = (4k^2 + 1 + 4k) - 1 = 4k^2 + 4k = 4k(k+1).$$

We distinguish two cases. If k is even, then $k(k+1)$ is even and so $4k(k+1)$ is divisible by 8.
If k is odd, then $k+1$ is even. Therefore $k(k+1)$ is again even and $4k(k+1)$ divisible by 8.

In both cases $n^2 - 1 = 4k(k+1)$ is divisible by 8. This concludes the proof. ■

It is now time to explore Conjecture 1.7. The question here is whether or not $n^2 + n + 41$ is a prime integer for *every* positive integer n. We know that when $n = 1, 2, 3, 4, 5, 6$ or 7 the answer is yes, but examples do not make a proof. At this point, we do not know whether the conjecture is true or false. Let us investigate the question further. Suppose that n is any positive integer; we must ask whether it is possible to factor $n^2 + n + 41$ as a product of two positive integers, neither of which is one.[3] When $n = 41$ such a factorization certainly exists, since we can write

$$41^2 + 41 + 41 = 41(41 + 1 + 1) = 41 \cdot 43.$$

Our *counterexample* shows that there exists at least one value of n for which $n^2 + n + 41$ is *not* prime. We have therefore disproved the conjecture that 'for all positive integers n, $n^2 + n + 41$ is prime,' and so Conjecture 1.7 is false!

The moral of the story is this: to show that a conjecture is true you must prove that it holds for all the cases in consideration, but to show that it is false a single counterexample suffices.

Conjectures: *True* or *False*?

Do your best to prove or disprove the following conjectures. Then revisit these problems at the end of the course to realize how much your proof skills have improved.

1. The sum of any three consecutive integers is even.
2. There exist integers m and n such that $7m + 5n = 4$.
3. Every common multiple of 6 and 10 is divisible by 60.
4. There exist integers x and y such that $6x + 9y = 10$.
5. For every positive real number x, $x + \frac{1}{x}$ is greater than or equal to 2.
6. If x is any real number, then $x^2 \geq x$.
7. If n is any integer, $n^2 + 5n$ must be even.
8. If x is any real number, then $|x| \geq -x$.
9. Consider the set $\mathbb{R}$ of all real numbers. For all x in $\mathbb{R}$, there exists y in $\mathbb{R}$ such that $x < y$.
10. Consider the set $\mathbb{R}$ of all real numbers. There exists x in $\mathbb{R}$ such that, for all y in $\mathbb{R}$, $x < y$.
11. The sets $A = \{n \in \mathbb{N} : n^2 < 25\}$ and $B = \{n^2 : n \in \mathbb{N} \text{ and } n < 5\}$ are equal. Here $\mathbb{N}$ denotes the set of natural numbers.

Now we know a little of what mathematics is about, it is time to practice some of it!

[3]Once again we rely on a definition: a positive integer is *prime* if it cannot be written as a product of two integers, both greater than one.

Discussion Questions

1. An integer is odd if it isn't even. Prove that an integer a is odd if and only if $a = 2b + 1$ for some integer b. That is, prove that all odd integers can be written this way, and that if an integer can be written this way, then it is odd.

2. When you do math, do you feel like you're engaging with the creative side of your personality?

3. Try to come up with some examples for each of the following definitions. Are these definitions problematic in some way? If so, try to fix the definition.

Definition 1.9. A rational number a/b is even if a and b are both even integers.

Definition 1.10. We say that an integer a is *very odd* if $a = bc$ for odd integers b and c.

Exercises

1.3.1 A rational number, roughly speaking, is what we think of as a fraction.

Definition. A real number r is a "rational number" if $r = \frac{p}{q}$ for some integer p and for some non-zero integer q.

(a) Are integers rational numbers?
(b) True or false: Each rational number r has a non-zero integer n so that rn is an integer.

1.3.2 Consider the following definitions.

Definition. An integer is "super-odd" if it is the product of an odd integer times itself (that is, it is the square of an odd integer).

Definition. An integer is "super-even" if it is the product of an even integer times itself (that is, it is the square of an even integer).

Using the definitions of the super-odd and super-even integers, come up with a conjecture of your own! Can you prove or disprove your conjecture?

1.3.3 Decide whether the following conjecture is true or false, and justify your reasoning.

Conjecture. *There is a smallest positive real number.*

1.4 Planning and Writing a Proof

Your main responsibility for the rest of this course is to write proofs. If you look back at any of the conjectures on page 10, there might be a question on your mind – *how do I prove (or disprove) this?* If you read a proof (or disproof) written by someone else, you might have the related question – *how did they ever come up with this proof?*

The source of a proof is often less magical than it appears: usually the original author of the proof experimented until they found something that worked. Most of that experimentation gets hidden in the final written proof. The proof itself should be written in the way that is easiest to read.

In order to bridge the gap between what goes into thinking about a proof and what the proof actually looks like, we recommend splitting the proof writing process into the following four steps.

Interpret Make sense of the statement. What is the proposition saying? Can you rephrase the claim in a way that is more clear to you? What are you trying to prove and what are you assuming? The most important part of this step is identifying the *logical structure* of the statement you're trying to prove. At this stage, you might ignore the particular details of the statement and instead focus on its logical meaning. Depending on the statement, it might be helpful to rephrase the statement as an if-then statement.

Brainstorm Now that you understand what the statement is saying, convince *yourself* that the statement is true. First, look up the relevant definitions.

Next, think of some instances where the conditions of this proposition are met. Try out some examples, and ask yourself what makes the claim work in those instances. Looking carefully at examples may help build intuition about *why* the claim is true and suggest a strategy to prove it.

You'll also want to review other theorems that relate to these definitions. What theorems involve the same definitions? Do you know any theorems that relate your assumptions to the conclusion? Have you seen a proof for a similar statement before?

This stage is a little like packing for a trip. Imagine laying out all the stuff you might want to bring, looking to get an overview of what you have available. Even if you don't end up bringing that fourth pair of shoes, maybe it was helpful to remind yourself of all the options.

Sketch This is the phase where you build the skeleton of your proof. Think again about what you are assuming and what you are are you trying to prove. What should the *first* step of the proof be? What about the *last* step? Write down some informal arguments for yourself to connect the first and last steps of your proof, feeling free to use shorthand. If you get stuck, try a different approach. Scribble out some drawings and draw as many arrows as you want to connect your ideas.

This *sketch* step is where most of the thinking happens and it will most likely be the longest step in the proof-writing process. This is also the step where you will be doing most of your calculations. You can be as messy as you want in this step because *you* should be the only one who ever reads this. Once you've learned a variety of different proof methods, this is a good stage at which to experiment with them and try to find which proof method will work best.

Prove Once you have proven the statement to yourself, it's time to prove the statement to the world. Now is when you go back through your proof sketch and translate it into a linear story, written in complete sentences. Here you should carefully word your explanations and you should avoid using shorthand. The result should be a clear, formal proof like the ones you read in a mathematics textbook, where symbols like $\Longrightarrow$ or $\therefore$ are replaced by words like "hence" or "therefore". Although you are providing a mathematical argument, your proof should read like prose.

Review Finally, you should *review* your proof. Assume that the reader is looking at the problem for the first time and has not read your sketch. Read your own proof with some skepticism and consider its readability and flow. Get read of unnecessary claims and revise the wording if necessary.

During this review step, try also reading your proof out loud. If you find yourself adding extra words that aren't written down, then include those words in the proof.

Let's use this framework to prove the following theorem.

Theorem 1.11. *A positive integer n is divisible by 5 when its last digit is 5.*

First, we need to make sense of the statement of the theorem.

Interpret. What is the logical structure of this statement we are trying to prove? In this case, it has deliberately been written in a less-than-straightforward way. Don't just read the words in the statement, but read it slowly and carefully enough that you understand what it is asserting. Once you understand what this sentence means, it is possible to translate it into an if-then statement.

- *Equivalent statement*: Let n be a positive integer. If the last digit of n is 5, then n is divisible by 5.

■

Not every statement can be rewritten as an if-then statement, but many can, including this statement we are trying to prove. Before moving on, be sure you agree that our statement really means the same thing as what we wrote above. Understanding the English meaning of the statement was essential to writing it as an if-then statement.

It's hopeless to try to memorize all the possibilities (in general, this course will involve less memorization than your earlier math classes). Instead of memorizing that a statement with the logical structure "P when Q" is the same as "if Q then P", you should instead think about the natural English meaning of the sentences, and make sure they are equivalent. (It might be helpful to try an example sentence that doesn't involve math; can you think of one in this case, using the structure "P when Q"?)

Now we move on to the brainstorm phase, where we gather examples and definitions related to our theorem.

Brainstorm. A key practice is to write out the relevant definitions. We are already comfortable with the notion of *divisibility*, but the precise notion of *digits* is probably less familiar.

- Definition of "divisible by 5": An integer n is *divisible* by 5 if there exists an integer k such that $n = 5k$. That is, n is a multiple of 5.
- Meaning of "digit": Each digit of n is a number between 0 and 9, representing how many 1's, 10's, 100's, etc. are "in" n. (For example, the number 671 has six hundreds, seven tens, and a one. That is, $671 = 6 \cdot 100 + 7 \cdot 10 + 1$.)

  More formally, if n is a k-digit number and its digits are $n_0, n_1, \ldots, n_{k-1}$, then

  $$n = n_{k-1} \cdot 10^{k-1} + n_{k-2} \cdot 10^{k-2} + \cdots + n_1 \cdot 10 + n_0. \tag{1.1}$$

  Thus, "last digit 5" means $n_0 = 5$.

Do we believe this statement is true? Let's try out some examples.

- $15 = 3 \cdot 5$ is divisible by 5.

- $245 = 49 \cdot 5$ is divisible by 5.

The theorem claims that *every* positive integer ending in 5 is a multiple of 5, or, as we rephrased it, *if* a positive integer ends in 5, *then* it is divisible by 5. How important is that assumption, that the last digit is 5? Here are some more examples.

- 13 is not divisible by 5.

- $20 = 4 \cdot 5$ is divisible by 5.

From these examples, we notice that not all positive integers are divisible by 5 (okay, we probably already knew that), and that some positive integers which do not end in 5 are divisible by 5 (e.g., 20).

■

The brainstorming phase is also a good time to write out relevant theorems. In this case, I don't think we know any theorems that will help us, but as we learn more theorems, we will have more options available to us.

For example, imagine we knew a theorem that said, "A number is divisible by 5 if and only if its last digit is 0 or 5." That theorem would instantly imply what we're trying to prove (and more). (**Warning!** If you stumble upon a one-sentence proof like this on an exam, probably your instructor doesn't want you using that theorem, and at the very least you should ask to confirm whether or not it's okay to use.)

As another example, imagine we knew a theorem that said, "If the last digit of a positive integer is 6, then that positive integer is divisible by 2." This theorem itself wouldn't help us prove our result, but its proof would very likely be helpful.

Next we have the *sketch* portion of the proof. Let's revisit our examples and experiment to try to discover *why* this statement is true. We will make use of our description above of what it means to have last digit equal to 5.

Our final proof needs to be written in complete sentences, but since the sketch portion is just for us, we can be less formal. Here is how a sketch of this proof might look.

Sketch. Our examples:

- $15 = 1 \cdot 10 + 5 = 1 \cdot \underbrace{2 \cdot 5}_{10} + 5 = 5 \cdot (2 + 1)$ (a multiple of 5).

- $245 = 2 \cdot 10^2 + 4 \cdot 10^1 + 5 = 2 \cdot 10 \cdot \underbrace{2 \cdot 5}_{10} + 4 \cdot \underbrace{2 \cdot 5}_{10} + 5 = 5 \cdot (40 + 8 + 1)$ (a multiple of 5).

(Commentary, not part of the sketch.

The key idea here is that we can pull out a 5 from each power of 10 and from the last digit (which is a 5). This will be our general strategy.

We now explore the general case, using the strategy suggested by the previous examples. At this stage, we focus on outlining the main ideas, and do not bother about writing down a formal proof. The bullet points here aren't essential; they are just there to make it a little easier for you to read.

End of commentary, back to the sketch.)

- $n = n_{k-1} \cdot 10^{k-1} + n_{k-2} \cdot 10^{k-2} + \cdots + n_1 \cdot 10 + 5 \quad$ (last digit 5)
- 10, 10^2, 10^3, ... are all multiples of 5.
- n_1, n_2, n_3, ... are not multiples of 5 (in general), but this is irrelevant to us.
- $n_{k-1} \cdot 10^{k-1} + n_{k-2} \cdot 10^{k-2} + \cdots + n_1 \cdot 10$ is a multiple of 5.
- n is a multiple of 5 because

$$n = \underbrace{n_{k-1} \cdot 10^{k-1} + n_{k-2} \cdot 10^{k-2} + \cdots + n_1 \cdot 10}_{5\cdot\text{ (an integer)}} + \underbrace{5}_{5\cdot 1} \quad \text{(last digit 5)}$$

- So 5 divides n.

■

The authors had a difficult time writing this "sketch" portion, because, on one hand, we want to explain what we're thinking. On the other hand, we want to show you what a sketch looks like in real life. The sketch is usually quite messy, and is nearly impossible for anyone else to follow. In fact, a common struggle for professional mathematicians is returning to a sketch they wrote weeks or months ago, and trying to figure out what their scratchwork meant.

Next, let's turn this sketch into a formal proof that a classmate would be able to follow. There is a certain writing style used in most formal mathematical proofs. We will use that style in the following, and you will get more used to it in the more proofs you read (and write).

Prove. We need to show that if the positive integer n has last digit 5, then n is divisible by 5. Write n in its base-10 expansion:

$$\begin{aligned} n &= n_{k-1} \cdot 10^{k-1} + n_{k-2} \cdot 10^{k-2} + \cdots + n_1 \cdot 10 + n_0 \\ &= n_{k-1} \cdot 10^{k-1} + n_{k-2} \cdot 10^{k-2} + \cdots + n_1 \cdot 10 + 5. \end{aligned}$$

Note that for all $j \geq 1$, 10^j is a multiple of 5. Hence, for all $j = 1, \ldots, k-1$, there exists an integer a_j such that we can write $10^j = a_j \cdot 5$. Thus, we get:

$$n = n_{k-1} \cdot (a_{k-1} \cdot 5) + n_{k-2} \cdot (a_{k-2} \cdot 5) + \cdots + n_1 \cdot (a_1 \cdot 5) + 5$$

(by associativity of multiplication)

$$= (n_{k-1} \cdot a_{k-1}) \cdot 5 + (n_{k-2} \cdot a_{k-2}) \cdot 5 + \cdots + (n_1 \cdot a_1) \cdot 5 + 1 \cdot 5$$

(by distributivity)

$$= (n_{k-1} \cdot a_{k-1} + n_{k-2} \cdot a_{k-2} + \cdots + n_1 \cdot a_1 + 1) \cdot 5.$$

Because the quantity $n_{k-1} \cdot a_{k-1} + n_{k-2} \cdot a_{k-2} + \cdots + n_1 \cdot a_1 + 1$ is an integer, this shows that n is a multiple of 5. Hence, n is divisible by 5. This completes the proof. ∎

Finally, we review the proof and try to make it more readable. At the beginning of your career as a proof writer, you will be required to justify every claim in your proof (hence, for example, the reference to associativity above). You may think we are being too picky, but this initial requirement is just an attempt to get you in the habit of explaining your claims. As you gain more practice with proof writing, you will be expected to include fewer details in your proof and will even be allowed to skip some simple steps that most readers may consider "trivial" (because they believe they are true without feeling the need for an explanation). Thus, a more experienced proof-writer may review the proof above and present a more compact version, like the one offered below.

Review. We need to show that if the positive integer n has last digit 5, then n is divisible by 5. Write n in its base-10 expansion:

$$\begin{aligned} n &= n_{k-1} \cdot 10^{k-1} + n_{k-2} \cdot 10^{k-2} + \cdots + n_1 \cdot 10 + n_0 \\ &= n_{k-1} \cdot 10^{k-1} + n_{k-2} \cdot 10^{k-2} + \cdots + n_1 \cdot 10 + 5. \end{aligned}$$

Since 10^j is a multiple of 5 for all $j \geq 1$, the sum of the first $k-1$ terms on the right-hand side is divisible by 5 (i.e., the terms from $n_{k-1} \cdot 10^{k-1}$ to $n_1 \cdot 10$). Of course, 5 is also divisible by 5. Thus, their sum n is also divisible by 5. This completes the proof. ∎

Some comments:

1. Whether in the longer version of the proof, or the shorter version of the proof, we didn't fully justify that the sum of $k-1$ terms which are divisible by 5 is divisible by 5. In fact, at this stage, we may not even have one of the necessary tools (mathematical induction) to make that claim 100% rigorous. In this proof, we are asking the reader to accept that it is "obvious".

2. If you ask 5 mathematicians to prove this statement, you will get 5 different proofs (like our two versions above), possibly vastly different. Some aspects will be consistent (for example, the use of complete sentences, and the fact that 10 is divisible by 5), but the proofs themselves could look quite different.

3. Don't expect to immediately be able to write a proof in the style of our proof; instead, focus on correctness, clarity, using complete sentences, and including all necessary steps. The writing style of your proofs will develop as you read and write more proofs.

Here is another example, illustrating the *interpreting*, *brainstorming* and *sketching* portions of our proof-writing process. We won't fully prove this proposition, but we will illustrate some first steps in the proof-discovery process.

Example. Consider the following proposition:

Every function $f : \mathbb{R} \to \mathbb{R}$ represented by a degree 2 polynomial is non-constant.

1. [Interpret] We rewrite this statement as an if-then statement. The result is quite similar to the

original statement (and of course, is identical in meaning, the whole point is to not change the meaning).

Answer: If a function $f : \mathbb{R} \to \mathbb{R}$ is represented by a degree 2 polynomial, then f is non-constant.

2. [Brainstorm] What are the most important definitions to recall for this problem? Write down those definitions carefully.

   Answer: "Degree 2 polynomial" and "non-constant function".

   A *degree 2 polynomial* is a polynomial which can be written in the form $n_2 x^2 + n_1 x + n_0$, where $n_2 \neq 0$.

   A function f is *non-constant* if there exist inputs a and b such that $f(a) \neq f(b)$.

   Thus the theorem says that if $f(x) = n_2 x^2 + n_1 x + n_0$ with $n_2 \neq 0$, then we can find two values a and b (with $a \neq b$) such that $f(a) \neq f(b)$.

   Make a few examples to convince yourself that the statement is true.

   - $f(x) = x^2 + 1$ is a degree 2 polynomial. It is not constant because $f(0) = 1$ but $f(1) = 2$. (A less formal justification that it's non-constant: the graph of $f(x)$ is not a horizontal line.)
   - $f(x) = x^2 - x$ is a degree 2 polynomial. It is not constant because $f(0) = 0$ but $f(-1) = 2$.

   Plan: Following the same strategy used in the previous examples, we can try to plug in some small values of x. Hopefully, the corresponding outputs will be distinct. (If not, we will try some more inputs.)

3. [Sketch] Let's try plugging in 0, 1, and -1 to our general polynomial $n_2 x^2 + n_1 x + n_0$. We get:

   - $f(0) = n_0$
   - $f(1) = n_2 + n_1 + n_0$
   - $f(-1) = n_2 - n_1 + n_0$.

   Can we find two values that have different outputs?

   Answer: If $f(1) \neq f(0)$, we are done. Otherwise, if $f(1) = f(0)$, consider $f(-1)$. We will show that $f(-1) \neq f(1)$, because if $f(-1) = f(1) = f(0)$, then f cannot have degree 2.

   Here are the details: suppose $f(1) = f(0)$, then $n_2 + n_1 = 0$. If $f(-1) = f(1)$, then $n_1 = -n_1$, so $n_1 = 0$. So if $f(0) = f(1) = f(-1)$, then $n_2 = 0$, but that means f is not a degree 2 polynomial. Since we assumed f is a degree 2 polynomial, we know that $f(0) = f(1) = f(-1)$ is impossible, so f is not a constant polynomial.

We are not quite ready to write out a formal proof using the suggested strategy, because that strategy is secretly using the proof technique "proof by contradiction" which we will cover soon. Instead, let's practice reviewing "proofs" by evaluating some alternative arguments to prove our proposition (which may, or may not, be logically correct).

Example. Consider the same statement as above.

Every function $f : \mathbb{R} \to \mathbb{R}$ represented by a degree 2 polynomial is non-constant.

Are the following arguments valid proofs? If not, why not?

- "Proof": Consider the degree 2 polynomial $f(x) = 3x^2 - 5x + 7$. This is not a constant polynomial, because $f(0) = 7$ and $f(1) = 5$.

  Is it a valid proof?

  Answer. This is not a valid proof, because it is providing a single example. A complete proof needs to work for *every* degree 2 polynomial, not just one.

- "Proof": A degree 2 polynomial is a parabola (concave up, if the leading coefficient is positive, and concave down, if the leading coefficient is negative). In either case, the graph is not a horizontal line, so it is non-constant.

  Is it a valid proof?

  Answer. There is nothing *incorrect* with this proof, but it also feels incomplete. Everything involving concave up or concave down seems like it would need to be justified (what is even the definition of these terms?) and justifying them would probably be more difficult than proving the original statement.

- "Proof": Let $f(x) = ax^2 + bx + c$ be represented by a degree 2 polynomial (thus $a \neq 0$), and assume it is constant. Because f is constant, then its derivative is zero, so $2ax + b = 0$ for all x. But this is only possible if $a = b = 0$. Given that $a = 0$, f is not a degree 2 polynomial.

  Is it a valid proof?

  Answer. Assuming we are allowed to use properties of derivatives in our proof, then this is a correct and complete proof. (Ask your professor if you're unsure what tools you are allowed to use in a proof.)

  You may be surprised by the approach we took in this proof: the way we proved that f is *not* constant is by showing that it cannot be otherwise; indeed, if we assumed that f were constant, we would reach some nonsense. As we will see shortly, this is an example of a "proof by contradiction".

- "Proof": The only constant polynomials are degree zero, and because $0 \neq 2$, a degree two polynomial cannot be a constant function.

  Is it a valid proof?

  Answer. Again, there is nothing incorrect in what's written here, but the "proof" is essentially stating as obvious something that immediately implies what we are trying to prove. Not only the statement "The only constant polynomials are degree zero" is not trivial, and hence requires a proof, but it is actually a more elaborate (and general) result than the proposition we are trying to prove.

  This last argument is definitely not a complete proof in our context. (If your proof seems too easy, for example if it's a midterm and you're given a whole page but only need one sentence, you should worry that you might be assuming too much.)

Reading Quiz

1. At which of the following parts of the proof-writing process is it acceptable to use shorthand? Select all that apply.
   (a) It is never acceptable to use shorthand when writing a proof.
   (b) It is always acceptable to use shorthand when writing a proof.
   (c) The sketching phase.
   (d) The proof phase.
2. True or False: Your proof sketch is for your eyes only.
3. Most of the hard work happens during the __________ phase.
   (a) Brainstorming
   (b) Sketching
   (c) Proof

Practice Problems

1.4.1 Use the steps discussed in this section to prove the following theorem.

Theorem 1.12. *If the sum of two primes is again prime, then one of the primes must be 2.*

Video Solution

Exercises

1.4.1 Take the sketch for the following theorem and turn it into a well-written proof.

Theorem 1.13. *If n is an integer greater than 2, then $n^2 - 1$ is composite.*

Sketch. $n^2 - 1 = (n-1)(n+1)$ ■

1.4.2 Critique the following proof. That is, fix any mathematical errors and suggest ways to make the proof more clear.

Prove or disprove: an integer is composite if and only if it has two distinct prime factors.

Proof. The claim is false. An integer is composite if it is not prime. For example, the number 6 is composite because $6 = 2 \cdot 3$. But $4 = 2 \cdot 2$. ■

1.4.3 What is wrong with the following proof? Explain why this error could be prevented by planning out a proof before writing it.

Theorem 1.14. *If $n \geq 3$ is an integer, then $n^2 > 2n + 1$.*

Proof. Let $n^2 > 2n + 1$. Then $0 < n^2 - 2n - 1 = (n-1)^2 - 2$, which is only true when $n \geq 3$. ■

1.4.4 We usually prove statements like "P if and only if Q" in two steps. First we prove that P implies Q and then we prove that Q implies P. Explain why we can also do this; first prove that P implies Q and then prove that $\neg P$ implies $\neg Q$.

1.4.5 Consider the following definition.

Definition: For integers n, a, and a positive integer d, we write $n \equiv a \pmod{d}$ if n is the sum of a and an integer multiple of d.

Now consider the statement:

$$\textit{If } n \equiv 1 \pmod{2} \textit{ and } m \equiv 1 \pmod{4}, \textit{ then } n^2 - m \textit{ is divisible by 4.}$$

Below is a sketch of a proof of this statement. Unfortunately, it is incorrect. Find the mistake(s) in the sketch and explain your reasoning.

Sketch:
$n = 1 + 2k,\ m = 1 + 4k,$
$1 + 4k + 4k^2 - (1 + 4k),$
$n^2 - m = 4k^2,\ 4 \text{ divides } 4(k^2).$

1.4.6 Consider the following claim:

Claim: *There are infinitely many prime numbers.*

The claim is true, but the proof given below contains one extraneous statement. Such a statement is a sentence that, if deleted, does not change the validity of the proof. Find the extraneous statement and briefly justify why it does not change the proof.

Proof. Assume there are exactly n prime numbers $2 \leq p_1, ..., p_n$ and consider the integer

$$N = p_1 \cdots p_n + 1.$$

N is an odd number since $p_1 = 2$ and $N = 2(p_2 \cdots p_n) + 1$. Certainly N is divisible by some prime p_j in the list since we are assuming that the list $p_1, ..., p_n$ contains all primes. Thus

$$N - p_1 p_2 \cdots p_n = 1$$

is divisible by that p_j. This is a contradiction since 1 is not divisible by an integer $p_j \geq 2$. ■

1.4.7 We give three attempts to prove the following statement:

For any positive real numbers a and b, $\frac{a+b}{2} \geq \sqrt{ab}$.

Below, three attempts of a proof are given.

Attempt 1: Assuming that $\frac{a+b}{2} \geq \sqrt{ab}$, we can multiply both sides by 2, which is positive, to get

$$a + b \geq 2\sqrt{ab}$$

Squaring, we get

$$\begin{aligned} (a+b)^2 &\geq 4ab \\ a^2 + 2ab + b^2 &\geq 4ab \\ a^2 - 2ab + b^2 &\geq 0 \\ (a-b)^2 &\geq 0 \end{aligned}$$

Since we have arrived at a true statement, the initial assumption must be true. Thus, $\frac{a+b}{2} \geq \sqrt{ab}$ as desired.

Attempt 2: For all positive numbers a and b, we have $(\sqrt{a} - \sqrt{b})^2 \geq 0$. Thus, $a - 2\sqrt{a}\sqrt{b} + b \geq 0$, which is equivalent to $a + b \geq 2\sqrt{ab}$. Dividing by 2, which is positive, we obtain $\frac{a+b}{2} \geq \sqrt{ab}$, as desired.

Attempt 3: Let $a = 18$ and $b = 2$. Then $\frac{a+b}{2} = 10$ and $\sqrt{ab} = 6$. $10 \geq 6$ so this is true.

(a) Decide which of the proof attempts below is best; describe as completely as possible why the attempt you chose is best.

(b) What are your criticisms of the other proof attempts? Be as specific and thorough as possible.

1.4.8 Given below is the proof of a statement. Determine which statement the proof is trying to show.

Proof. Let m and n be integers. Assume that $m - n$ is odd, say $m - n = 2t + 1$ for some integer t. We consider two cases:
If n is odd, then $n = 2s + 1$ for some integer s. Thus

$$m = (m - n) + n = (2t + 1) + (2s + 1) = 2(t + s + 1).$$

Because $t + s + 1$ is an integer, this shows that m is even.
If n is even, then $n = 2s$ for some integer s. Thus

$$m = (m - n) + n = (2t + 1) + 2(s) = 2(t + s) + 1.$$

Because $t + s$ is an integer, this shows that m is odd.

Hence m and n have opposite parity. ■

1.4.9 Consider the following definition.

Definition: Let f real-valued function on the set of real numbers. If $f(z) = 0$, then we call z a **zero** of f.

Suppose f is a function and S is your favorite subset of the real numbers. If we want to prove that the set of all zeros of f is contained in a set S, what can we assume and what do we want to show?

2 Logic and the Language of Proofs

In order to read and construct proofs, we need to start with the language in which they are written: *logic.* Logic is to mathematics what grammar is to English. Section 2.1 will not look particularly mathematical, but we'll quickly get to work in Section 2.3 using logic in a mathematical context.

2.1 Propositions

Definition 2.1. A **proposition** or **statement** is an expression that is either true or false.

Examples.

1. $17 - 24 = 7$.
2. 39^2 is an odd integer.
3. The moon is made of cheese.
4. Every cloud has a silver lining.
5. God exists.

In order to make sense, these propositions require a clear *definition* of every concept they contain. There are many concepts of God in many cultures, but once it is decided *which* we are talking about, it is clear that They either exist or do not. This example illustrates that a statement need not be indisputably true or false, or even determinable, in order to qualify as a proposition. Mostly when people argue over propositions and statements, what they are really disagreeing about are definitions!
Note that any expression that is neither true nor false is not a proposition. *January 1st* is not a proposition, neither is *Green.*

Truth Tables

One often has to deal with abstract propositions; those where you do not know the truth or falsity, or indeed when you don't explicitly know the proposition! In such cases it can be convenient to represent the combinations of propositions in a tabular format. For instance, if we have two propositions (P and Q), or even three (P, Q and R) then all possible combinations of truth T and falsehood F are represented in the following tables:

P	Q
T	T
T	F
F	T
F	F

P	Q	R
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

The mathematician in you should be looking for patterns and asking: how many rows would a truth table corresponding to n propositions have, and can I *prove* my assertion? Right now it is hard to prove that the answer is 2^n: induction (Chapter 5) makes this very easy.

Connecting Propositions: Conjunction, Disjunction and Negation

We now *define* how to combine propositions in natural ways, modeled on the words *and, or* and *not.*

Definition 2.2. Let P and Q be propositions. The **conjunction** (AND, $\wedge$) of P and Q, the **disjunction** (OR, $\vee$) of P and Q, and the **negation** (NOT, $\neg$, $\sim$, $\overline{\ }$) of P are defined by the following truth tables,

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

P	$\neg P$
T	F
F	T

It is usually better to use *and, or* and *not* rather than *conjunction, disjunction* and *negation*: the latter may make you sound educated, but at the risk of being misunderstood!

Example. Let P, Q and R be the following propositions:

- P. Irvine is a city in California.
- Q. Irvine is a town in Ayrshire, Scotland.
- R. Irvine has seven letters.

Clearly P is true while R is false. If you happen to know someone from Scotland, you might know that Q is true.[a] We can now compute the following (increasingly grotesque) combinations…

$P \wedge Q$	$P \vee Q$	$P \wedge R$	$\neg R$	$(\neg R) \wedge P$	$\neg(R \vee P)$	$(\neg P) \vee [((\neg R) \vee P) \wedge Q]$
T	T	F	T	T	F	T

[a] The second syllable is pronounced like the i in bin or win. Indeed the first Californian antecedent of the Irvine family which gave its name to UCI was an Ulster-Scotsman named James Irvine (1827–1886). Probably the family name was originally pronounced in the Scottish manner.

How did we establish these facts? Some are quick, and can be done in your head. Consider, for instance, the statement $(\neg R) \wedge P$. Because R is false, $\neg R$ is true. Thus $(\neg R) \wedge P$ is the conjunction of two true statements, hence it is true. Similarly, we can argue that $R \vee P$ is true (because R is false and P is true), so the negation $\neg(R \vee P)$ is false.
Establishing the truth value of the final proposition $(\neg P) \vee [((\neg R) \vee P) \wedge Q]$ requires more work. You may want to set up a truth table with several auxiliary columns to help you compute:

P	Q	R	$\neg P$	$\neg R$	$(\neg R) \vee P$	$((\neg R) \vee P) \wedge Q$	$(\neg P) \vee [((\neg R) \vee P) \wedge Q]$
T	T	F	F	T	T	T	T

The importance of parentheses in a logical expressions cannot be stressed enough. For example, try building the truth table for the propositions $P \vee (Q \wedge R)$ and $(P \vee Q) \wedge R$. Are they the same?

Conditional and Biconditional Connectives

In order to logically set up proofs, we need to see how propositions can lead one to another.

Definition 2.3. The **conditional** ($\implies$) and **biconditional** ($\iff$) *connectives* have the truth tables

P	Q	$P \implies Q$
T	T	T
T	F	F
F	T	T
F	F	T

P	Q	$P \iff Q$
T	T	T
T	F	F
F	T	F
F	F	T

For the proposition $P \implies Q$, we call P the **hypothesis** and Q the **conclusion**.

Observe that the expressions $P \implies Q$ and $P \iff Q$ are themselves *propositions*: they are sentences which are either true or false!

Synonyms

$\implies$ and $\iff$ can be read in many different ways:

$P \implies Q$	$P \iff Q$
P implies Q	P if and only if Q
Q if P	P iff Q
P only if Q	P and Q are equivalent
P is sufficient for Q	P is necessary and sufficient for Q
Q is necessary for P	

Example. The following propositions all mean exactly the same thing:

- If you are born in Rome, then you are Italian.
- You are Italian if you are born in Rome.
- You are born in Rome only if you are Italian.
- Being born in Rome is sufficient to be Italian.
- Being Italian is necessary for being born in Rome.

Are you comfortable with what P and Q are here?

The biconditional connective should be easy to remember: $P \iff Q$ is true precisely when P and Q have identical truth states. It is harder to make sense of the conditional connective. One way of thinking about it is to consider what it means for an implication to be *false*. If $P \implies Q$ is false, it is impossible to create a logical argument which assumes P and concludes Q. The second row of $P \implies Q$ encapsulates the fact that it should be impossible for truth ever to logically imply falsehood.

Aside. Why is $F \implies T$ considered *true?*

This is the most immediately confusing part of the truth table for the conditional connective. One way that may help to remember this is to think of the implication $P \implies Q$ as making a promise. For example, suppose your teacher says: "if the class earns a B average on the midterm exam, then I will buy donuts for the class." Under what circumstances will your teacher have lied to you? Only in the case that it is true that the class earned a B average, but the teacher failed to provide donuts for the class.

Here is a mathematical example, written with an English translation at the side.

$$\begin{aligned} 7 = 3 &\implies 0 \cdot 7 = 0 \cdot 3 && \text{(If } 7 = 3\text{, then 0 times 7 equals 0 times 3)} \\ &\implies 0 = 0 && \text{(then 0 equals 0)} \end{aligned}$$

Thus $7 = 3 \implies 0 = 0$. Logically speaking this is a perfectly correct argument, thus the *implication* is true. The argument makes us uncomfortable because $7 = 3$ is clearly false.

Theorems and Direct Proofs

Truth tables and connectives are very abstract. To apply them to mathematics we need the following basic notions of theorem and proof.

Definition 2.4. A **theorem** is a justified assertion that some statement of the form $P \implies Q$ is true. A **proof** is an argument that justifies the truth of a theorem.

Think back to the truth table for $P \implies Q$ in Definition 2.3. Suppose that the hypothesis P is true and that $P \implies Q$ is true: that is, $P \implies Q$ is a *theorem*. We must be in the *first row* of the truth table, and so the conclusion Q is also true. This is how we think about proving basic theorems. In a *direct proof* we start by assuming the hypothesis (P) is true and make a logical argument ($P \implies Q$) which asserts that the conclusion (Q) is true. As such, it often convenient to rewrite the statement of a theorem as an implication of the form $P \implies Q$. Here is a very simple theorem which we prove directly.

Theorem 2.5. *The product of two odd integers is odd.*

The first thing to do is to write the theorem in terms of propositions and connectives: that is, in the form $P \implies Q$.

- P is 'x and y are odd integers.' This is our assumption, the hypothesis.
- Q is 'The product of x and y is odd.' This is what we want to show, the conclusion.
- Showing that $P \implies Q$ is true, that (the truth of) P implies (the truth of) Q requires an argument. This is the *proof.*

Proof. Let x and y be *any* two odd integers. We want to show that product $x \cdot y$ is an odd integer. By definition, an integer is odd if it can be written in the form $2k+1$ for *some* integer k. Thus there must be integers n, m such that $x = 2n+1$ and $y = 2m+1$. We compute:

$$x \cdot y = (2n+1)(2m+1) = 4mn + 2n + 2m + 1 = 2(2mn + n + m) + 1.$$

Because $2mn + n + m$ is an integer, this shows that $x \cdot y$ is an odd integer. ■

It is common to place a symbol (in this case ■) at the end of a proof to tell the reader that your argument is complete. Traditionally the letters Q.E.D. (from the Latin *quod erat demonstrandum*, literally 'which is what had to be demonstrated') were used, but this has gone out of style.You may also feel that you want to write more, or less than the above. This is a difficult thing to judge. What do you feel is a convincing argument? Test your argument on your classmates. The appropriate level of detail will depend on your readership: a middle school student will need more detail than a graduate student! At the moment, the best guide is to write for someone with the same mathematical sophistication as yourself. If, in three weeks' time, you can return to what you've written and understand it, then it's probably good!

The Converse and Contrapositive

The following constructions are used continually in mathematics: it is vitally important to know the difference between them.

Definition 2.6. The **converse** of an implication $P \implies Q$ is the reversed implication $Q \implies P$. The **contrapositive** of $P \implies Q$ is $\neg Q \implies \neg P$.

In general, we cannot say anything about the truth value of the converse of a true implication. The contrapositive of a true implication is, however, *always* true. Actually, even more is true, an implication and its contrapositive always have the same truth value. This is a common enough phenomenon that we give it its own name.

Definition 2.7. We say two propositions are **logically equivalent** if they have the same truth table.

This definition is a bit vague (what does having the same truth table mean?) We could give a more rigorous definition, but instead hope that the following examples will make the definition clear.

Example. We prove that the expressions $(P \implies Q) \wedge (Q \implies P)$ and $P \iff Q$ are logically equivalent by computing their truth tables:

P	Q	$P \implies Q$	$Q \implies P$	$(P \implies Q) \wedge (Q \implies P)$
T	T	T	T	**T**
T	F	F	T	**F**
F	T	T	F	**F**
F	F	T	T	**T**

P	Q	$P \iff Q$
T	T	**T**
T	F	**F**
F	T	**F**
F	F	**T**

Notice that the bolded columns are the same in each table.

Note that when comparing truth tables, one should make sure the inputs (e.g. the columns for P and Q in the above example) are in the same order in both tables.

Theorem 2.8. *The contrapositive of an implication is logically equivalent the original implication.*

Proof. Simply use our definitions of negation and implication to compute the truth table:

P	Q	$P \implies Q$	$\neg Q$	$\neg P$	$\neg Q \implies \neg P$
T	T	**T**	F	F	**T**
T	F	**F**	T	F	**F**
F	T	**T**	F	T	**T**
F	F	**T**	T	T	**T**

Since the truth states in the third and sixth columns (in bold) are identical, we see that $P \implies Q$ and its contrapositive $\neg Q \implies \neg P$ are logically equivalent. ∎

Example. Let P and Q be the following statements:

P. Claudia is holding a peach.

Q. Claudia is holding a piece of fruit.

The implication $P \implies Q$ is true, since all peaches are fruit. As a sentence, we have:

If Claudia is holding a peach, then Claudia is holding a piece of fruit.

The *converse* of $P \implies Q$ is the sentence:

If Claudia is holding a piece of fruit, then Claudia is holding a peach.

This is palpably false: Claudia could be holding an apple!

The *contrapositive* of $P \implies Q$ is the following sentence:

If Claudia is *not* holding any fruit, then she is *not* holding a peach.

This is clearly true.

Proof by Contrapositive

The fact that $P \implies Q$ and $\neg Q \implies \neg P$ are logically equivalent allows us, when convenient, to prove $P \implies Q$ by instead proving its contrapositive. As an example, consider another basic theorem.

Theorem 2.9. *Let x and y be integers. If $x + y$ is odd, then exactly one of x or y is odd.*

The theorem is an implication of the form $P \implies Q$ where

P. The sum $x + y$ of integers x and y is odd.

Q. Exactly one of x or y is odd.

A direct proof would require that we assume P to be true and logically deduce the truth of Q. For instance we might start our argument with:

Suppose that $x + y = 2n + 1$ for some integer n

The problem is that this doesn't really tell us anything about x and y, which we need to think about in order to demonstrate the truth of Q. Instead we consider the negations of our propositions:

$\neg Q$. x and y are both even or both odd (they have the same parity).

$\neg P$. The sum $x + y$ of integers x and y is even.

Since $P \implies Q$ is logically equivalent to the seemingly simpler contrapositive $(\neg Q) \implies (\neg P)$, we choose to prove the latter. This is, by Theorem 2.8, equivalent to proving the original implication.

Proof. Assume that x and y have the same parity. There are two cases: x and y are both even, or both odd.

Case 1: Let $x = 2m$ and $y = 2n$ be even. Then $x + y = 2(m + n)$ is even.

Case 2: Let $x = 2m + 1$ and $y = 2n + 1$ be odd. Then $x + y = 2(m + n + 1)$ is even.

In both cases $x + y$ is even, and the result is proved. ∎

De Morgan's Laws

In order to perform proofs by contrapositive (and later by contradiction) it is necessary to compute the negations of propositions. The most helpful results in this regard are attributable to Augustus de Morgan, a very famous 19th century logician.

Theorem 2.10 (de Morgan's laws)**.** *Let P and Q be any propositions. Then:*

1. *$\neg(P \wedge Q)$ is logically equivalent to $\neg P \vee \neg Q$.*
2. *$\neg(P \vee Q)$ is logically equivalent to $\neg P \wedge \neg Q$.*

Here is a proof of the first law. Try the second on your own.

Proof.

P	Q	$P \wedge Q$	$\neg(P \wedge Q)$	$\neg P$	$\neg Q$	$\neg P \vee \neg Q$
T	T	T	**F**	F	F	**F**
T	F	F	**T**	F	T	**T**
F	T	F	**T**	T	F	**T**
F	F	F	**T**	T	T	**T**

Simply observe that the fourth and seventh columns are identical. ∎

It is worth pausing to observe how similar the two laws are, and how concise. There is some beauty here. With a written example the laws are much easier to comprehend.

Example. (Of the first law) Suppose that of a morning you can choose (or not) to ride the subway to work, and you can choose (or not) to have a cup of coffee. Consider the following sentence:

I rode the subway *and* I had coffee.

What would it mean for this sentence to be *false?* Any sentence which asserts the falsehood of the above is a suitable *negation.* For example:

I *didn't* ride the subway *or* I *didn't* have coffee.

Note that the mathematical use of *or* includes the possibility that you neither rode the subway nor had coffee.

You will see de Morgan's laws again when we encounter sets.

Aside. Think about the meaning!
In the previous example we saw how negation switches *and* to *or.* This is true only when *and* denotes a conjunction between two propositions. Before applying De Morgan's laws, think about the *meaning* of the sentence. For example, the negation of

Mark and Mary have the same height.

is the proposition:

Mark and Mary do not have the same height.

If you blindly appeal to De Morgan's laws you might end up with the following piece of nonsense:

Mark *or* Mary do not have the same height.

Logical rules are wonderfully concise, but very easy to misuse. Always think about the meaning of a sentence and you shouldn't go wrong.

Negating Conditionals

You will often want to understand the negation of a statement. In particular, it is important to understand the negation of a conditional $P \implies Q$. Is it enough to say 'P doesn't imply Q'? And what could this mean? To answer the question you can use truth tables, or just think.
Here is the truth table for $P \implies Q$ and its negation: recall that negation simply swaps T and F.

P	Q	$P \implies Q$	$\neg(P \implies Q)$
T	T	T	F
T	F	F	T
F	T	T	F
F	F	T	F

The only time there is a T in the final column is when *both* P is true *and* Q is false. We have therefore proved the following:

Theorem 2.11. *$\neg(P \implies Q)$ is logically equivalent to $P \wedge \neg Q$ (read 'P and not Q').*

Now *think* in words rather than calculate. What is the negation of the following implication?

It's the morning therefore I'll have coffee.

Hopefully it is clear that the negation is:

It's the morning *and* I *won't* have coffee.

The implication *'therefore'* has disappeared and the expression *'and won't'* is in its place.

Warning! The negation of $P \implies Q$ is *not a conditional.* In particular it is *neither* of the following:

The converse, $Q \implies P$.

The contrapositive of the converse, $\neg P \implies \neg Q$.

If you are unsure about this, write down the truth tables and compare.

Example. Let x be an integer. What is the negation of the following sentence?

If x is even, then x^2 is even.

Written in terms of propositions, we wish to negate $P \implies Q$, where P and Q are:

P. x is even.

Q. x^2 is even.

The negation of $P \implies Q$ is $P \wedge \neg Q$, namely:

x is even and x^2 is odd.

This is very different to $\neg P \implies \neg Q$ (if x is odd then x^2 is odd).

Keep yourself straight by thinking about the meaning of these sentences. It should be obvious that 'x even $\implies x^2$ even' is true. It negation should therefore be *false.* The fact that it is false should make reading the negation feel a little uncomfortable.

Tautologies and Contradictions

We finish this section with two related concepts that are helpful for understanding proofs.

Definition 2.12. A **tautology** is a logical expression that is always true, regardless of what the component statements might be.
A **contradiction** is a logical expression that is always false.

The easiest way to detect these is simply to construct a truth table. We remark that two propositions φ and ψ are logically equivalent if and only if $\varphi \iff \psi$ is a tautology.

Examples. 1. $P \wedge (\neg P)$ is a very simple contradiction:

P	$\neg P$	$P \wedge (\neg P)$
T	F	F
F	T	F

Whatever the proposition P is, it cannot be true at the same time as its negation.

2. $(P \wedge (P \implies Q)) \implies Q$ is a tautology. This is essentially how we understand a direct proof: if P is true and we have a correct argument $P \implies Q$, then Q must also be true.

P	Q	$P \implies Q$	$P \wedge (P \implies Q)$	$(P \wedge (P \implies Q)) \implies Q$
T	T	T	T	T
T	F	F	F	T
F	T	T	F	T
F	F	T	F	T

Aside. Algebraic Logic

One can study logic in a more algebraic manner. De Morgan's Laws are algebraic. Here are a few of the other basic laws of logic.

Law of Double Negation :

$\neg(\neg P)$ is logically equivalent to P

Commutative laws :

$P \wedge Q$ is logically equivalent to $Q \wedge P$

$P \vee Q$ is logically equivalent to $Q \vee P$

Associative laws :

$(P \wedge Q) \wedge R$ is logically equivalent to $P \wedge (Q \wedge R)$

$(P \vee Q) \vee R$ is logically equivalent to $P \vee (Q \vee R)$

Distributive laws :

$(P \wedge Q) \vee R$ is logically equivalent to $(P \vee R) \wedge (Q \vee R)$

$(P \vee Q) \wedge R$ is logically equivalent to $(P \wedge R) \vee (Q \wedge R)$

You can check them all with truth tables. Using these rules, one can answer questions, such as deciding when an expression is a tautology, without laboriously creating truth tables. It is even fun! Such an approach is appropriate when you are considering abstract propositions, say in a formal logic course. In this text our primary interest with logic lies in using it to prove theorems. When one has an explicit theorem it is important to keep the meanings of all propositions clear. By relying too much on abstract laws like the above, it is easy to lose the meaning and write nonsense!

Reading Quiz

1. A *tautology* is a proposition which ____________________

    (a) is false no matter what the truth value of its component propositions.
    (b) is only true when all of its component propositions are true.
    (c) is never false, no matter what the truth value of its component propositions.
    (d) is built only using the connectives $\wedge, \vee$.

2. A *contradiction* is a proposition which ____________________

    (a) is false no matter what the truth value of its component propositions.
    (b) is only true when all of its component propositions are false.
    (c) is never false, no matter what the truth value of its component propositions.
    (d) is built only using the connective $\neg$.

3. The *contrapositive* of the conditional $P \implies Q$ is the conditional _________________

    (a) $\neg P \implies Q$
    (b) $\neg Q \implies \neg P$
    (c) $Q \implies P$
    (d) $P \implies \neg Q$

4. True or False: The *converse* of $P \implies Q$ is always logically equivalent to $P \implies Q$.

5. The *negation* of the conditional "if I study at least 25 hours per week, then I will be successful" is the proposition _________________

    (a) "I study at least 25 hours per week, but I am not successful."
    (b) "Either I study less than 25 hours per week, or I am successful."
    (c) "Either I study at least 25 hours per week, or I am not successful."
    (d) 'If I am successful, then I will study at least 25 hours per week."

6. De Morgan's laws state that:

    $\neg(P \vee Q)$ is logically equivalent to (1)_______________
    $\neg(P \wedge Q)$ is logically equivalent to (2)_______________

    (a) (1) $\neg(P \implies Q)$, (2) $(\neg P \vee \neg Q)$
    (b) (1) $(\neg P \wedge \neg Q)$, (2) $(P \vee Q)$
    (c) (1) $(\neg P \wedge Q)$, (2) $(P \vee \neg Q)$
    (d) (1) $(\neg P \wedge \neg Q)$, (2) $(\neg P \vee \neg Q)$

Practice Problems

2.1.1 Suppose that "If Colin was early, then no-one was playing pool" is a true statement.

(a) What is its contrapositive of this statement? Is it true?

(b) What is the converse? Is it true?

(c) What can we conclude (if anything?) if we discover each of the following? *Treat the two scenarios separately.*

(i) Someone was playing pool.

(ii) Colin was late.

Video Solution

2.1.2 Prove that $P \lor \neg Q$ is logically equivalent to $\neg P \implies (\neg P \land \neg Q)$.

Video Solution

2.1.3 Define the connective $\uparrow$ (called the *Sheffer stroke*, or *NAND*) by the following truth table:

P	Q	$P \uparrow Q$
T	T	F
T	F	T
F	T	T
F	F	T

(a) Prove $P \uparrow Q$ is logically equivalent to $\neg(P \land Q)$.

(b) Find an expression built using only P and the connective $\uparrow$ which is logically equivalent to $\neg P$.

(c) Find an expression built using only P, Q, and the connective $\uparrow$ which is logically equivalent to $P \land Q$.

Video Solution

Exercises

2.1.1 Express each of the following statements in the "If ..., then ..." form. There are many possible correct answers.

(a) You must eat your dinner if you want to grow.

(b) Being a multiple of 12 is a sufficient condition for a number to be even.

(c) It is necessary for you to pass your exams in order for you to obtain a degree.

(d) A triangle is equilateral only if all its sides have the same length.

2.1.2 Suppose that "x is an even integer" and "y is an irrational number" are true statements and that "$z \geq 3$" is a false statement. Which of the following are true?
Hint: Label each of the given statements, and think about each of the following using connectives.

(a) If x is an even integer, then $z \geq 3$.

(b) If $z \geq 3$, then y is an irrational number.

(c) If $z \geq 3$ or x is an even integer, then y is an irrational number.

(d) If y is an irrational number and x is an even integer, then $z \geq 3$.

2.1.3 Write the negation, the converse and the contrapositive of the following claim:

If A and B are invertible matrices, then AB is a square matrix and $\det(AB) \neq 0$.

Write your answers in sentences, like the original.

2.1.4 Orange County has two competing transport plans under consideration: widening the 405 freeway and constructing light rail down its median. A local politician is asked, "Would you like to see the 405 widened or would you like to see light rail constructed?" The politician wants to sound positive, but to avoid being tied to one project. What is their response?
(*Hint: Think about how the word 'OR' is used in logic...*)

2.1.5 Consider the proposition P given below:

If the integer m is greater than 3, the integer $2m$ is not prime.

(a) Rewrite P using the word 'necessary.'

(b) Rewrite P using the word 'sufficient.'

(c) Write the negation, the converse and the contrapositive of P.

Write your answers in sentences, like the originals.

2.1.6 Let A be a square matrix. Consider the proposition Q given below:

For A to be invertible, it is necessary and sufficient that A is non-singular.

(a) Rewrite Q as a biconditional.

(b) Write the negation of Q. (Hint: Your answer should be the disjunction of two sentences.)

2.1.7 Let m and n be two integers. Consider the statement:

m and n are not both even.

(a) Your friend writes this statement as

$m \wedge n$ are not both even.

What are some issues with what your friend wrote, if any?

(b) Then your friend claims that the negation of this statement is

m or n is odd.

Is this the correct negation? If not, what is the correct negation?

2.1.8 Construct the truth tables for the propositions $P \vee (Q \wedge R)$ and $(P \vee Q) \wedge R$. Are they the same?

2.1.9 Complete the truth table for the following proposition.

P	Q	R	$P \implies Q$	$(P \implies Q) \wedge \neg R$	$((P \implies Q) \wedge \neg R) \implies P$
T	T	T			
T	T	F			
T	F	T			
T	F	F			
F	T	T			
F	T	F			
F	F	T			
F	F	F			

2.1.10 Apply de Morgan's laws to the result of Theorem 2.11 to prove that $P \implies Q$ is logically equivalent to $\neg P \vee Q$. Do not use truth tables for this exercise.

2.1.11 Prove the law of *double negation*, i.e., that

$\neg(\neg P)$ is logically equivalent to P.

2.1.12 Prove that

(a) $(P \vee P)$ is logically equivalent to P

(b) $(P \wedge P)$ is logically equivalent to P

This is known as *idempotence* for $\vee$ and $\wedge$, respectively.

2.1.13 Prove that

(a) $P \vee (P \wedge Q)$ is logically equivalent to P

(b) $P \wedge (P \vee Q)$ is logically equivalent to P

These are known as the *absorption laws*.

2.1.14 Prove or disprove: $\neg P \vee \neg Q$ is logically equivalent to $P \implies (P \wedge \neg Q)$.

2.1.15 Recall that a contradiction is a combination of statements that is always false, regardless of the truth values of the original statements. A combination of statements that is always true is called a tautology.

(a) Is $(P \wedge \neg P) \implies Q$ a tautology, a contradiction, or neither?

(b) Prove that $((P \vee Q) \wedge \neg P) \wedge \neg Q$ is a contradiction.

(c) Prove that $(\neg P \wedge Q) \vee (P \wedge \neg Q) \iff \neg(P \iff Q)$ is a tautology.

2.1.16 Prove or disprove: $(P \wedge \neg Q \implies F) \iff (P \implies Q)$ is a tautology. Here F represents a *contradiction:* some proposition which is always false.

2.1.17 (a) Suppose that 'f is a linear function and b is a zero of f' is a false statement. What can we conclude if we discover each of the following? *Treat the two scenarios separately.*

i. f is a linear function.

ii. f is a linear function if and only if b is a zero of f.

(b) Suppose that 'If *Amy* likes art, then no one likes history.' is a true statement.

i. What is the contrapositive of this statement? Is it true?

ii. What is the converse of this statement? Is it true?

iii. What can we conclude (if anything?) if we discover each of the following? *Treat the two scenarios separately.*

A. Someone is likes history.

B. Amy does not like art.

2.1.18 Suppose that the following statements are true:

- Every octagon is magical.
- If a polygon is not a rectangle, then is it not a square.
- A polygon is a square, if it is magical.

Is it true that 'Octagons are rectangles'? Explain your answer.
(*Hint: try rewriting each of the statements as an implication.*)

2.1.19 (a) Use a truth table to prove the distributive law:

$(P \wedge Q) \vee R$ is logically equivalent to $(P \vee R) \wedge (Q \vee R)$

(b) Use logical algebra (see the aside on page 33) to prove that

$$((P \implies R) \wedge (Q \implies R)) \iff ((P \vee Q) \implies R)$$

is a tautology. (*Hint: start by using the result of question 10*)

2.1.20 (a) Do there exists propositions P, Q such that both $P \implies Q$ and its converse are true?

(b) Do there exist propositions P, Q such that both $P \implies Q$ and its converse are false?

Justify your answers by giving an example or a proof that no such examples exist.

2.1.21 (a) Suppose we have propositions P and Q such that both P and $\neg P$ are sufficient for Q. What, if anything, can be said about the tuth value of Q?

(b) Find truth values for P and Q which make the following expression false.

$$(\neg P \wedge \neg Q) \vee Q$$

2.1.22 (Hammack's *Book of Proof*, Section 2.5, Exercise 11.) Suppose P is false and that the statement $(R \implies S) \iff (P \wedge Q)$ is true. Find the truth values of R and S. (This can be done with or without a truth table).

2.1.23 Let R be the proposition "The summit of Mount Everest is underwater". Suppose that S is a proposition such that $(R \vee S) \iff (R \wedge S)$ is false.

(a) What can you say about S?

(b) What if, instead, $(R \vee S) \iff (R \wedge S)$ is true?

Hopefully it is obvious to you that R is false...

2.1.24 Complete the following.

(a) Let F be a contradiction and R any proposition. Prove $F \wedge R$ is a contradiction.

(b) Fill in the following proof of the fact that $Q \wedge \neg(P \implies Q)$ is a contradiction.

Proof. We give a chain of logical equivalences which simplifies the statement:

$Q \wedge \neg(P \implies Q)$	is logically equivalent to $Q \wedge (P \wedge \neg Q)$	(____________)
	is logically equivalent to ____________	(Commutative law)
	is logically equivalent to $(Q \wedge \neg Q) \wedge P$	(____________)

By part (a), since __________ is a contradiction, so is $(Q \wedge \neg Q) \wedge P$. Hence $Q \wedge \neg(P \implies Q)$ is a contradiction. ∎

2.1.25 Define the connective $\downarrow$ (called the *Quine dagger*, or *NOR*) by the following truth table:

P	Q	$P \downarrow Q$
T	T	F
T	F	F
F	T	F
F	F	T

(a) Prove $P \downarrow Q$ is logically equivalent to $\neg(P \vee Q)$.

(b) Find an expression built using only P and the connective $\downarrow$ which is logically equivalent to $\neg P$.

(c) Find an expression built using only P, Q, and the connective $\downarrow$ which is logically equivalent to $P \wedge Q$. [Hint: do the problem for $P \vee Q$ first, then use De Morgan's laws.]

2.1.26 (Hard) Suppose that P, Q are propositions. Argue that *any* of the 16 possible truth tables

P	Q	?
T	T	T/F
T	F	T/F
F	T	T/F
F	F	T/F

represents an expression ? created using only P and Q and the operations $\wedge, \vee, \neg$. Can you extend your argument to show that any truth table with any number of inputs represents some logical expression?

2.2 Propositional Functions and Quantifiers

While the logic of propositions from Section 2.1 is a fairly straightforward place to start our explorations, it is not powerful enough to express the majority of statements one encounters when actually doing mathematics. For one, it lacks the ability to deal with *variables*, which are of central importance in mathematics. For example, the expressions "17 is a prime number greater than 2", "5 is a prime number greater than 2", and "32 is a prime number greater than 2" are really just three instances of the same thing. Namely

"x is a prime number greater than 2"

with 17, 2, and 32 plugged in for x, respectively. In propositional logic, however, these three would constitute three wholly different propositions, obscuring their relationship.

Moreover, propositional logic fails to capture standard patterns of argument which we wish to consider. Consider propositions P and Q. If we know that $P \implies Q$ is true, and we know P is true, then we *must* conclude that Q is true (if you don't believe this, try to prove it yourself!) Since this holds no matter what propositions P and Q actually are, we call this a *valid argument*. On the other hand, an argument of the form P is true, Q is true, therefore R is true is not valid as it could be the case that P and Q are true and R false if you choose P, Q, and R to be certain statements! Now consider the following:

All prime numbers greater than 2 are odd.
17 is a prime number greater than 2.
Therefore 17 is odd.

This is certainly correct reasoning, however in propositional logic this argument takes the form of: P is true, Q is true, therefore R is true. But this (1) is not valid and (2) does not capture the true form of the argument. A better way to translate this argument would be:

All A's are B.
x is an A.
Therefore x is a B.

It turns out this is a valid argument, it does not depend on what A and B actually are. To study this in more depth, we will need to move beyond propositions to *propositional functions* and *quantifiers*.

Definition 2.13. A **propositional function** is a family of propositions which depend on one or more variables. The collection of objects allowed to be substituted in for variables in a propositional function is its **domain**.

For instance if $P(x)$ is a propositional function depending on a single variable x, then for each object a in the domain of P, $P(a)$ is a proposition.

Example. Suppose that x is allowed to be any real number. We could define the propositional function $P(x)$ by $x^2 > 4$.

In this example $P(5)$ is true, whilst $P(-1)$ is false. More generally, $P(x)$ is true for some values of

x (namely $x > 2$ or $x < -2$) and false for others ($-2 \leq x \leq 2$).

Example. Suppose that x is allowed to be any integer. Define the propositional function $P(x)$ by "x is a prime number" and $Q(x)$ by "$x > 2$".

The expression "x is a prime number greater than 2" can then be translated as $P(x) \wedge Q(x)$. Thus $P(17) \wedge Q(17)$ is true, whilst $P(2) \wedge Q(2)$ and $P(32) \wedge Q(32)$ are false.

At the beginning of the section we considered the expression "All prime numbers greater than 2 are odd". We saw in the example above how to translate "x is a prime number greater than 2" into logic. To translate the rest of the expression, we need to deal with the "all" and the "is odd" parts. Here, "all" is an example of a *quantifier*, something that tells us how many things satisfy some propositional function. This one is called the *universal quantifier* as it says everything (in the domain) satisfies some propositional function.

Let's try to translate "x is odd" into logic now. Recall that a number x is odd if it can be written as $2k + 1$ *for some* integer k. In other words *there exists* and integer k such that $x = 2k + 1$. Here we see another type of quantifier, the *existential quantifier*, which posits the existence of some (at least one) thing that satisfies some propositional function. The English language has all sorts of quantifiers (all, some, many, few, etc.) but in mathematics we primarily deal with just two.

Definition 2.14. The **universal quantifier** is denoted $\forall$ (read "for all" or "for every").

Let $P(x)$ be a propositional function. Then

$$\forall x\, P(x) \qquad \text{(read: for all } x,\ P(x) \text{ is true)}$$

is a proposition which is true if and only if $P(x_0)$ is true for all x_0 in the domain of P.

Definition 2.15. The **existential quantifier** is denoted $\exists$ (read "there exists").

Let $P(x)$ be a propositional function. Then

$$\exists x\, P(x) \qquad \text{(read: there is an } x, \text{ such that } P(x) \text{ is true)}$$

is a proposition which is true if and only if there exists some x_0 in the domain of P such that $P(x_0)$ is true.

We pause briefly to introduce some notation to help speed things along. We use $\mathbb{N}$ to denote the positive integers, $\mathbb{Z}$ the integers, $\mathbb{R}$ the real numbers, and $\in$ for 'is a member of the set'. Thus $2 \in \mathbb{Z}$ is read as '2 is a member of the set of integers', or more concisely, '2 is an integer'. We will properly cover this notation in Chapter 3.

Example. Recall the above example where, for each real number x, $P(x)$ is the proposition $x^2 > 4$. Consider the quantified propositions:

- $\forall x\, P(x)$ is *false*, since $P(x)$ is not true for all $x \in \mathbb{R}$. In particular $P(-1)$ is false.
- $\exists x\, P(x)$ is *true*, since there is at least one $x \in \mathbb{R}$ for which $P(x)$ is true, namely $x = 5$.

Definition 2.16. A **counterexample** to $\forall x\, P(x)$ is a single element x_0 in the domain of P such that $P(x_0)$ is false.
An **example** of $\exists x\, P(x)$ is a single element x_0 in the domain of P such that $P(x_0)$ is true.

Clearly $x_0 = -1$ is a counterexample to $\forall x\, (x^2 > 4)$, while $x_0 = 5$ is an example of $\exists x\, (x^2 > 4)$.

Example. Here is a slightly more complicated example with a propositional function with two variables. Let $R(x, y)$ be given by $x = 2y + 1$ where we agree that x and y are to be integers. Then $\exists y\, R(x, y)$ asserts that there exists some integer, let's call it k, such that $x = 2k + 1$. In other words, $\exists y\, R(x, y)$ asserts that x is odd. Let $O(x)$ be $\exists y\, R(x, y)$. Note that $O(x)$ is a propositional *function*. It still depends on x!

As a test to see if you are following along, check to make sure the following make sense:

$O(5)$ is true
$O(24)$ is false
$\forall x\, O(x)$ is false
$\exists x\, O(x)$ is true.

Mathematics is filled with compound expressions of this type, where quantifiers and propositional functions are combined to create more complicated propositional functions.

Two Common Translations

There are two constructions involving quantifiers which are common enough that we point them out here. It can be useful to understand the underlying logical structure of statements of the form "all A's are B's" and "there is an A which is a B". The sentences "all primes greater than 2 are odd" and "there exists an odd prime" are somewhat natural examples of these types of statements.

Remark. Let $P(x)$ and $Q(x)$ be propositional functions. The statement

All A's are B's

is really a statement of the following form, where $P(x)$ is the propositional function "x is an A" and $Q(x)$ is the propositional function "x is a B":

Everything which satisfies $P(x)$ also satisfies $Q(x)$

which can be written as

$$\forall x\,(P(x) \implies Q(x)).$$

Similarly, the statement

There is an A which is a B

is really

There is something satisfying $P(x)$ which also satisfies $Q(x)$

which can be written as

$$\exists x\,(P(x) \wedge Q(x)).$$

Example. "All humans are mortal" is written in logic as $\forall x\,(P(x) \implies Q(x))$ where $P(x)$ is "x is a human" and $Q(x)$ is "x is mortal".

Example. We let $P(x)$ be "x is a prime number", $Q(x)$ be $x > 2$, and $O(x)$ be "x is odd". Then "all primes greater than 2 are odd" can be written as

$$\forall x\,[(P(x) \wedge Q(x)) \implies O(x)].$$

"There is an odd prime" can be written as

$$\exists x\,(P(x) \wedge O(x)).$$

Bounded Quantifiers

Often we wish to make explicit the domain of a propositional function which we are quantifying, or to restrict our quantifiers to smaller parts of the domain. We can accomplish this by the use of *bounded quantifiers*. We introduce this via examples.

Example. Consider the statement: "every real number is a cube". By what we said above, we could translate this into logic as

$$\forall x\, Q(x)$$

where $Q(x)$ is "x is a cube" and has domain the real numbers[a]. If we want to emphasize that we are quantifying over the real numbers, we often write something like

$$\forall x \in \mathbb{R}\, Q(x).$$

In reality, mathematicians rarely write out specific statements using notation like $P(x)$ or $Q(x)$ (these letters are usually reserved for standing in for abstract propositional functions). In this example, "x is a cube" can be written as $\exists y \in \mathbb{R}\,(x = y^3)$. So "every real number is a cube" is

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

[a] We could also write this as $\forall x\,(P(x) \implies Q(x))$ where $P(x)$ is "x is a real number".

Example. While the statement "every real number is a square" is false, we can make it a true statement by restricting our attention to just *positive* reals: "every *positive* real number is a square" is true. We could write this as

$$\forall x > 0\,\exists y \in \mathbb{R}\,(x = y^2).$$

Here, the condition $x > 0$ in the $\forall$ quantifier restricts the quantifier the smaller domain of just those real numbers which are > 0, i.e., positive real numbers.

Aside. Clarity versus Concision

As with all forms of art, different practitioners of mathematics have different tastes. Some write very concisely, keeping words to a minimum. Some write almost entirely in English. Some use a hybrid of quantifiers and English, aiming for a balance between brevity and clarity. For example, consider the famous *sum of four squares* theorem:

English	Every positive integer may be written as the sum of the squares of four integers
Extreme Logic	$(\forall n \in \mathbb{N})(\exists a, b, c, d \in \mathbb{Z})(n = a^2 + b^2 + c^2 + d^2)$
Hybrid	$\forall n \in \mathbb{N}$, $\exists a, b, c, d \in \mathbb{Z}$ such that $n = a^2 + b^2 + c^2 + d^2$

The purpose of writing mathematics is to help the reader understand what you've written *without* you being there to explain it. Your presentation style has an enormous effect on whether you are successful! A good rule is to write in sentences, replacing words with symbols only when it makes things more readable while simultaneously preserving the flow of the sentence.

Remark. In some sense, we don't really need bounded quantifiers. For example, one can view $\forall x \in \mathbb{R}\, P(x)$ as simply shorthand for $\forall x\,(x \in \mathbb{R} \implies P(x))$ and $\exists x > 0\, P(x)$ as shorthand for $\exists x(x > 0 \wedge P(x))$. This pattern holds in general: you can replace a bounded $\forall$ quantifier with an unbounded one by putting the condition at the front of an implication and a bounded $\exists$ quantifier by putting the condition in a conjunction.
In practice, however, this can get very messy, especially in statements with many quantifiers! Getting comfortable with bounded quantifiers can help you write cleaner statements.

Negating Quantified Propositions

Perhaps the most important skill to have regarding quantifiers (in this course) is knowing how to negate them.

Theorem 2.17. *For any propositional function $P(x)$, we have:*

1. $\neg(\forall x\, P(x))$ *is logically equivalent to* $\exists x\, \neg P(x)$.
2. $\neg(\exists x\, P(x))$ *is logically equivalent to* $\forall x\, \neg P(x)$.

In essence, negation swaps the quantifiers $\forall \leftrightarrow \exists$. Like with all theorems, if you want to understand it, you should unpack it, write it in English, and come up with some examples.

1. The negation of '$P(x)$ is true for all x' is, '$P(x)$ is false for some x.'
2. The negation of '$P(x)$ is true for some x' is, '$P(x)$ is always false.'

Examples. Here are two examples, numbered corresponding to the parts of Theorem 2.17.

1. The negation of the statement, 'Everyone owns a bicycle' is:

   Somebody does not own a bicycle.

   It is extremely pedantic, but symbolically we might write:

   $$\neg\Big[\forall \text{ people } x,\ x \text{ owns a bicycle}\Big] \iff \exists \text{ a person } x \text{ such that } x \text{ does not own a bicycle.}$$

2. Suppose that x is a real number and consider the quantified proposition:

   $$\exists x \text{ such that } \sin x = 4.$$

   This has the form $\exists x\, P(x)$, and therefore has negation $\forall x\, \neg P(x)$. Explicitly, the negation is:

   $$\forall x \text{ we have } \sin x \neq 4.$$

   Note how we introduced the words *we have* to make the sentence read more clearly.

Once you're comfortable negating simple propositions and quantifiers, negating multiple quantifiers is easy. Just follow the rules, think, and take your time.

Example. Show that the following statement about the real numbers is false.

$$\forall x\, \exists y \text{ such that } xy = 3.$$

The negation of this expression follows the rules for switching quantifiers and negating the final statement:

$$\exists x, \text{ such that } \forall y \text{ we have } xy \neq 3.$$

It is easy to see that the negated statement is true:

Proof. Let $x = 0$, then, regardless of y, we have $xy = 0 \neq 3$. ∎

Because the negation is true, the original statement is false.

Warning! When negating bounded quantifiers, do **not** change the condition. Here is a couple examples to illustrate. We leave the reason as to why to not change the condition to the exercises.

Example.

1. The negation of "every yellow car has four doors" is **not** "there exists a non-yellow car without four doors". The correct negation is "there is a yellow car without four doors".
2. Consider the statement "every positive real number is a square" which we saw we could write as

$$\forall x > 0\, \exists y \in \mathbb{R}\, (x = y^2).$$

Note that this is a *true* statement and so its negation must be false. So something like

$$\exists x \leq 0\, \forall y \in \mathbb{R}\, (x \neq y^2)$$

could *not* be the correct negation as this is also a true statement. Instead, simply negate the quantifiers as you normally would, but leave the condition as is:

$$\exists x > 0\, \forall y \in \mathbb{R}\, (x \neq y^2).$$

Advice when Negating: Hidden and Excess Quantifiers

Theorem 2.17 seems very simple, but it is easy to misuse. Here are some points to consider when negating quantifiers.

1. Don't forget the *meaning* of the sentence. Use the logical rules in Theorem 2.17, but also think it out in words. You should get the same result. Think about the finished sentence and read it aloud: if it *sounds* like the opposite of what you started with then it probably is!
2. The symbol $\nexists$ for 'does not exist' is much abused. Very occasionally its use is appropriate, but it too often demonstrates laziness or a lack of understanding. Avoid using it unless absolutely necessary.
3. Beware of hidden quantifiers! Sometimes a quantifier is not explicitly stated. This is especially the case with the universal quantifier and is very common when a statement contains an implication. Consider the following very easy theorem.

$$\text{If } n \text{ is an odd integer, then } n^2 \text{ is odd.} \tag{$*$}$$

This is really a statement about *all* integers. There is a hidden quantifier that's been suppressed in the interests of readability. Instead, the theorem could have been written

$$\forall n \in \mathbb{Z}\, (n \text{ is odd} \implies n^2 \text{ is odd}).$$

In this form we can negate by combining the rules in Theorems 2.11 and 2.17. The pattern is

$$\neg\left[\forall n\,(P(n) \implies Q(n))\right] \quad \text{is equivalent to} \quad \exists n\,(P(n) \land \neg Q(n)).$$

The negation of $(*)$ is therefore,

$\exists n \in \mathbb{Z}$ such that n is odd and n^2 is even.

Given that $(*)$ is a theorem, its negation is, of course, false!

Here is a harder example of a hidden quantifier, this time from Linear Algebra. You do not have to know what a vector is to work with this definition. We are purely concerned with how to negate an abstract statement.

Definition 2.18. Vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are **linearly independent** if

$$a\mathbf{x} + b\mathbf{y} + c\mathbf{z} = \mathbf{0} \implies a = b = c = 0$$

The implication is a statement about *all* real numbers a, b, c. We could instead have written

$\forall a, b, c \in \mathbb{R}$ we have $a\mathbf{x} + b\mathbf{y} + c\mathbf{z} = \mathbf{0} \implies a = b = c = 0.$

To negate the definition, we must also negate the hidden quantifier. The result is the definition of what it means for vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ to be *linearly dependent*:

$\exists a, b, c$ not all zero such that $a\mathbf{x} + b\mathbf{y} + c\mathbf{z} = \mathbf{0}$

The final challenge here is recalling how to negate an implication: recall Theorem 2.11, and note that the negation of $a = b = c = 0$ is that *at least one* of a, b, c is non-zero.

Putting it all together: the definition of continuity

You might have seen the strict definition of continuity in a calculus class.[4] It combines multiple quantifiers, a hidden quantifier and an implication. The purpose of this example isn't to teach you the subtleties of continuity. Just as with the linear independence example, we simply want to be able to read and negate such expressions.

Definition 2.19. Suppose that f is a function whose domain and codomain are sets of real numbers. We say that f is **continuous** at $x = a$ if,

$$\forall \varepsilon > 0\ \exists \delta > 0 \text{ such that } |x - a| < \delta \implies |f(x) - f(a)| < \varepsilon. \tag{$*$}$$

The implication is a statement about *all* real numbers x which satisfy some property, so we once again have a hidden quantifier:

$\forall \varepsilon > 0\ \exists \delta > 0$ such that $\underline{\forall x \in \mathbb{R}}\ |x - a| < \delta \implies |f(x) - f(a)| < \varepsilon.$

We can now use our rules to state what it means for f to be *discontinuous* at $x = a$:

$\exists \varepsilon > 0$ such that $\forall \delta > 0\ \exists x \in \mathbb{R}$ such that $|x - a| < \delta$ and $|f(x) - f(a)| \geq \varepsilon.$

Warning! Remember, the negation of $\forall \varepsilon > 0$ is *not* $\exists \varepsilon \leq 0$. Only the ultimate proposition[5] is negated! For an example of this definition in use, see the exercises.

[4] If not, you will have plenty time to get used to it in an upper-division Analysis course...

[5] In this case the ultimate proposition is $|x - a| < \delta \implies |f(x) - f(a)| < \varepsilon$.

The Order of Quantifiers Matters!

We conclude this section with an important observation: the order of quantifiers matters critically! Consider, for example, the following propositions:

1. For every person x, there exists a person y such that y is a friend of x.
2. There exists a person y such that, for every person x, y is a friend of x.

Assuming that x and y always represent people, we can rewrite the sentences as follows:

1. $\forall x\, \exists y$ such that y is a friend of x.
2. $\exists y$ such that $\forall x$ we have that y is a friend of x.

All we've done is to switch the order of the two quantifiers! How does this affect the meaning? Written entirely in English, the statements become:

1. Everyone has at least one friend.
2. There is someone who is friends with everybody.

Quite different! The critical observation is that if $\exists y$ comes after x, then y is *allowed to depend on* x. Each person might have a friend, but that friend is likely to be different depending on the person. If $\forall x$ comes after y, then x cannot depend on y.

Play around with the pairs of examples below. What are the meanings? Which ones are true?

- $\forall$ days x, $\exists$ a person y such that y was born on day x.
- $\exists$ a person y such that, $\forall$ days x, y was born on day x.

- $\forall$ circles x, $\exists$ a point y such that y is the center of x.
- $\exists$ a point y such that, $\forall$ circles x, y is the center of x.

- $\forall\, x \in \mathbb{Z}\ \exists y \in \mathbb{Z}$ such that $y \leq x$.
- $\exists y \in \mathbb{Z}$ such that, $\forall x \in \mathbb{Z}\quad y \leq x$.

What happens in the last two examples if we replace the integers $\mathbb{Z}$ with the positive integers $\mathbb{N}$?

Reading Quiz

1. Let $P(x)$ be $x^2 - 1 = 0$ with domain all real numbers. Which of the following are true? Select all that apply

   (a) $P(1)$

   (b) $P(-1)$

   (c) $P(3)$

   (d) $P(x)$

   (e) $\forall x\, P(x)$

(f) $\exists x\, P(x)$

(g) $\neg\forall x\, P(x)$

2. A value x_0 for which $P(x_0)$ is *false* is known as a(n) ____________________

   (a) example.

   (b) counterexample.

   (c) realization.

   (d) solution.

3. Which of the following are equivalent to the given expression?

   $$\neg\forall x\, \exists y\, P(x, y)$$

   (a) $\exists x\, \forall y\, P(x, y)$

   (b) $\neg\exists x\, \forall y\, P(x, y)$

   (c) $\exists x\, \forall y\, \neg P(x, y)$

   (d) $\forall x\, \exists y\, \neg P(x, y)$

4. True or False: the order of quantifiers in an expression can always be switched without changing the meaning of the expression.

Practice Problems

2.2.1 Write each of the following using propositional functions and quantifiers. Make sure to define any propositional functions you are using.

   (a) Every class has an instructor.

   (b) For all real numbers x and y, if x and y are positive, then there exists a positive integer n such that $nx > y$.

   (c) There exists a real number which is positive, and is less than $1/n$, for each positive integer n.

   Video Solution

2.2.2 Negate the following.

   (a) Every class has an instructor.

   (b) For all real numbers x and y, if x and y are positive, then there exists a positive integer n such that $nx > y$.

   (c) There exists a real number which is positive, and is less than $1/n$, for each positive integer n.

   Video Solution

2.2.3 Here are four propositions. Which are true and which false? Justify your answers.

   (a) $\forall x \in \mathbb{R},\ \exists y \in \mathbb{R}$ such that $y^4 = 4x$.

(b) $\exists y \in \mathbb{R}$ such that $\forall x \in \mathbb{R}$ we have $y^4 = 4x$.

(c) $\forall y \in \mathbb{R}$, $\exists x \in \mathbb{R}$ such that $y^4 = 4x$.

(d) $\exists x \in \mathbb{R}$ such that $\forall y \in \mathbb{R}$ we have $y^4 = 4x$.

Video Solution

Exercises

2.2.1 For each of the following sentences, rewrite the sentence using quantifiers. Then write the negation (using both words and quantifiers)

(a) All mathematics exams are hard.

(b) No football players are from San Diego.

(c) There is a odd number that is a perfect square.

2.2.2 Suppose that $P(x)$, $Q(y)$ and $R(x,y,z)$ are propositional functions. Compute the negation of the following quantified propositions:

(a) $\forall x\ \exists y\ P(x) \wedge Q(y)$

(b) $\forall x\ \exists y\ \forall z\, R(x,y,z)$

2.2.3 Suppose someone claims that the negation of

$$x^2 > 0 \implies x > 0$$

is '$x^2 > 0$ and $x \leq 0$.' Why is this incorrect? What is the correct negation?

2.2.4 Consider the propositional function

$$P(x,y,z): \quad (x-3)^2 + (y-2)^2 + (z-7)^2 > 0$$

where the domain of each of the variables x, y and z is $\mathbb{R}$.

(a) Express the quantified statement $\forall x \in \mathbb{R}, \forall y \in \mathbb{R}, \forall z \in \mathbb{R}, P(x,y,z)$ in words.

(b) Is the quantified statement in (a) true or false? Explain.

(c) Express the negation of the quantified statement in (a) in symbols.

(d) Express the negation of the quantified statement in (a) in words.

(e) Is the negation of the quantified statement in (a) true or false? Explain.

2.2.5 The following statements are about positive real numbers. Which one is true? Explain your answer.

(a) $\forall x\ \exists y$ such that $xy < y^2$.

(b) $\exists x$ such that $\forall y\ \ xy < y^2$.

2.2.6 Which of the following statements are true? Explain.

(a) $\exists$ a married person x such that $\forall$ married people y, x is married to y.

(b) $\forall$ married people x, $\exists$ a married person y such that x is married to y.

2.2.7 A function f is said to be *decreasing* if:

$$x \le y \implies f(x) \ge f(y).$$

(a) There is a hidden quantifier in the definition: what is it?

(b) State what it means for f not to be decreasing.

(c) Give an example to demonstrate the fact that *not decreasing* and *increasing* do not mean the same thing!

2.2.8 Are the following True or False? Give some explanation for why you chose your answer.

(a) For every two points A and B in the plane, there exists a circle on which both A and B lie.

(b) There exists a circle in the plane on which lie any two points A and B.

2.2.9 You are given the following definition (*you do not have to know what is meant by a field).*

Let x be an element of a field $\mathbb{F}$. An *inverse* of x is an element y in $\mathbb{F}$ such that $xy = 1$.

Consider the following proposition:

All non-zero elements in a field have an inverse.

(a) Restate the proposition using both of the quantifiers $\forall$ and $\exists$.

(b) Find the negation of the proposition, again using quantifiers.

2.2.10 Consider the following proposition.

$$\forall m, n \in \mathbb{R}, \quad m > n \implies m^2 > n^2. \tag{$\dagger$}$$

(a) What is the negation of ($\dagger$)?

(b) Prove that ($\dagger$) is *false.*

(c) Suppose you rewrite the proposition as follows

$$\forall m, n \in A, \quad m > n \implies m^2 > n^2.$$

What is the largest collection (set) of real numbers A for which the proposition is *true*? Justify your answer.

2.2.11 Let $P(x)$ be a propositional function and n a positive integer.

(a) Define the quantifier $\exists^{\ge n}$ so that the proposition $\exists^{\ge n} x\, P(x)$ is true if and only if there are at least n elements in the domain of $P(x)$ which make $P(x)$ true. Find an expression in which the only quantifiers are $\forall$ and $\exists$ which has the same meaning as $\exists^{\ge n} x\, P(x)$.

(b) Define the quantifier $\exists^{=n}$ so that the proposition $\exists^{=n} x\, P(x)$ is true if and only if there are exactly n elements in the domain of $P(x)$ which make $P(x)$ true. Find an expression in which the only quantifiers are $\forall$ and $\exists$ which has the same meaning as $\exists^{=n} x\, P(x)$.

2.2.12 Recall from calculus the definitions of the limit of a sequence $(x_n) = (x_1, x_2, x_3, \ldots)$.

'x_n *diverges to* ∞' means: $\forall M > 0$, $\exists N \in \mathbb{N}$ such that $n > N \implies x_n > M$.

'x_n *converges to* L' means: $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $n > N \implies |x_n - L| < \varepsilon$.

Here we assume that all elements of (x_n) are real numbers.

(a) State what it means for a sequence x_n not to diverge to ∞. *Beware of the hidden quantifier!*

(b) State what it means for a sequence x_n not to converge to L.

(c) State what it means for a sequence x_n not to converge at all.

(d) (Hard) Prove, using the definition, that the sequence defined by $x_n = n$ diverges to ∞.

(e) (Hard) Prove that the sequence defined by $x_n = \frac{1}{n}$ converges to zero.

[You may want to revisit the last two parts after reading the following sections.]

2.3 Methods of Proof

There are four standard methods for proving a theorem $P \implies Q$. In practice, long proofs will use several such arguments joined together. We have already discussed the first two types of proof in Section 2.1.

Direct Assume P is true and deduce that Q is true.

Contrapositive Assume $\neg Q$ and deduce $\neg P$. This is enough since the contrapositive $\neg Q \implies \neg P$ is logically equavalent to $P \implies Q$.

Contradiction Assume that P and $\neg Q$ are true and deduce a *contradiction*. Since $P \wedge \neg Q$ implies a contradiction, it follows that $P \wedge \neg Q$ must be *false*. By Theorem 2.11, we see that $P \implies Q$ is true.

Induction This has a completely different flavor: we will consider it in Chapter 5.

Each of the methods has advantages and disadvantages. For instance, the direct method has the advantage of a straightforward logical flow. The contrapositive method is useful when the *negations* $\neg P$, $\neg Q$ are simpler than P, Q themselves. This is often the case when one or both statements involve the *non-existence* of something. Working with their negations might give you the existence of ingredients with which you can calculate. Proof by contradiction has a similar advantage: assuming both P and $\neg Q$ gives you two pieces of information with which you can calculate. Logically speaking there is no difference between the three methods, beyond how you visualize your argument.

To illustrate the difference between direct proof, proof by contrapositive, and proof by contradiction, we prove the same simple theorem in three different ways.

Theorem 2.20. *Suppose that x is an integer. If $3x + 5$ is even, then $3x$ is odd.*

Direct Proof. We show that if $3x + 5$ is even then $3x$ is odd.

Assume that $3x + 5$ is even, then $3x + 5 = 2n$ for some integer n. Hence

$$3x = 2n - 5 = 2(n - 3) + 1.$$

This is clearly odd, because it is of the form 'an even integer plus one.' ∎

Contrapositive Proof. We show that if $3x$ is even then $3x + 5$ is odd.

Assume that $3x$ is even, and write $3x = 2n$ for some integer n. Then

$$3x + 5 = 2n + 5 = 2(n + 2) + 1.$$

This is odd, because $n + 2$ is an integer. ∎

Contradiction Proof. We assume that $3x + 5$ and $3x$ are both even, and we deduce a falsehood. Write $3x + 5 = 2m$ and $3x = 2n$ for some integers m and n. Then

$$5 = (3x + 5) - 3x = 2m - 2n = 2(m - n).$$

Since $m - n$ is an integer, this says that 5 is even: a contradiction. ∎

Some simple proofs

Theorem 2.21. *Let $m, n \in \mathbb{Z}$. Both m and n are odd if and only if the product mn is odd.*

There are really two theorems here:

($\Rightarrow$) If m and n are both odd integers, then the product mn is odd.

($\Leftarrow$) If the product mn of two integers is odd, then both m and n are odd.

Often when there are two directions you'll have to prove them separately. Here we give a direct proof for ($\Rightarrow$) and a contapositive proof for ($\Leftarrow$).

Proof. ($\Rightarrow$) Let m and n be odd. Then $m = 2k + 1$ and $n = 2l + 1$ for some $k, l \in \mathbb{Z}$. Then

$$mn = (2k+1)(2l+1) = 4kl + 2k + 2l + 1 = 2(2kl + k + l) + 1.$$

This is odd, because $2kl + k + l \in \mathbb{Z}$.

($\Leftarrow$) Suppose that the integers m and n are *not* both odd. That is, assume that *at least one* of m and n is even. We show that the product mn is even. Without loss of generality,[a] we may assume that n is even, from which $n = 2k$ for some integer k. Then,

$$mn = m(2k) = 2(mk) \quad \text{is even.}$$

∎

[a] See 'Potential Mistakes' below for what this means.

In the second part of the proof, we did not need to consider whether m was even or odd: if n is even, the product mn is even regardless. The second part would have been very difficult to prove directly. For instance, you might have tried to start a direct proof with:

Assume that mn is odd, then $mn = 2k + 1$ for some integer k. Then...

We are stuck!

Theorem 2.22. *If $3x + 5$ is even, then x is odd.*

We can prove this directly, by the contrapositive method, or by contradiction. We'll do all of them, so you can appreciate the difference.

Direct Proof. Simply quote the two previous theorems. Because $3x + 5$ is even, $3x$ must be odd by Theorem 2.20. Now, since $3x$ is odd, both 3 and x are odd by Theorem 2.21. ■

Contrapositive Proof. Suppose that x is even. Then $x = 2m$ for some integer m and we get

$$3x + 5 = 6m + 5 = 2(3m + 2) + 1.$$

Because $3m + 2 \in \mathbb{Z}$, we have $3x + 5$ odd. ■

Contradiction Proof. Suppose that both $3x + 5$ and x are even. We can write $3x + 5 = 2m$ and $x = 2k$ for some integers m and k. Then

$$5 = (3x + 5) - 3x = 2m - 6k = 2(m - 3k)$$

is even. Contradiction. ■

Selecting a method of proof is often a matter of taste. You should be able to see the advantages and disadvantages of the various approaches. The direct proof is more logically straightforward, but it depends on two previous results. The contrapositive and the contradiction arguments are quicker and more self-contained, but they require a greater level of comfort with logic. Consider who you are writing for before you decide to present a slick difficult proof over a slow simple one.[6] For even more variety, here is a direct proof of Theorem 2.22 that does not use any previous result.

Alternative Direct Proof. Suppose $3x + 5$ is even, so $3x + 5 = 2n$ for some integer n. Then

$$\begin{aligned} x &= (3x + 5) - 2x - 5 = 2n - 2x - 5 \\ &= 2(x - n - 3) + 1 \end{aligned}$$

is odd. ■

The fact that such variety is possible just makes proving theorems even more fun!

Common Mistake 1. Generality and 'Without Loss of Generality'

There are many common mistakes in the writing of proofs that you should be careful to avoid. Here are two incorrect 'proofs' of the $\implies$ direction of Theorem 2.21.

[6] The Hungarian mathematician Paul Erdős used to refer to simple, elegant proofs as being 'from the Book,' as if the Almighty had a book of perfect proofs of which mere mortals might occasionally be permitted a glimpse. Of course, as with all matters spiritual, one person's Book may be very different to another's…

Fake Proof 1. $m = 3$ and $n = 5$ are both odd, and so $mn = 15$ is odd. ■

This is an *example* of the theorem, not a proof. Examples are critical to helping you understand and believe what a theorem says, but they are no substitute for a proof! Recall the discussion in the Introduction on the usage of the word *proof* in English.

Fake Proof 2. Let $m = 2k + 1$ and $n = 2k + 1$ be odd. Then,

$$mn = (2k + 1)(2k + 1) = 2(2k^2 + 2k) + 1$$

is odd. ■

The problem with this second 'proof' is that it is insufficiently general. m and n are supposed to be *any* odd integers, but by setting both of them equal to $2k + 1$, we've chosen m and n to be the same! Notice how the correct proof uses $m = 2k + 1$ and $n = 2l + 1$, where we place no restriction on the integers k and l.

By *generality* we mean that we must make sure to consider all possibilities encompassed by the hypothesis. The phrase *Without Loss of Generality*, often shorted to WLOG, is used when a choice is made which might at first appear to restrict things but, in fact, does not.

Think back to how this was used in the the proof of Theorem 2.21. Since the integers m and n appear symmetrically in the Theorem, if at least one of them is even, then we lose nothing by assuming that the second integer n is even.

The phrase WLOG is used to pre-empt a challenge to a proof in the sense of *Fake Proof 2*, as if to say to the reader:

'You might be tempted to object that my argument is not general enough. However, I've thought about it, and there is no problem.'

Common Mistake 2. Incorrect use of the equal sign Remember that propositions should be joined by connectives: i.e., by $\implies$ or $\iff$. It is very common to see students write something like

$$m \text{ is odd} = m = 2k + 1 \text{ for some integer } k$$

This is extremely confusing! If this is part of a longer argument, things will become very difficult to follow. Since 'm is odd' and '$m = 2k + 1$ for some integer k' are both *propositions*, they should be linked by a *connective*. We should instead write

$$m \text{ is odd} \iff m = 2k + 1 \text{ for some integer } k$$

Common Mistake 3. Becoming distracted by algebra Here is a palpably ludicrous 'theorem' which illustrates another potential mistake.

Theorem (Fake Theorem)**.** *The only number is zero.*

Fake Proof. Let x be any number and let $y = x$, then

$$\begin{aligned} x = y &\implies x^2 = xy && \text{(Multiply both sides by } x\text{)} \\ &\implies x^2 - y^2 = xy - y^2 && \text{(Subtract } y^2 \text{ from both sides)} \\ &\implies (x-y)(x+y) = (x-y)y && \text{(Factorize)} \\ &\implies x + y = y && \text{(Divide both sides by } x - y\text{)} \\ &\implies x = 0 \end{aligned}$$

■

Everything is fine up to the third line, but then we divide by $x - y$, which is zero! Don't let yourself become so enamoured of logical manipulations that you forget to check the basics.

More simple proofs

We continue with more straightforward proofs. None of these results are particularly important, they are just exercises in deciding how to present an argument.

Theorem 2.23. *Suppose that $x \in \mathbb{R}$. Then $x^3 + 2x^2 - 3x - 10 = 0 \implies x = 2$.*

We can prove this theorem using any of the three methods. All rely on your ability to factorize the polynomial:

$$x^3 + 2x^2 - 3x - 10 = (x-2)(x^2 + 4x + 5) = (x-2)\Big[(x+2)^2 + 1\Big],$$

and partly on your knowledge that $ab = 0 \iff a = 0$ or $b = 0$ (proof in the exercises).

Direct Proof. If $x^3 + 2x^2 - 3x - 10 = 0$, then $(x-2)[(x+2)^2 + 1] = 0$. Hence at least one of the factors $x - 2$ or $(x+2)^2 + 1$ is zero.
In the first case we conclude that $x = 2$.
The second case is impossible, since $(x+2)^2 \geq 0 \implies (x+2)^2 + 1 > 0$.
Therefore $x = 2$ is the only solution. ■

Contrapositive Proof. Suppose that $x \neq 2$. Then $x^3 + 2x^2 - 3x - 10 = (x-2)[(x+2)^2 + 1] \neq 0$ since neither of the factors is zero. ■

Contradiction Proof. Suppose that $x^3 + 2x^2 - 3x - 10 = 0$ and $x \neq 2$. Then

$$0 = x^3 + 2x^2 - 3x - 10 = (x-2)[(x+2)^2 + 1].$$

Since $x \neq 2$, we have $x - 2 \neq 0$.
It follows that $(x+2)^2 + 1$ must be zero. However, $(x+2)^2 + 1 \geq 1$ for all real numbers x, so we have a contradiction. ■

On balance, the contrapositive proof is probably the most elegant, but you can decide for yourself.

Common Mistake 4. Being excessively logical The statement of Theorem 2.23 is an implication $P \implies Q$ where P and Q are:

$$P. \quad x^3 + 2x^2 - 3x - 10 = 0, \qquad\qquad Q. \quad x = 2.$$

You can make life very hard for yourself by being overly logical. For instance, you may wish take a third proposition $R.\ \ x \in \mathbb{R}$, and state the theorem as $R \implies (P \implies Q)$. This is the way of pain! It's easier to assume, as a universal constraint, that you're always dealing with real numbers; you can then ignore said constraint within the argument.
Indeed, one can always append a third proposition to the front of any theorem, namely, "all math I already know." Try to resist the temptation to be so logical that your arguments become unreadable. The goal is to convince the reader that the theorem is true, not to confuse them!

Reading Quiz

1. In a *proof by contrapositive* of $P \implies Q$, we assume that (1) ________________ and deduce that (2) ________________.
   (a) (1) $\neg Q$ is true, (2) P is true
   (b) (1) Q is false, (2) P is true
   (c) (1) $\neg P$ is true, (2) $\neg Q$ is true
   (d) (1) $\neg Q$ is true, (2) $\neg P$ is true
2. A *proof by contradiction* of $P \implies Q$ begins by assuming that ____________________________.
   (a) $\neg P \vee Q$ is true
   (b) $P \wedge \neg Q$ is true
   (c) $P \implies Q$ is true
   (d) $Q \implies P$ is false
3. In which of the following situations would it be correct to invoke *without loss of generality*? Select all that apply.
   (a) Suppose we are attempting to prove that for two integers m and n, if either one is even, then so is the product. Without loss of generality we can assume that n is even.

(b) We are trying to prove that for two integers m and n, if both are odd, then so is the product. Without loss of generality, we can assume that both m and n are equal to $2k + 1$ for some integer k.

(c) Attempting to prove that if m is even and n is odd, then mn is even. Without loss of generality can be used to assume that $m = 2$.

(d) Attempting to prove that if three boxes are painted either green or gold, there must be two boxes which are painted the same color. Without loss of generality can be used to assume that the first box is painted green.

Practice Problems

2.3.1 Let x and y be integers. Prove: For $x^2 + y^2$ to be even, it is necessary that x and y have the same parity (i.e. both even or both odd).

Video Solution

2.3.2 Let n be an integer. Prove that, in order for n to be odd, it is sufficient that its ones digit is either 1, 3, 5, 7, or 9.

Video Solution

Exercises

2.3.1 Show that for any given integers a, b, c, if a is even and b is odd, then $7a - ab + 12c + b^2 + 4$ is odd.

2.3.2 Augustus de Morgan satisfied his own problem:

I turn(ed) x years of age in the year x^2.

(a) Given that de Morgan died in 1871, and that he wasn't the beneficiary of some miraculous anti-aging treatment, find the year in which he was born.

(b) Suppose you have an acquaintance who satisfies the same problem. How old will they turn this year?

Give a formal argument which justifies that you are correct.

2.3.3 Suppose you are teleported to a world in which the law of double negation does not necessarily hold, i.e., it is not the case that $\neg\neg P$ is equivalent to P. Would you be able to carry out proofs by contradiction in this world?

2.3.4 Prove that if n is a positive integer greater than 1, then $n! + 2$ is even.

Here $n!$ denotes the factorial *of the integer n. Look up the definition if you forgot about it.*

2.3.5 Consider the following proposition, where x is assumed to be a real number.

$$x^3 - 3x^2 - 2x + 6 = 0 \implies x = 3 \qquad (*)$$

(a) Is the proposition $(*)$ true or false? Justify your answer. Is its converse true?

(b) Repeat part (a) for the proposition

$$x^3 - 3x^2 - 2x + 6 = 0 \implies x \neq 3$$

(c) Does anything change about the truth status of $(*)$ if we assume that it is a statement about *rational numbers* x? Explain.

2.3.6 (a) Let $x \in \mathbb{Z}$. Prove that

$$5x + 3 \text{ is even if and only if } 7x - 2 \text{ is odd.}$$

Can you conclude anything about $7x - 2$ if $5x + 3$ is odd?

(b) Prove or disprove: An integer n is even if and only if n^3 is even.

2.3.7 Let n and m be positive integers. Prove $n^2 m$ is even if and only if n and m are not both odd.

2.3.8 Let x and y be integers. Prove $x^2 + y^2$ is even **if and only if** x and y have the same parity (i.e. both even or both odd).

2.3.9 Let n be an integer. Prove $n^2 + n + 58$ is even.

2.3.10 Below is the proof of a result. What result is being proved?

Proof. Assume that x is odd. Then $x = 2k + 1$ for some integer k. Then

$$2x^2 - 3x - 4 = 2(2k+1)^2 - 3(2k+1) - 4 = 8k^2 + 2k - 5 = 2(4k^2 + k - 3) + 1.$$

Since $4k^2 + k - 3$ is an integer, $2x^2 - 3x - 4$ is odd. ∎

2.3.11 Here is another proof. What is the result this time?

Proof. Assume, without loss of generality, that x and y are even. Then $x = 2a$ and $y = 2b$ for some integers a, b. Therefore,

$$xy + xz + yz = (2a)(2b) + (2a)z + (2b)z = 2(2ab + az + bz).$$

Since $2ab + az + bz$ is an integer, $xy + xz + yz$ is even. ∎

2.3.12 Consider the following proof of the fact that (for m an integer) if m^2 is even, then m is even. Is proceeding by contradiction necessary? Can you re-write the proof so that it doesn't use contradiction?

Proof. Suppose, for contradiction, that m^2 is even, but m is odd. Say $m = 2k + 1$ for some integer k. Then

$$m^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$$

is odd, contradicting that m^2 is even. Thus m must be even. ■

2.3.13 (a) Prove that if x and y are positive real numbers, then $\sqrt{x+y} \neq \sqrt{x} + \sqrt{y}$. *Argue by contradiction.*

(b) Prove that if x and y are positive real numbers, then $\sqrt{x+y} \neq \sqrt{x} + \sqrt{y}$. *Argue by contrapositive.*

2.3.14 Prove that $ab = 0 \iff a = 0$ or $b = 0$.

2.3.15 You meet three men, Corey, Jansen, and Vogel, each of whom is a either Truthteller or a Liar. Truthtellers speak only the truth; Liars speak only lies. You ask Corey whether he is a Truthteller or a Liar. Corey answers with his back turned, so you cannot hear what he says.

"What did he say?" you ask Jansen.

Jansen replies: "Corey says he is a Truthteller."

Vogel says: "Jansen is lying."

Is Vogel a Truthteller or a Liar? Explain your answer.

2.3.16 Assume that Mary's father lives in California. Consider the following implication P:

If Mary's father is an oilman and does not have any friends in Wisconsin, then Mary plays tennis or basketball, or she appeared in at least one article of a December 1997 New York Times newspaper edition.

(a) Find the contrapositive of P.

(b) Find the converse of P.

(c) Find the negation of P.

(d) Imagine you are a detective and want to find the truth value of P. Describe your action-strategy in full detail.

2.3.17 Suppose we have three proposition P, Q, and R, and we want to prove that all three are equivalent, i.e., $P \iff Q$, $Q \iff R$, and $P \iff R$.

(a) Prove that to establish these three equivalences, it is enough to show $P \iff Q$ and $Q \iff R$.

(b) Prove that to establish these three equivalences, it is enough to show $P \implies Q$, $Q \implies R$, and $R \implies P$.

2.3.18 Numbers of the form $\frac{k(k+1)}{2}$, where k is a positive integer, are called *triangular numbers*. Prove that n is the square of an odd number if and only if $\frac{n-1}{8}$ is triangular.

2.4 More Methods of Proof

Definition-Pushing

The next example concerns divisibility. Before you can prove a theorem, it is critical that you know the *meaning* of all of the words in its statement. We therefore state the definition of divisibility.

Definition 2.24. Let n and p be integers. We say that n is **divisible by** p (or p **divides** n) if $n = pk$ for some integer k. If n is divisible by p, we write $p \mid n$.

Now we can present a theorem.

Theorem 2.25. *If $n \in \mathbb{Z}$ is divisible by $p \in \mathbb{N}$, then n^2 is divisible by p^2.*

Proof. We prove directly. Let n be divisible by p. Then $n = pk$ for some $k \in \mathbb{Z}$. Then $n^2 = p^2k^2$, and so n^2 is divisible by p^2. ∎

This is an example of a *definition-pushing* proof. If you simply state the the definition of everything important in the theorem, the proof will often be staring you in the face.

Proof by Cases

The next proof is also in the definition-pushing vein. However, it requires that we consider several cases. The relevant definition here is that of *remainder.*

Definition 2.26. An integer n is said to have remainder $r = 0, 1,$ or 2 upon division by 3 if we can write $n = 3k + r$ for some integer k.

With a little thought, it should be clear that every integer is of the form $3k$, $3k + 1$ or $3k + 2$. This is analogous to how all integers are either even ($2k$) or odd $(2k + 1)$. We will consider remainders more carefully in Chapter 4.

Theorem 2.27. *If n is an integer, then n^2 has remainder 0 or 1 upon dividing by 3.*

Proof. We again prove directly. There are three cases: n has remainder 0, 1 or 2 upon dividing by 3.

(a) If n has remainder 0, then $n = 3m$ for some $m \in \mathbb{Z}$ and so $n^2 = 9m^2 = 3(3m^2)$ has remainder 0.

(b) If n has remainder 1, then $n = 3m + 1$ for some $m \in \mathbb{Z}$ and so

$$n^2 = 9m^2 + 6m + 1 = 3(3m^2 + 2m) + 1 \quad \text{has remainder 1.}$$

(c) If n has remainder 2, then $n = 3m + 2$ for some $m \in \mathbb{Z}$ and so

$$n^2 = 9m^2 + 12m + 4 = 3(3m^2 + 4m + 1) + 1 \quad \text{has remainder 1.}$$

Thus n^2 has remainder 0 or 1. ■

Proving Universal Statements

You have already seen many examples of proving universal statements in Section 2.3, albeit where the universal quantifier was hidden. For example, Theorem 2.22 could be written explicitly with the universal quantifier: $\forall x \in \mathbb{Z}$, if x is odd, then $3x+5$ is odd. Proving this version with the explicit universal quantifier would go much the same way as with the version with the hidden quantifier.

Sometimes theorems are not written explicitly in the form $\forall x\,(P(x) \implies Q(x))$. Here is one example, called the *Arithmetic Mean - Geometric Mean Inequality*, or *AM-GM* for short.

Theorem 2.28. *For all positive real numbers x and y, we have*

$$\frac{x+y}{2} \geq \sqrt{xy}.$$

This theorem is written in the form $\forall x, y > 0,\ P(x,y)$ where $P(x,y)$ is the statement $\frac{x+y}{2} \geq \sqrt{xy}$. It may seem unclear how to approach proving such a theorem as our proof methods so far have mostly focused on proving *conditional* statements. So one way to start would be to rewrite this theorem as a conditional. From Section 2.2, we know that $\forall x, y > 0,\ P(x,y)$ is equivalent to $\forall x, y\,((x > 0 \wedge y > 0) \implies P(x,y))$. So we are essentially proving:

If $x, y > 0$, then $\dfrac{x+y}{2} \geq \sqrt{xy}$.

First we give a direct proof: note how the implication signs are stacked to make the argument clearer.

Direct Proof. Clearly $(x-y)^2 \geq 0$ with equality if and only if $x = y$. Now multiply out:

$$\begin{aligned} x^2 - 2xy + y^2 \geq 0 &\iff (x^2 + 2xy + y^2) - 4xy \geq 0 \\ &\iff x^2 + 2xy + y^2 \geq 4xy \\ &\iff (x+y)^2 \geq 4xy \\ &\iff x + y \geq 2\sqrt{xy} \qquad (*) \\ &\iff \frac{x+y}{2} \geq \sqrt{xy}. \end{aligned}$$

The square-root in $(*)$ is well-defined because $x+y$ is positive, which is true because we assumed x and y are positive.[a] ■

[a] Moreover, the inequality is preserved since the function $f(t) = t^2$ is *increasing* when t is positive.

The following contradiction proof incorporates exactly the same calculation, but is laid out in a different order. This is not always possible, and you have to take great care when trying it. You will likely agree that the direct proof is easier to follow. Note that for contradiction we assume that

$$\neg \forall x, y > 0,\ \frac{x+y}{2} \geq \sqrt{xy}$$

is true. Pushing the negation through, our contradiction assumption becomes

$$\exists x, y > 0, \frac{x+y}{2} < \sqrt{xy}.$$

Contradiction Proof. Suppose that $\frac{x+y}{2} < \sqrt{xy}$ for some $x, y > 0$. Since $x + y \geq 0$, this is true if and only if $(x+y)^2 < 4xy$. Now multiply out and rearrange:

$$\begin{aligned}
(x+y)^2 < 4xy &\iff x^2 + 2xy + y^2 < 4xy \\
&\iff x^2 - 2xy + y^2 < 0 \\
&\iff (x-y)^2 < 0.
\end{aligned}$$

Since squares of real numbers are non-negative, this is a contradiction. Thus $\frac{x+y}{2} \geq \sqrt{xy}$. ∎

Proving Existential Statements

Now we look at proving existential statements of the form $\exists x\, P(x)$. The most simple approach is to come up with some object x_0 such that $P(x_0)$ is true. Such an x_0 is called an *example* or *witness* of $P(x)$.

Theorem 2.29. *Let m and n be two fixed integers with $m < n$. Then there exists a rational number r such that $m < r < n$.*

Proof. Let $r = \frac{m+n}{2}$ be the arithmetic mean of m and n. Then r is a rational number as it is of the form p/q for an integer p and nonzero integer q. Since $m < n$, we have $2m < m + n < 2n$ and dividing by 2 yields $m < r < n$. ∎

This is an example of a *constructive* proof as we explicitly construct the number which makes the proposition true. Later we will we give an example of a *non-constructive* proof of an existential statement where we show that an example must exist, even though we do not know what it is explicitly.

Disproving Universal Statements

What does it take to show that some universal statement is false? If a statement has the form $\forall x\, P(x)$, then by Theorem 2.17, its negation is $\exists x\, \neg P(x)$. In other words, *dis*proving the universal statement $\forall x\, P(x)$ is the same as *proving* the existential statement $\exists x\, \neg P(x)$. This amounts to providing a *counterexample*, i.e., finding some x_0 such that $P(x_0)$ is false.

Example. Disprove the following statement. If all four sides of a quadrilateral have the same length, then the quadrilateral must be a square.

Proof. All four sides of a square certainly have the same length, but remember that all four angles in a square must be congruent as well. Disproving this statement then amounts to finding some quadrilateral with four congruent sides, but without four congruent angles.

We can easily construct such a figure in the following way. Start with any non-right isosceles triangle $\triangle ABC$ where the sides $\overline{AB}$ and $\overline{AC}$ are congruent. Now place an identical copy of $\triangle ABC$, call it $\triangle A'B'C'$, on top of $\triangle ABC$ and reflect it through the side $\overline{BC}$. We're left with the quadrilateral $ABA'C$. The sides $\overline{AB}$ and $\overline{AC}$ are congruent to each other since we started with an isosceles triangle, and these are both congruent to $\overline{BA'}$ and $\overline{A'C}$ since these were obtained by copying $\overline{AB}$ and $\overline{AC}$. Finally, this figure is *not* a square because angle $\angle BAC$ not a right angle. You might remember this figure as a *rhombus*.

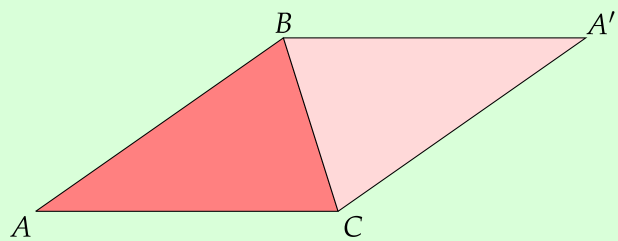


Figure 2: Square of side length $a + b$.

■

Disproving Existential Statements: Non-existence Proofs

When a theorem claims that something does not exist, it is generally a good idea to try a contrapositive or a contradiction proof. This is since 'does not exist' is already a *negative* statement. A contradiction or contrapositive proof of a theorem $P \implies Q$ already involves the negated statement $\neg Q$. If Q states that something does not exist, then $\neg Q$ states that it does, which often gives you something to play with! To see this in action, consider the following result.

Theorem 2.30. *The equation* $x^{17} + 12x^3 + 13x + 3 = 0$ *has no positive (real number) solutions.*

First we interpret the theorem as an implication: throughout we assume that x is a real number.

For all x, if $x^{17} + 12x^3 + 13x + 3 = 0$, then $x \leq 0$.

The theorem is now in the form $\forall x(\ P(x) \implies Q(x)\ )$, with:

$$P(x): \quad x^{17} + 12x^3 + 13x + 3 = 0, \qquad\qquad Q(x): \quad x \leq 0.$$

The negation of $Q(x)$ is simply '$x > 0$.' To prove the theorem by contradiction, we assume $\exists x\,(P(x) \wedge \neg Q(x))$, and deduce a contradiction.

Proof. Assume that a real number x satisfies $x^{17} + 12x^3 + 13x + 3 = 0$, and that $x > 0$. Because all terms on the left hand side are positive, we have $x^{17} + 12x^3 + 13x + 3 > 0$. A contradiction. ■

Note how quickly the proof is written: it assumes that any reader is familiar with the underlying logic of a contradiction proof without it needing to be spelled out. The discussion we undertook before writing the proof would be considered scratch work: you shouldn't include it a final write-up.

If you want to extend the above result, and you can recall the Intermediate and Mean Value Theorems from Calculus, you should be able to prove that there is exactly one (necessarily negative!) solution to the above polynomial equation.

Combining and Subdividing Theorems

Sometimes it is useful to break a proof into pieces, akin to viewing a computer program as a collection of subroutines that you combine for some greater purpose. Usually the intention is to make the proof of a difficult result more readable, but you may also wish to emphasize the importance of certain aspects of your work. Mathematics does this by using *lemmas* and *corollaries.*

> *Lemma:* a theorem whose importance you want to downplay. Often the result is individually unimportant, but becomes more useful when referenced in the proof of a larger theorem.
>
> *Corollary:* a theorem which follows quickly from another result. Corollaries can be used to draw attention to a particular aspect or a special case of a theorem.

In many mathematical papers the word *theorem* is reserved only for the most important results, everything else being presented as a lemma or corollary. The choice of what to call a result is entirely one of presentation. If you want your paper to be more readable, or to highlight what you think is important, then lemmas and corollaries are your friends!
Here is a famous example of a lemma at work.

Lemma 2.31. *Suppose that $n \in \mathbb{Z}$. Then n^2 is even $\iff$ n is even.*

Prove this yourself: the ($\Rightarrow$) direction is easiest using the contrapositive method, while the ($\Leftarrow$) direction works well directly.

Theorem 2.32. $\sqrt{2}$ *is irrational.*

This is tricky for a few reasons. The theorem does not appear to be of the form $P \implies Q$, , nor does it seem to be any of the forms covered in this section. However, let us look a little closer at what is being claimed. Remember *irrational* simply means *not rational*, i.e., not of the form m, n for $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. Thus saying that $\sqrt{2}$ is irrational is equivalent to saying that there are no $m, n \in \mathbb{N}$ such that $\sqrt{2} = \frac{m}{n}$. Phrased like this, we can see that the theorem statement is really a non-existence statement, and so it makes sense to try a proof by contradiction.

Proof. Suppose that $\sqrt{2} = \frac{m}{n}$ for some $m, n \in \mathbb{N}$. Without loss of generality, we may assume that m, n have no common factors.

Then $m^2 = 2n^2$ which says that m^2 is even.

By Lemma 2.31 we have that m is even.

Thus $m = 2k$ for some $k \in \mathbb{Z}$.

But now, $n^2 = 2k^2$, from which (Lemma 2.31 again) we see that n is even.

Now m and n have a common factor of 2. This is a contradiction. ∎

First observe how Lemma 2.31 was used to make the proof easier to read. Without the lemma, the essential shape of the proof would have been less clear.

Now try to make sense of the proof. In the first line we invoke the definition of *rational*, being the *ratio* of two integers. The main challenge comes immediately afterwards. Once we assume that $\sqrt{2} = \frac{m}{n}$, we can immediately insist that m, n have no common factors. Indeed this is no significant restriction *once we assume that m, n exist*, that is *once we assume that $\sqrt{2}$ is rational.* It is important to realize that the 'no common factors' assumption is *not* the assumption being contradicted. Because of this subtlety, we include the phrase 'without loss of generality' so that the reader is forced to think carefully, and does not jump to the wrong conclusion.

It might seem difficult to completely understand, but if we hadn't made the observation, our calculation could have continued forever, telling us nothing!

$$m^2 = 2n^2 \implies n^2 = 2k^2 \implies k^2 = 2l^2 \implies \cdots$$

If you find this approach difficult, you may prefer an alternative proof given in the exercises.

Non-constructive Proofs

We saw earlier that we usually prove an existential statement by coming up with an explicit object and showing that it has the desired properties. Here is an example that shows that this can be a bit subtle.

Theorem 2.33. *There are irrational numbers a and b such that a^b is a rational number.*

Proof. By Theorem 2.32, we know that $\sqrt{2}$ is irrational. Consider the number $(\sqrt{2})^{\sqrt{2}}$. There are two possibilities. If $(\sqrt{2})^{\sqrt{2}}$ is rational, then we can take $a = b = \sqrt{2}$ in the theorem statement and we're done. On the other hand, if $(\sqrt{2})^{\sqrt{2}}$ is irrational, then we have

$$\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = (\sqrt{2})^{\sqrt{2}\cdot\sqrt{2}} = (\sqrt{2})^2 = 2.$$

In this case, we can take $a = (\sqrt{2})^{\sqrt{2}}$ and $b = \sqrt{2}$ in the theorem statement. In both cases we have found irrational numbers a and b where a^b is rational. ∎

The critical, and perhaps unsatisfying, part of this proof is that we have *not* said whether or not $(\sqrt{2})^{\sqrt{2}}$ is rational. We then say that this proof is *non-constructive* since it does not actually contain explicit (and unconditional) irrational a and b with a^b rational. Instead, we have shown that the theorem is true *whether or not* $(\sqrt{2})^{\sqrt{2}}$ is rational.

Prime Numbers

Here is another famous result, dating back at least to the Ancient Greeks (Euclid's *Elements*, Proposition IX.20). As ever, we need a solid definition before we try to prove anything.

Definition 2.34. An integer ≥ 2 is *prime* if the only positive integers it is divisible by are itself and 1.

The first few primes are $2, 3, 5, 7, 11, 13, 17, 19, \ldots$ It follows[7] from the definition that all positive integers ≥ 2 are either primes or *composites* (products of primes). In particular, every integer ≥ 2 is divisible by at least one prime. We may now state Euclid's result.

Theorem 2.35. *There are infinitely many prime numbers.*

Proof. We prove by contradiction. Assume there are exactly n prime numbers $p_1, \ldots, p_n$ and consider the integer

$$\Pi := p_1 \cdots p_n + 1.$$

Certainly Π is divisible by some prime: since we are assuming that the list $p_1, \ldots, p_n$ contains *all* the primes, Π must be divisible by some prime p_i in the list. However, the product $p_1 \cdots p_n$ is clearly divisible by p_i, whence so is the difference[a]

$$\Pi - p_1 \cdots p_n = 1.$$

We conclude that 1 is divisible by the prime p_i, contradicting[b] the fact that $p_i \geq 2$. ∎

[a] Is this obvious? Can you prove it?!

[b] Euclid's original argument was not strictly by contradiction. Instead he asserted that, given any list of primes $p_1, \ldots, p_n$, the number Π must be divisible by a new prime not in his list.

Reading Quiz

2.4.1 When proving a non-existence statement, i.e., proving that something *does not* exist, proof by contradiction is often useful because ________________________.

(a) contradiction is more powerful than a direct proof.

(b) direct and contrapositive proofs are too complicated.

[7] This is not completely obvious: we will prove it much later in Theorem 5.19.

(c) it allows one to assume such an object exists, hence giving an object that can be manipulated.

(d) it allows one to assume such an object does not exist, which is exactly what the problem is asking for.

2.4.2 True or False: When proving a universal statement like $\forall x\, P(x)$, it suffices to give an explicit example of an x for which $P(x)$ holds.

2.4.3 True or False: When proving an existential statement like $\exists x\, P(x)$, it suffices to give an explicit example of an x for which $P(x)$ holds.

2.4.4 In the proof that $\sqrt{2}$ is irrational, we started by assuming that $\sqrt{2} = \frac{m}{n}$ for integers m and n *with no common factors*. Why is this justified?

(a) Because no pair of two integers ever has a common factor.

(b) Because any rational number $\frac{m}{n}$ can be seen, by canceling the common factors of m and n, to be equal to a rational $\frac{m'}{n'}$ where m' and n' have no common factors.

(c) It is not justified, we have lost generality by making this assumption.

(d) Because $\sqrt{2}$ is irrational.

Practice Problems

2.4.1 Prove or disprove the following conjectures:

(a) The sum of any 3 consecutive integers is divisible by 3.

(b) The sum of any 4 consecutive integers is divisible by 4.

(c) The product of any 3 consecutive integers is divisible by 6.

Video Solution

2.4.2 Critique the following proof. If the proof adequately demonstrates why the statement is true, explain why. Otherwise, identify any errors and explain how to correct them.

Theorem 2.36. *If x is a positive real number, then $x > 1$ if and only if $1/x < 1$.*

Proof. Suppose $1/x < 1$. Since x is positive, multiplying both sides of this inequality by x does not reverse the inequality and we obtain $1 < x$. ■

Video Solution

Exercises

2.4.1 Prove or disprove the following conjectures.

(a) There is an even integer which can be expressed as the sum of three even integers.

(b) Every even integer can be expressed as the sum of three even integers.

(c) There is an odd integer which can be expressed as the sum of two odd integers.

(d) Every odd integer can be expressed as the sum of three odd integers.

To get a feel about whether a claim is true or false, try out some examples. If you believe a claim is false, provide a specific counterexample. If you believe a claim is true, give a formal proof.

2.4.2 Let P be the proposition: 'Every positive integer is divisible by thirteen.'

(a) Write P using quantifiers.

(b) What is the negation of P?

(c) Is P true or false? Prove your assertion.

2.4.3 (a) Prove or disprove: There exist integers m and n such that $2m - 3n = 15$.

(b) Prove or disprove: There exist integers m and n such that $6m - 3n = 11$.

2.4.4 Prove or disprove: There exist a line L in $\mathbb{R}^2$ such that, for all points $A, B \in \mathbb{R}^2$, we have A, B lie on L.

2.4.5 Prove that between any two distinct rational numbers there exists another rational number.

2.4.6 Consider the statement:

For any non-zero rational number r and any irrational number t, rt is irrational.

(a) Translate this statement into logic using quantifiers and propositional functions.

(b) Prove the statement.

2.4.7 Let p be an odd integer. Prove that $x^2 - x - p = 0$ has no *integer* solutions.

2.4.8 Prove: For every positive integer n, $n^2 + n + 3$ is an odd integer greater than or equal to 5.
There are two claims here: $n^2 + n + 3$ is odd, and $n^2 + n + 3 \geq 5$.

2.4.9 In this question, you should use the following definition of the rational numbers.

Definition. A real number x is *rational* if it may be written in the form $x = \frac{p}{q}$ where p is an integer and q is a positive integer. x is *irrational* if it is not rational.

Prove or disprove the following conjectures.

Conjecture (1). *If x and y are real numbers such that $3x + 5y$ is irrational, then at least one of x and y is irrational.*

Conjecture (2). *If x and y are rational numbers, then $3x + 4xy + 2y$ is rational.*

Conjecture (3). *If x and y are irrational numbers, then $3x + 4xy + 2y$ is irrational.*

2.4.10 *(Snake-like integers)* Let's say that an integer y is *Snake-like* if and only if there is some integer k such that $y = (6k)^2 + 9$.

(a) Give three examples and three non-examples of Snake-like integers.

(b) Given $y \in \mathbb{Z}$, compute the negation of the statement, 'y is Snake-like.'

(c) Show that every Snake-like integer is a multiple of 9.

(d) Show that the statements, 'n is Snake-like,' and, 'n is a multiple of nine,' are not equivalent.

2.4.11 Prove that it is never the case that $x^2 = 2y^2$ for integers x and y.

2.4.12 Here is an alternative argument that $\sqrt{2}$ is irrational. Suppose that $\sqrt{2} = \frac{m}{n}$ where $m, n \in \mathbb{N}$. This time we don't assume that m, n have no common factors.

(a) m, n satisfy the equation $m^2 = 2n^2$. Prove that there exist positive integers m_1, n_1 which satisfy the following three conditions:

$$m_1^2 = 2n_1^2, \qquad m_1 < m, \qquad n_1 < n.$$

(b) Show that there exist two sequences of decreasing positive integers $m > m_1 > m_2 > \cdots$ and $n > n_1 > n_2 > \cdots$ which satisfy $m_i^2 = 2n_i^2$ for all $i \in \mathbb{N}$.

(c) Is it possible to have an infinite sequence of decreasing *positive* integers? Why not? Show that we obtain a contradiction and thus conclude that $\sqrt{2} \notin \mathbb{Q}$.

This is an example of the *method of infinite descent*, which is very important in number theory.

2.4.13 The real numbers have the *Archimedean property*, that is, for any positive real numbers x and y, there exists a positive integer n such that $nx > y$. Use this fact to show that there do not exist any positive real numbers which are less than $1/n$ for all positive integers n.

2.4.14 Consider the following proof of the fact that every real number is less than some positive integer:

> *Proof.* Consider a real number x. For example, $x = 19.7$. Then $x < 20$ and 20 is a positive integer. ■

What is wrong with this proof? Give a correct proof. [Hint: use the previous exercise.]

2.4.15 Here is an extension of question 5. Let $\lceil x \rceil$, the *ceiling* of x, denote the smallest integer greater than or equal to x. E.g. $\lceil 3.2 \rceil = 4$, $\lceil 7 \rceil = 7$ and $\lceil -8.4 \rceil = -8$.

(a) Suppose that x and y are real numbers with $x < y$. Use the ceiling function to show that there exists a positive integer n for which $n(y - x) > 1$.

(b) Prove or disprove: $\forall x, y \in \mathbb{R}$ with $x < y$, $\exists m, n \in \mathbb{Z}$ for which $nx < m < ny$.

(c) Use parts (a) and (b) to prove that between any two real numbers there exists a rational number.

(d) (Hard) Is it true that between any two real numbers there exists an *irrational* number? If so, prove it.

2.4.16 Suppose that x, y, z are real numbers such that $x + y + z = 1$. Prove

$$(1 - x)(1 - y)(1 - z) \geq 8xyz.$$

[Hint: find a way to apply the AM-GM inequality.]

2.4.17 You are given the following facts.

(a) All polynomials are continuous.

(b) (Intermediate Value Theorem) If f is continuous on $[a, b]$ and L lies between $f(a)$ and $f(b)$, then $f(x) = L$ for some $x \in (a, b)$.

(c) If $f'(x) > 0$ on an interval, then f is an increasing function.

Use these facts to give a formal proof that $x^{17} + 12x^3 + 13x + 3 = 0$ has *exactly one solution* x, and that x lies in the interval $(-1, 0)$.

2.4.18 (Hard) This question uses Definition 2.19.

(a) Prove, directly from the definition, that $f(x) = x^2$ is continuous at $x = 0$. *If you are given* $\epsilon > 0$, *what should* δ *be?*

(b) Prove that $g(x) = \begin{cases} 1 + x & \text{if } x \geq 0, \\ x & \text{if } x < 0, \end{cases}$ is discontinuous at $x = 0$.

(c) (Very hard) Let $h(x) = \begin{cases} x & \text{if } x \text{ is rational,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$ Prove that f is continuous *only* at $x = 0$.

2.4.19 (Hard) In this question we prove Rolle's Theorem from calculus:

> Suppose f is continuous on $[a, b]$, differentiable on (a, b), and $f(a) = f(b) = 0$, then $\exists c \in (a, b)$ such that $f'(c) = 0$.

As you work through the question, think about where the hypotheses are used and why we need them.

(a) Recall the Extreme Value Theorem. The function f is continuous on $[a, b]$, so f is bounded and attains its bounds. Otherwise said,

$$\exists m, M \in [a, b] \text{ such that } \forall x \in [a, b] \text{ we have } f(m) \leq f(x) \leq f(M).$$

Suppose that $f(m) = f(M)$. Why is the conclusion of Rolle's Theorem obvious in this case?

(b) Now suppose that $f(m) \neq f(M)$. Argue that *at least one* of the following cases holds:

$$f(M) > 0 \quad \text{or} \quad f(m) < 0.$$

(c) Without loss of generality, we may assume that $f(M) > 0$. By considering the function $-f$, explain why.

(d) Assume $f(M) > 0$. Then $M \neq a$ and $M \neq b$. Consider the difference quotient,

$$\frac{f(M+h) - f(M)}{h}.$$

Show that if $0 < |h| < \min\{M - a, b - M\}$ then the difference quotient is well-defined (exists and makes sense).

(e) Suppose that $0 < h < b - M$. Show that

$$\frac{f(M+h) - f(M)}{h} \leq 0.$$

How do we know that $L^+ := \lim\limits_{h \to 0^+} \frac{f(M+h)-f(M)}{h}$ exists? What can you conclude about L^+?

(f) Repeat part (d) for $L^- := \lim\limits_{h \to 0^-} \frac{f(M+h)-f(M)}{h}$.

(g) Conclude that $L^+ = L^- = 0$. Why have we completed the proof?

3 Sets and Functions

Sets are the fundamental building blocks of mathematics. In the sub-discipline of Set Theory, mathematicians define all basic notions, including number, addition, function, etc., purely in terms of sets. In such a system it can take over 100 pages of discussion to *prove* that $1 + 1 = 2$! We will not be anything like so rigorous. Indeed, before one can accept that such formality has its place in mathematics, a level of familiarity with sets and their basic operations is necessary.

3.1 Set Notation and Describing a Set

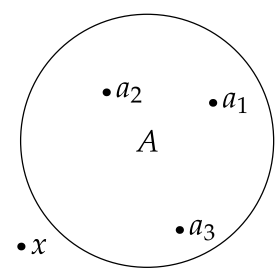
We start with a naïve notion: a set is a collection of objects.[8]

Definition 3.1. A **set** is a collection of objects.

If x is an object in a set A, we write $x \in A$ and say that x is an **element** or **member** of A. On the other hand, if x is a member of some other set B, but not of A, we write $x \notin A$.

Sets A and B are described as **equal**, written $A = B$, if they have exactly the same elements.

When thinking abstractly about sets, you may find *Venn diagrams* useful. A set is visualized as a region in the plane and, if necessary, members of the set can be thought of as dots in this region. This is most useful when one has to think about multiple, possibly over-lapping, sets. The graphic represents a set A with at least three elements a_1, a_2, a_3. The element x does not lie in A.



Notation and Conventions

We use capital letters for sets, e.g. A, B, C, S, and lower-case letters for elements. It is conventional, though not required, to denote an abstract element of a set by the corresponding lower-case letter: thus $a \in A$, $b \in B$, etc.
Curly brackets $\{\ ,\ \}$ are used to bookend the elements of a set: for instance, if we wrote

$$S = \{3, 5, f, \alpha, \beta\}$$

then we'd say, 'S is the set whose elements are 3, 5, f, α and β.'
The order in which we list the elements of a set is irrelevant, thus

$$S = \{\beta, f, 5, \alpha, 3\} = \{f, \alpha, 3, \beta, 5\}.$$

Listing all the elements in such a fashion is known as *roster notation.*

By contrast, *set-builder notation* describes the elements of a set by starting with a larger set and restricting to those elements which satisfy some property. The symbols $|$ or $:$ are used as a short-hand

[8] For this course, our notion is enough. It eventually became clear that some collections of objects cannot be considered sets, and the search for a completely rigorous definition began; thus was Axiomatic Set Theory born.

for 'such that.' Which symbol you use depends partly on taste, although the context may make one clearer to read.[9] For example, if $S = \{3, 5, f, \alpha, \beta\}$ is the set defined above, we could write,

$$\{s \in S \mid s \text{ is a Greek letter}\} = \{s \in S : s \text{ is a Greek letter}\} = \{\alpha, \beta\}$$

We would read: 'The set of elements s in S such that s is a Greek letter is the set $\{\alpha, \beta\}$.'

More generally, if S is a set and P is a propositional function whose domain is S, then we can define a new set

$$A := \{s \in S : P(s) \text{ is true}\}$$

Example. Let $A = \{2, 4, 6\}$ and $B = \{1, 2, 5, 6\}$. There are many options for how to write A and B in set-builder notation. For example, we could write

$$A = \{2n \in \mathbb{Z} : n = 1,\ 2 \text{ or } 3\} \quad \text{and} \quad B = \{n \in \mathbb{Z} \mid 1 \le n \le 6 \text{ and } n \ne 3, 4\}.$$

We now practice the opposite skill by converting five sets from set-builder to roster notation.

$$S_1 = \{a \in A : a \text{ is divisible by } 4\} = \{4\}$$

$$S_2 = \{b \in B : b \text{ is odd}\} = \{1, 5\}$$

$$S_3 = \{a \in A \mid a \in B\} = \{2, 6\}$$

$$S_4 = \{a \in A : a \notin B\} = \{4\}$$

$$S_5 = \{b \in B \mid b \text{ is odd and } b - 1 \in A\} = \{5\}$$

Take your time getting used to this notation. Can you find an alternative description in set-builder notation for the sets $S_1, \ldots, S_5$ above? It is *crucial* that you can translate between various descriptions of a set or you won't be able to read much mathematics!

Sets of Numbers

Common sets of numbers are written in the Blackboard Bold typeface.

$\mathbb{N} = \mathbb{Z}^+ = \text{natural numbers} = \{1, 2, 3, 4, \ldots\}$

$\mathbb{N}_0 = \{0, 1, 2, 3, 4, \ldots\}$

$\mathbb{Z} = \text{integers} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$

$\mathbb{Q} = \text{rational numbers} = \{\frac{m}{n} : m \in \mathbb{Z} \text{ and } n \in \mathbb{N}\} = \{\frac{a}{b} : a, b \in \mathbb{Z} \text{ and } b \ne 0\}$

$\mathbb{R} = \text{real numbers}$

$\mathbb{R} \setminus \mathbb{Q} = \text{irrational numbers}$ (read '$\mathbb{R}$ minus $\mathbb{Q}$')

$\mathbb{C} = \text{complex numbers} = \{x + iy : x, y \in \mathbb{R},\ \text{where } i = \sqrt{-1}\}$

$\mathbb{Z}_{\ge n} = \text{integers} \ge n = \{n, n+1, n+2, n+3, \ldots\}$

[9]See Choice of Notation, below.

$$n\mathbb{Z} = \text{multiples of } n = \{\ldots, -3n, -2n, -n, 0, n, 2n, 3n, \ldots\}$$

Where there are multiple choices of notation, we will tend to use the first in the list: for example $\mathbb{N}_0$ is preferred to $\mathbb{Z}_{\geq 0}$. The use of a subscript 0 to include zero and a superscript $\pm$ to restrict to positive or negative numbers is standard.

Examples. $7 \in \mathbb{Z}, \qquad \pi \in \mathbb{R}, \qquad \pi \notin \mathbb{Q}, \qquad \sqrt{-5} \in \mathbb{C}, \qquad -e^2 \in \mathbb{R}^-.$

There are often many different ways to represent the same set in set-builder notation. For example, the set of even numbers may be written in multiple ways: think about the English translations.

$$\begin{aligned} 2\mathbb{Z} &= \{2n \in \mathbb{Z} : n \in \mathbb{Z}\} && \text{(The set of integers of the form } 2n \text{ such that } n \text{ is an integer)} \\ &= \{n \in \mathbb{Z} : \exists k \in \mathbb{Z},\ n = 2k\} && \text{(The set of integers which are a multiple of 2)} \\ &= \{n \in \mathbb{Z} :\ 2 \,|\, n\} && \text{(The set of integers which are divisible by 2)} \end{aligned}$$

Can you find any other ways to describe the even numbers using basic set notation?

The notation $n\mathbb{Z}$ is most commonly used when n is a natural number, but it can also be used for other n. For example

$$\tfrac{1}{2}\mathbb{Z} = \left\{\tfrac{1}{2}x : x \in \mathbb{Z}\right\} = \left\{m, m + \tfrac{1}{2} : m \in \mathbb{Z}\right\}$$

is the set of multiples of $\frac{1}{2}$ (comprising the integers and half-integers). The notation can also be extended: for example $2\mathbb{Z} + 1$ would denote the odd integers.

Aside. Choice of Notation

The notations $|$ and $:$ for 'such that' give you leeway in case one these symbols is being used to mean something else. For example, the final expression (above) for the even numbers is much cleaner than the alternative

$$2\mathbb{Z} = \{n \in \mathbb{Z} \mid 2 \mid n\}.$$

In other situations the opposite is true. In Section 3.4 we shall consider functions. If you recall the concept of an even function from calculus, we could denote the set of such as

$$\{f : \mathbb{R} \to \mathbb{R} : \forall x\, f(x) = f(-x)\} \qquad \text{or} \qquad \{f : \mathbb{R} \to \mathbb{R} \mid \forall x\, f(x) = f(-x)\}.$$

In this case the latter notation is clearly superior.

Examples. 3.1.1 Write the set $A = \{x \in \mathbb{R} : x^2 + 3x + 2 = 0\}$ in roster notation.

We are looking for the set of all real number solutions to the quadratic equation $x^2 + 3x + 2 = 0$. A simple factorization tells us that $x^2 + 3x + 2 = (x+1)(x+2)$, whence $A = \{-1, -2\}$.

3.1.2 Use the set $B = \{0, 1, 2, 3, \ldots, 24\}$ to describe $C = \{n \in \mathbb{Z} : n^2 - 3 \in B\}$ in roster notation.

We see that

$$n^2 - 3 \in B \iff n^2 \in \{3, 4, 5, \ldots, 25, 26, 27\}$$

Since n must be an integer in order to be an element of C, it follows that

$$C = \{\pm 2, \pm 3, \pm 4, \pm 5\}.$$

3.1.3 It is often harder to convert from roster to set-builder notation, as you might be required to spot a pattern, and many choices could be available. For example, if

$$D = \left\{\frac{1}{6}, \frac{1}{20}, \frac{1}{42}, \frac{1}{72}, \frac{1}{110}, \frac{1}{156}, \ldots\right\},$$

you might consider it reasonable to write

$$D = \left\{\frac{1}{2n(2n+1)} : n \in \mathbb{N}\right\}.$$

Of course the ellipses $(\ldots)$ might not indicate that the elements of the set continue in the way you expect. For larger sets, the concision and clarity of set-builder notation makes it much preferred!

3.1.4 Are the following sets equal?

$$E = \{n^2 + 2 \in \mathbb{Z} : n \text{ is an odd integer}\}, \qquad F = \{n \in \mathbb{Z} : n^2 + 2 \text{ is an odd integer}\}.$$

It may help to first construct a table listing some of the values of $n^2 + 2$:

n	n^2	n^2+2
± 1	1	3
± 3	9	11
± 5	25	27
± 7	49	51
± 9	81	83
$\vdots$	$\vdots$	$\vdots$

The set E consists of those integers of the form $n^2 + 2$ where n is an odd integer. By the table,

$$E = \{3, 11, 27, 51, 83, \ldots\}.$$

On the other hand, F includes all those integers n such that n^2+2 is odd. It is easy to see that

$$n^2+2 \text{ is odd} \iff n^2 \text{ is odd} \iff n \text{ is odd.}$$

Thus F is simply the set of all odd integers:

$$F = \{\pm 1, \pm 3, \pm 5, \pm 7, \ldots\} = 2\mathbb{Z}+1.$$

Plainly the two sets are not equal.

Intervals

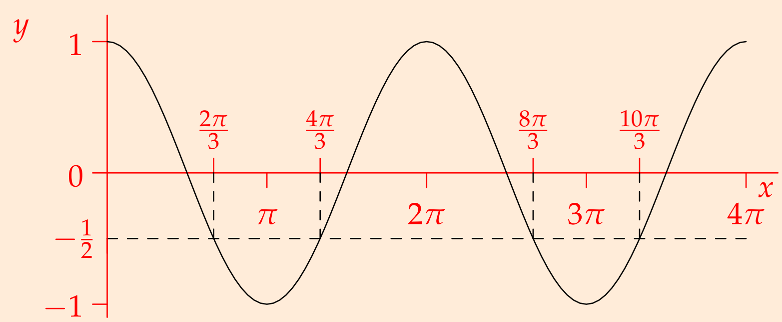
Interval notation is useful when discussing collections of *real numbers.* You should be familiar from calculus with the words *open* and *closed* with regard to intervals. For example,

$$(0,1) = \{x \in \mathbb{R} : 0 < x < 1\}, \qquad \text{(Open interval)}$$
$$[0,1] = \{x \in \mathbb{R} : 0 \le x \le 1\}, \qquad \text{(Closed interval)}$$
$$(0,1] = \{x \in \mathbb{R} : 0 < x \le 1\}. \qquad \text{(Half-open interval)}$$

When writing intervals with $\pm\infty$ use an open bracket at the infinite end(s): $[1,\infty) = \{x \in \mathbb{R} : x \ge 1\}$. This is since the symbols $\pm\infty$ do not represent real numbers and so are not members of any interval.

Example. Recall some basic trigonometry. Consider the set of solutions to the equation $\cos x = -\frac{1}{2}$ where x lies in the interval $[0,4\pi]$. This set can be written in set-builder and roster notation as

$$\left\{x \in [0,4\pi] : \cos x = -\frac{1}{2}\right\} = \left\{\frac{2\pi}{3}, \frac{4\pi}{3}, \frac{8\pi}{3}, \frac{10\pi}{3}\right\}$$



Cardinality and the Empty Set

Definition 3.2. A set A is **finite** if it contains a finite number of elements: this number is the set's **cardinality**, written $|A|$. If A contains infinitely many elements, it said to be an **infinite set**.

Examples. 3.1.1 Let $A = \{a, b, \alpha, \gamma, \sqrt{2}\}$, then $|A| = 5$.

3.1.2 Let $B = \Big\{4, \{1,2\}, \{3\}\Big\}$. It is important to note that the *elements/members* of B are 4, $\{1,2\}$ and $\{3\}$, two of which are themselves sets. Therefore $|B| = 3$. The set $\{1,2\}$ is an object in its own right, and can therefore be placed in a set along with other objects.[a]

[a] The fact that a set (containing objects) is also an object might seem confusing, but you should be familiar with the same problem in English. Consider the following sentences: 'UCI *are* constructing a laboratory' and 'UCI *is* constructing a laboratory.' In the first case we are thinking of UCI as a collection of individuals, in the latter case UCI is a single object. Opinions differ in various modes of English as to which is grammatically correct.

Cardinality is a very simple concept for finite sets. For infinite sets, such as the natural numbers $\mathbb{N}$, the concept of cardinality much more subtle. In Chapter 8 we will consider what cardinality means for infinite sets and meet several bizarre and fun consequences. For the present, cardinality only has meaning for finite sets.

To round things off we need a symbol to denote a set that contains nothing at all!

Axiom. There exists a set $\emptyset$ with no elements (cardinality zero: $|\emptyset| = 0$). We call $\emptyset$ the **empty set**.

There are many *representations* of the empty set. For example $\{x \in \mathbb{N} : x^2 + 3x + 2 = 0\}$ and $\{n \in \mathbb{N} : n < 0\}$ are both empty. Despite this, we will see in Theorem 3.5 that there is only one set with no elements, so that all representations actually denote the *same set* $\emptyset$. Note also that $|A| \in \mathbb{N}$ for any *finite non-empty* set A.

Aside. Axioms

An axiom is a basic assumption; something that we need in order to do mathematics, but cannot prove. This is the cheat by which mathematicians can be 100% sure that something is true: a result is proved based on the assumption of several axioms. With regard to the empty set axiom, it probably seems bizarre that we can assume the existence of some set that has nothing in it. Regardless, mathematicians have universally agreed that we need the empty set in order to do the rest of mathematics. The assumption that set-builder notation always defines a new set is another axiom.

Reading Questions

3.1.1 Which of the following describe the following set? Select all that apply.

$$\{0, 1, 2, 3, 4\}$$

(a) $\{x \in \mathbb{N}_0 : x \leq 4\}$
(b) $\{x \in \mathbb{Q} : x \in [0,4]\}$
(c) $\{x \in \mathbb{Z} : x \in [0,4)\}$
(d) $\{x \in \mathbb{Z} : x \in [0,4]\}$

3.1.2 What is the cardinality of the set $\{\text{cat}, \{1,2\}, 2\}$?

(a) 2
(b) 3
(c) 4
(d) it is an infinite set

3.1.3 True or False: An open interval contains its endpoints.

3.1.4 True or False: $\{1,2,3\} = \{3,1,2\}$.

3.1.5 Which of the following sets are empty?

(a) $\{x \in \mathbb{R} : x^2 < 0\}$
(b) $\{x \in \mathbb{R} : x^2 \leq 0\}$
(c) $\{x \in \mathbb{N} : x \in [0.5, 0.75)\}$
(d) $[1,1]$

Practice Problems

3.1.1 Write each of the following sets in roster notation (i.e. list their elements).

(a) $\{x \in \mathbb{R} : x^2 - 5x + 4 = -2\}$
(b) $\{x \in \mathbb{Q} : 2x \in \mathbb{Z}\}$
(c) $\{n^2 - 1 \in \mathbb{Z} : n \in \{-3, -1, 1, 3\}\}$
(d) $\{x \in 2\mathbb{Z} + 1 : x \in (0, 10]\}$

Video Solution

3.1.2 Write each of the following sets in set-builder notation.

(a) $\{\ldots, -8, -3, 2, 7, 12, 17, \ldots\}$
(b) $\{2, 3, 5, 7, 11, 13\}$
(c) $\{1, \frac{1}{4}, \frac{1}{16}, \frac{1}{64}, \ldots\}$

Video Solution

3.1.3 Let

$$A = \{0, \{0\}, \{1,2\}, \{0, \{1,3\}\}\}$$

Answer True or False for each of the following:

(a) $0 \in A$
(b) $\{0\} \in A$
(c) $1 \in A$
(d) $\{1\} \in A$
(e) $\{1,3\} \in A$
(f) $\{0, \{1,3\}\} \in A$

What is the cardinality of A?

Video Solution

Exercises

3.1.1 Describe the following sets in roster notation: that is, list their elements.

(a) $\{x \in \mathbb{N} : x^2 \leq 5x\}$.
(b) $\{x^2 \in \mathbb{R} : x^2 - 3x + 2 = 0\}$.
(c) $\{n \in \{-4, -3, -2, -1, 0, 1, \ldots, 21\} : 4 \,|\, n^2\}$ (*does* : *or* | *denote the condition?*)
(d) $\{x \in \frac{1}{2}\mathbb{Z} : 0 \leq x \leq 4 \text{ and } 4x^2 \in 2\mathbb{Z} + 1\}$

3.1.2 Describe the following sets in set-builder notation (*look for a pattern*).

(a) $\{\ldots, -3, 0, 3, 6, 9, \ldots\}$
(b) $\{-3, 1, 5, 9, 13, \ldots\}$
(c) $\{1, \frac{1}{3}, \frac{1}{7}, \frac{1}{15}, \frac{1}{31}, \ldots\}$

3.1.3 Let

$$\begin{aligned} A &= \emptyset \\ B &= \{A\} \\ C &= \{\{A\}\} \\ D &= \{A, \{0\}, \{0, 1\}\} \end{aligned}$$

Answer True or False for each of the following:

(a) $0 \in A$
(b) $A \in B$
(c) $A \in C$
(d) $B \in C$
(e) $A \in D$
(f) $B \in D$
(g) $0 \in D$
(h) $\{0\} \in D$
(i) $\{1\} \in D$

3.1.4 Each of the following sets of real numbers is a single interval. Determine the interval.

(a) $\{x \in \mathbb{R} : x > 5 \text{ and } x \leq 19\}$
(b) $\{x \in \mathbb{R} : x \not\leq 5 \text{ or } x \leq 19\}$
(c) $\{x^2 \in \mathbb{R} : x \neq 0\}$
(d) $\{x \in \mathbb{R}^- : x^2 \geq 16 \text{ and } x^3 \leq 27\}$

3.1.5 Can you describe the set $\{x \in \mathbb{Z} : -3 \leq x < 77\}$ in interval notation? Why/why not?

3.1.6 Compare the sets $A = \{3x \in \mathbb{Z} : x \in 2\mathbb{Z}\}$ and $B = \{x \in \mathbb{Z} : 6 \mid (x - 12)\}$. Are they equal?

3.1.7 What is the cardinality of the following set? What are the elements?

$$\Big\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\Big\}.$$

3.1.8 Let $A = \{1,2,3,4\}$, and let B be the set

$$B = \Big\{\{x, y\} : x, y \in A\Big\}.$$

(a) Describe B in roster notation.

(b) Now compute the cardinality of the sets

$$C = \Big\{\{x, \{y\}\} : x, y \in A\Big\}$$

and

$$D = \Bigg\{\Big\{\{x, \{y\}\} : x, y \in A\Big\}\Bigg\}.$$

Compare them to $|B|$.

3.1.9 Prove or disprove the following conjectures.

(a) There exists $x \in \mathbb{R} \setminus \mathbb{Q}$ such that $x^2 \in \mathbb{Q}$.

(b) For all $x \in \mathbb{R} \setminus \mathbb{Q}$ we have $x^2 \in \mathbb{Q}$.

3.2 Subsets

In this section we consider the most basic manner in which two sets can be related.

Definition 3.3. If A and B are sets such that every element of A is also an element of B, then we say that A is a **subset** of B and write $A \subseteq B$.
A is a **proper subset** of B if it is a subset which is not equal. This can be written $A \subsetneq B$.[a]

[a] We will religiously stick to this notation. When reading other texts, note that some authors prefer $A \subset B$ for proper subset. Others use $\subset$ for any subset, whether proper or not.

The concept of subset provides us with an extremely important characterization of equality.

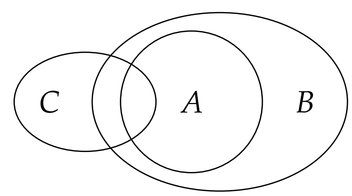
Theorem 3.4. *Two sets are equal if and only if they are each a subset of the other. Equivalently*

$$A = B \iff A \subseteq B \text{ and } B \subseteq A.$$

Proof. Recall that two sets A and B are equal if and only if they have the same elements. But this is if and only if every element of A is also an element of B and vice versa. ∎

You will often need to *prove* that two sets are equal: showing that each is a subset of the other is a very common way to accomplish this.

Venn diagrams are particularly useful for visualizing subset relations. The graphic on the right depicts three sets A, B, C: it should be clear that the only valid subset relation between the three is $A \subseteq B$.



Set-builder notation implicitly uses the concept of subset: the notation $X = \{y \in Y : P(y)\}$ describes a set X as the subset of some other set Y, all of whose elements satisfy the property $P(y)$. The previous section contained many examples that were subsets of the set of real numbers $\mathbb{R}$. Here are some other examples of subsets.

Examples. 3.2.1 $\mathbb{N} = \{n \in \mathbb{Z} : n > 0\}$. This is clearly a subset of $\mathbb{Z}$.

3.2.2 $\{x \in \mathbb{R} : x^2 - 1 = 0\} \subseteq \{y \in \mathbb{R} : y^2 \in \mathbb{N}\}$.
To make sense of this relationship, convert to roster notation: we obtain

$$\{-1, 1\} \subseteq \{\pm\sqrt{1}, \pm\sqrt{2}, \pm\sqrt{3}, \pm\sqrt{4}, \ldots\}.$$

3.2.3 If m and n are positive integers, then $m\mathbb{Z} \subseteq n\mathbb{Z} \iff n \mid m$. Make sure you're comfortable with this! For example, $4\mathbb{Z} \subseteq 2\mathbb{Z}$ since every multiple of 4 is also a multiple of 2.

Here we collect several results relating to subsets.

Theorem 3.5. *3.2.1 If $|A| = 0$, then $A = \emptyset$* *(Uniqueness of the empty set)*

3.2.2 For any set A, we have $\emptyset \subseteq A$ and $A \subseteq A$ *(Trivial and non-proper subsets)*

3.2.3 If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$ *(Transititvity of subsets)*

Proof. 3.2.1 Let A be a set with cardinality zero, i.e., with no elements. $\emptyset$ has no members, therefore $\emptyset \subseteq A$ is trivial: there is nothing to check to see that all elements of $\emptyset$ are also elements of A! The argument for $A \subseteq \emptyset$ is identical. By Theorem 3.4 we see that $A = \emptyset$.

3.2.2 Let A be any set. $\emptyset \subseteq A$ follows by the argument in 1. To prove that $A \subseteq A$ we must show that all elements of A are also elements of A. But this is completely obvious!

3.2.3 Assume that A is a subset of B and that B is a subset of C. We must show that all elements of A are also elements of C. Let $a \in A$. Since $A \subseteq B$ we know that $a \in B$. Since $B \subseteq C$ and $a \in B$, we conclude that $a \in C$. This shows that every element of A belongs to C. Hence $A \subseteq C$. ■

As a final observation, to which we will return in Theorem 3.15 and in Chapter 8, your intuition should tell you that, for finite sets, subsets have smaller cardinality:

$$A \subseteq B \implies |A| \leq |B|.$$

More generally, consider replacing the terms in Theorem 3.5 according to the following table:

$\subseteq$	$\leq$
$\emptyset$	0
sets A, B, C	non-negative integers
cardinality	absolute value

The results should seem completely natural! Recognizing the similarities between a new concept and a familiar one, essentially spotting patterns, is perhaps the most necessary skill in mathematics.

Reading Quiz

3.2.1 True or False: Every set has a proper subset.

3.2.2 True or False: $\{\mathbb{R}\} \subseteq \{\{\mathbb{R}\}\}$.

3.2.3 How many subsets does the set $A = \{0, 1\}$ have?

(a) 1
(b) 2
(c) 3
(d) 4

3.2.4 $A = B$ if and only if

(a) $A \subseteq B$
(b) $A \subseteq B$ and if $x \notin A$, then $x \notin B$.
(c) $B \subseteq A$
(d) $A \subsetneq B$ and B is finite.

Practice Problems

3.2.1 Suppose $A \subseteq B \subseteq C$ and $A = C$. Show $A = B$ and $B = C$.
Video Solution

3.2.2 Let

$$A = \{0, \{0\}, \{1,2\}, \{0, \{1,3\}\}\}$$

Answer True or False for each of the following:

(a) $\emptyset \subsetneq A$
(b) $\{0\} \subseteq A$
(c) $\{\{0\}\} \subseteq A$
(d) $\{1,2\} \subseteq A$
(e) $\{\{1,2\}\} \subseteq A$
(f) $\{0, \{0\}, \{1,3\}\} \subseteq A$
(g) $A \subsetneq A$

Video Solution

Exercises

3.2.1 Let A, B, C, D be the following sets.

$$A = \{-4, 1, 2, 4, 10\}$$
$$B = \{m \in \mathbb{Z} : |m| \leq 12\}$$
$$C = \{n \in \mathbb{Z} : 3 \mid (n^2 - 1)\}$$
$$D = \{t \in \mathbb{Z} : t^2 + 3 \in [4, 20)\}$$

Of the 12 possible subset relations $A \subseteq B,\ A \subseteq C, \ldots D \subseteq C$, which are true and which false?

3.2.2 (a) Let $A = \{x \in \mathbb{R} : x^3 + x^2 - x - 1 = 0\}$ and $B = \{x \in \mathbb{R} : x^4 - 5x^2 + 4 = 0\}$. Are either of the relations $A \subseteq B$ or $B \subseteq A$ true? Explain.
(b) Let $A = \{4n \colon n \in \mathbb{Z}\}$ and $B = \{k \in \mathbb{Z} \colon 3k + 5 \text{ is odd}\}$. Prove or disprove: $A \subseteq B$.

3.2.3 (a) Order the following sets according to which are subsets of which:

$$\mathbb{R}, \quad \mathbb{Z}, \quad \mathbb{N}_0, \quad \mathbb{N}, \quad \mathbb{Q}, \quad \mathbb{C}$$

(b) (Hammack's *Book of Proof*, Section 1.3, Exercise 14) True or False? $\mathbb{R}^2 \subseteq \mathbb{R}^3$. Explain your answer.

3.2.4 For which values of $x > 0$ is the following claim true?

$$[0, x] \subseteq [0, x^2]$$

Prove your assertion.

3.2.5 (a) Write down all *proper* subsets of $\{1, 2, 3\}$

(b) (Hammack's *Book of Proof*, Section 1.3, Exercise 6) List all subsets of $\{\mathbb{R}, \mathbb{Q}, \mathbb{N}\}$.

3.2.6 Write down all subsets of $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$.

3.2.7 Let $A = \{1, 2, \{1, 2\}, \{3\}\}$ and $B = \{1, 2\}$. Answer True or False for each of the following:

(a) $B \in A$

(b) $B \subseteq A$

(c) $3 \in A$

(d) $\{3\} \subseteq A$

(e) $\{3\} \in A$

(f) $\emptyset \subseteq A$

(g) $\emptyset \in A$

3.2.8 Let $A = \{0, 2, 4, 6, 8, 10\}$. Write the set $\{X \subseteq A : |X| = 2\}$ in roster notation.

3.2.9 Suppose $A \subseteq B \subseteq C \subseteq A$. Show $A = B = C$.

3.2.10 Fill in the blanks in the following proof of the fact that $A \subsetneq B$ and $B \subsetneq C$ implies $A \subsetneq C$.

Proof. Recall that $X \subsetneq Y$ means $X \subseteq Y$ *and* __________. So $A \subsetneq B$ and $B \subsetneq C$ means __________ and __________ which by Theorem 3.5 (insert ref here) gives $A \subseteq C$. All that remains to show is that A_____C. But if this is not true, then __________. But this would mean $A = B$ because ____________________, contradicting the fact that _______________. Thus we conclude that $A \subsetneq C$. ■

3.2.11 For the following proof sketch, determine the result that is being proved, and then turn the sketch into a formal proof.

Proof. ($\Leftarrow$) If $m = nk$ $\exists k \in \mathbb{Z}$, then $\forall x \in m\mathbb{Z}$, we have $x = mj = (nk)j \in n\mathbb{Z}$. Thus $m\mathbb{Z} \subseteq n\mathbb{Z}$.

($\Rightarrow$) Suppose $m\mathbb{Z} \subseteq n\mathbb{Z}$, then $m \in n\mathbb{Z}$, i.e. $n \mid m$. ■

3.2.12 Given $A \subseteq \mathbb{Z}$ and $x \in \mathbb{Z}$, we say that x is A-mirrored if and only if $-x \in A$. We also define:

$$M_A := \{x \in \mathbb{Z} \colon x \text{ is } A\text{-mirrored}\}.$$

(a) What is the negation of 'x is A-mirrored.'

(b) Find M_B for $B = \{0, 1, -6, -7, 7, 100\}$.

(c) Assume that $A \subseteq \mathbb{Z}$ is closed under addition (i.e., for all $x, y \in A$, we have $x + y \in A$). Show that M_A is closed under addition.

(d) In your own words, under which conditions is $A = M_A$?

3.2.13 Define the set [1] by:

$$[1] = \{x \in \mathbb{Z} \colon 5 \mid (x-1)\}.$$

(a) Describe the set [1] in roster notation.

(b) Compute the set $M_{[1]}$, as defined in Exercise 3.2.12

(c) Are the sets [1] and $M_{[1]}$ equal? Prove/Disprove.

(d) Now consider the set $[10] = \{x \in \mathbb{Z} \colon 5 \mid (x - 10)\}$. Are the sets [10] and $M_{[10]}$ equal? Prove/Disprove.

3.2.14 (a) Give a formal proof of the fact that $A \subseteq B \implies |A| \leq |B|$ for finite sets. *Resist the temptation to look at Theorem 3.15: it is far more technical than you need for this!*

(b) Explain why $|A| \leq |B| \;\not\Longrightarrow\; A \subseteq B$.

3.2.15 Consider the set $A = \{a, b, c, d\}$.

(a) How many subsets of A are there of cardinality 0, 1, 2, 3, and 4, respectively. Do you notice any patterns?

(b) Completely expand the polynomial $(1 + x)^4$. What do you notice about the coefficients?

3.2.16 Let A be a set. We define the *power set* of A, $\mathcal{P}(A)$, to be the set of all subsets of A:

$$\mathcal{P}(A) = \{X : X \subseteq A\}.$$

(a) Compute $\mathcal{P}(A)$ where $A = \{1, a, 5\}$.

(b) Prove or disprove: for any set A, $A \in \mathcal{P}(A)$.

(c) Prove or disprove: for any set A, $A \subseteq \mathcal{P}(A)$.

(d) Give an explicit example of a set A such that $A \neq \emptyset$ and $A \subseteq \mathcal{P}(A)$.

3.3 Unions, Intersections, and Complements

In this section we construct new sets from old, modeled precisely on the logical concepts of *and*, *or*, and *not*. For the duration of this section, suppose that $\mathcal{U}$ is some *universal set*, of which every set mentioned subsequently is a subset.[10]

First we consider the set construction modeled on *not*.

Definition 3.6. Let $A \subseteq \mathcal{U}$ be a set. The **complement** of A is the set

$$A^{\mathsf{C}} = \{x \in \mathcal{U} : x \notin A\}.$$

This can also be written $\mathcal{U} \setminus A$, $\mathcal{U} - A$, A', or $\overline{A}$.

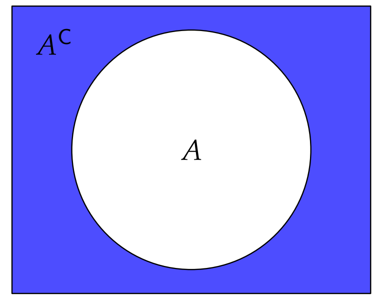
The Venn diagram is drawn on the right: A is represented by a circular region, while the rectangle represents the universal set $\mathcal{U}$. The complement A^{C} is the blue shaded region.

If $B \subseteq \mathcal{U}$ is some other set, then the **complement of** A **relative** B is

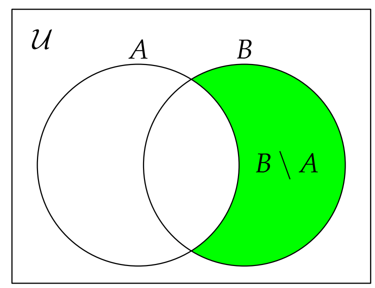
$$B \setminus A = \{x \in B : x \notin A\}.$$

The set $B \setminus A$ is also called B **minus** A. For its Venn diagram, we represent A and B as overlapping circular regions. The complement $B \setminus A$ is the green shaded region.

Note that $A^{\mathsf{C}} = \mathcal{U} \setminus A$, so that the two definitions correspond.



A^{C}: everything not in A



$B \setminus A$: everything in B but not in A

Example. Let $\mathcal{U} = \{1,2,3,4,5\}$, $A = \{1,2,3\}$, and $B = \{2,3,4\}$. Then

$$A^{\mathsf{C}} = \{4,5\}, \qquad B^{\mathsf{C}} = \{1,5\}, \qquad B \setminus A = \{4\}, \qquad A \setminus B = \{1\}.$$

Now we construct sets based on *or* and *and*.

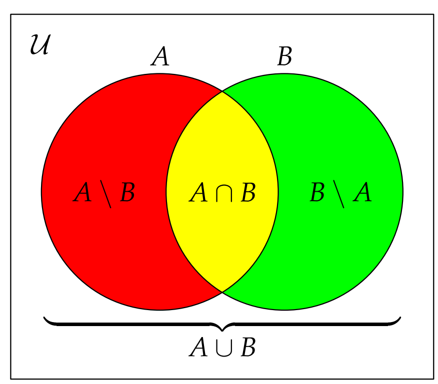
Definition 3.7. The **union** of A and B is the set

$$A \cup B = \{x \in \mathcal{U} : x \in A \text{ or } x \in B\}.$$

The **intersection** of A and B is the set

$$A \cap B = \{x \in \mathcal{U} : x \in A \text{ and } x \in B\}.$$

We say that A and B are **disjoint** if $A \cap B = \emptyset$.



In the Venn diagram, the sets A and B are again depicted as overlapping circles. Although it doesn't

[10] This is necessary so that the definitions to come made using set-builder notation really define sets.

constitute a proof, the diagram makes it clear that

$$A = (A \setminus B) \cup (A \cap B) \quad \text{and} \quad B = (B \setminus A) \cup (A \cap B).$$

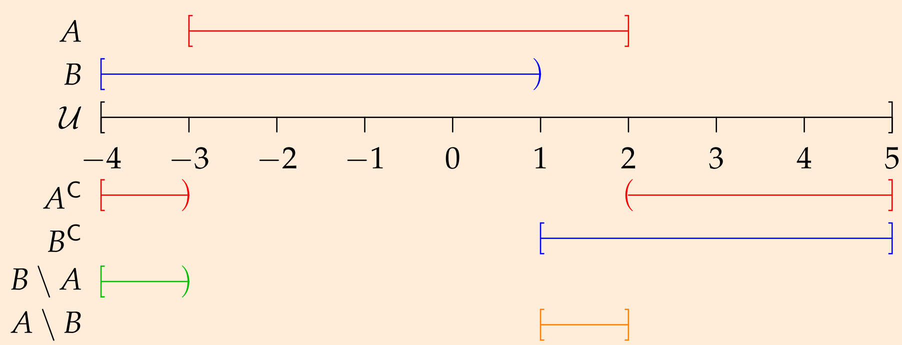
'Or' is used in the logical sense: $A \cup B$ is the collection of all elements that lie in A, in B, or in both. Now observe the notational pattern: $\cup$ looks very similar to the logic symbol $\vee$ from Chapter 2. The symbols $\cap$ and $\wedge$ are also similar. This should help you remember which symbol to use when!

Examples. 3.3.1 Let $\mathcal{U} = \{\text{fish, dog, cat, hamster}\}$, $A = \{\text{fish, cat}\}$, and $B = \{\text{dog, cat}\}$. Then,

$$A \cup B = \{\text{fish, dog, cat}\}, \qquad A \cap B = \{\text{cat}\}.$$

3.3.2 Using interval notation, let $\mathcal{U} = [-4, 5]$, $A = [-3, 2]$, and $B = [-4, 1)$. Then

$$A^{\mathsf{C}} = [-4, -3) \cup (2, 5], \qquad B^{\mathsf{C}} = [1, 5], \qquad B \setminus A = [-4, -3), \qquad A \setminus B = [1, 2].$$



3.3.3 Let $A = (-\infty, 3)$ and $B = [-2, \infty)$ in interval notation. Then $A \cup B = \mathbb{R}$ and $A \cap B = [-2, 3)$.

We didn't mention the universal set in the final example, though it seems reasonable to assume that $\mathcal{U} = \mathbb{R}$. In practice $\mathcal{U}$ is rarely made explicit, and is often assumed to be the smallest suitable uncomplicated set. When dealing with sets of real numbers this typically means $\mathcal{U} = \mathbb{R}$. In other situations $\mathcal{U} = \mathbb{Z}$ or $\mathcal{U} = \{0, 1, 2, 3, \ldots, n-1\}$ might be more appropriate.

The next theorem comprises the basic rules of set algebra.

Theorem 3.8. *Let A, B, C be sets. Then:*

3.3.1 $\emptyset \cup A = A$ *and* $\emptyset \cap A = \emptyset$.

3.3.2 $A \cap B \subseteq A \subseteq A \cup B$.

3.3.3 $A \cup B = B \cup A$ *and* $A \cap B = B \cap A$.

3.3.4 $A \cup (B \cup C) = (A \cup B) \cup C$ *and* $A \cap (B \cap C) = (A \cap B) \cap C$.

3.3.5 $A \cup A = A \cap A = A$.

3.3.6 $A \subseteq B \implies A \cup C \subseteq B \cup C$ *and* $A \cap C \subseteq B \cap C$.

You should be able to prove each of these properties directly from Definitions 3.3 and 3.7. Don't memorize the proofs: with a little practice working with sets, each of these results should feel completely obvious. It is more important that you are able to *vizualize* the laws using Venn diagrams. A Venn diagram does not constitute a formal proof, though it is extremely helpful for clarification. Here we prove only second result: think about how the Venn diagram in Definition 3.7 illustrates the result. Some of the other proofs are in the Exercises.

Proof of 2. There are two results here: $A \cap B \subseteq A$ and $A \subseteq A \cup B$. We show each separately, along with some of our reasoning.

Suppose that $x \in A \cap B$.	(Must show $x \in A \cap B \Rightarrow x \in A$)
Then $x \in A$ and $x \in B$.	(Definition of intersection)
But then $x \in A$, whence $A \cap B \subseteq A$	(Definition of subset)
Now let $y \in A$.	(Must show $y \in A \Rightarrow y \in A \cup B$)

Then '$y \in A$ or $y \in B$' is true, from which we conclude that $y \in A \cup B$.
Thus $A \subseteq A \cup B$. ∎

The following theorem describes how complements interact with other set operations.

Theorem 3.9. *Let A, B be sets. Then:*

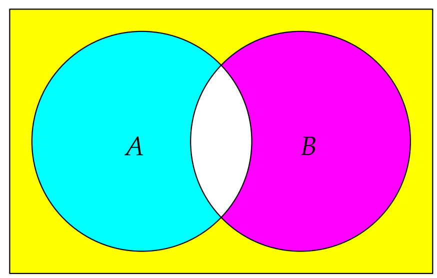
3.3.1 $(A \cap B)^{\mathsf{C}} = A^{\mathsf{C}} \cup B^{\mathsf{C}}$.

3.3.2 $(A \cup B)^{\mathsf{C}} = A^{\mathsf{C}} \cap B^{\mathsf{C}}$.

3.3.3 $(A^{\mathsf{C}})^{\mathsf{C}} = A$.

3.3.4 $A \setminus B = A \cap B^{\mathsf{C}}$.

3.3.5 $A \subseteq B \iff B^{\mathsf{C}} \subseteq A^{\mathsf{C}}$.



$(A \cap B)^{\mathsf{C}} = A^{\mathsf{C}} \cup B^{\mathsf{C}}$

Again: don't memorize these laws! Draw Venn diagrams to help with visualization.

Proof of 1. We start by trying to show that the left hand side is a subset of the right hand side.

$$\begin{aligned}
x \in (A \cap B)^{\mathsf{C}} &\implies x \notin A \cap B \\
&\implies x \text{ is not a member of } \textit{both } A \text{ and } B \\
&\implies x \text{ is not in } \textit{at least one } \text{of } A \text{ and } B \\
&\implies x \notin A \text{ or } x \notin B \\
&\implies x \in A^{\mathsf{C}} \text{ or } x \in B^{\mathsf{C}} \\
&\implies x \in A^{\mathsf{C}} \cup B^{\mathsf{C}}
\end{aligned}$$

With a little thinking, we realize that all of the $\implies$ arrows may be replaced with if and only if arrows $\iff$ without compromising the argument. We've therefore shown that the sets $(A \cap B)^{\mathsf{C}}$ and $A^{\mathsf{C}} \cup B^{\mathsf{C}}$ have the same elements, and are thus equal. ∎

We were lucky with our proof. Showing that both sides are subsets of each other would have been tedious, but we found a quicker proof by carefully laying out one direction. This happens more often than you might expect. Just be careful: you can't always make conditional connectives biconditional.

Parts 1. and 2. of the theorem are known as *De Morgan's laws*, just as the equivalent statements in logic: Theorem 2.10. Indeed, we could rephrase our proof in that language.

Alternative Proof of 1.

$$\begin{aligned}
x \in (A \cap B)^{\mathsf{C}} &\iff \neg[x \in A \cap B] \\
&\iff \neg[x \in A \text{ and } x \in B] \\
&\iff \neg[x \in A] \text{ or } \neg[x \in B] && \text{(De Morgan's first law)} \\
&\iff x \in A^{\mathsf{C}} \text{ or } x \in B^{\mathsf{C}} \\
&\iff x \in A^{\mathsf{C}} \cup B^{\mathsf{C}}
\end{aligned}$$

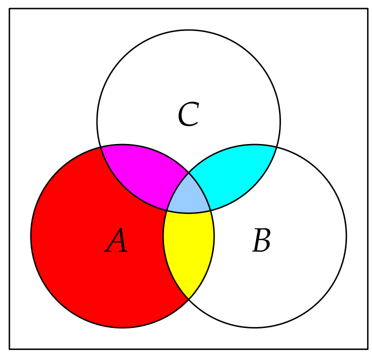
■

Finally, we have two results which describe the interaction of unions and intersections.

Theorem 3.10 (Distributive laws)**.** *For any sets A, B, C:*

3.3.1 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

3.3.2 $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$



We prove only the second result. The method is the standard approach: show that each side is a subset of the other. We do both directions this time, though with a little work and the cost of some clarity, you might be able to slim down the proof. The Venn diagram on the right illustrates the second result: simply add the colored regions.

Proof. ($\subseteq$) Let $x \in A \cup (B \cap C)$. Then $x \in A$ or $x \in B \cap C$. There are two cases:

(a) If $x \in A$, then $x \in A \cup B$ and $x \in A \cup C$ by Theorem 3.8, part 2.

(b) If $x \in B \cap C$, then $x \in B$ and $x \in C$. It follows that $x \in A \cup B$ and $x \in A \cup C$, again by Theorem 3.8.

In both cases $x \in (A \cup B) \cap (A \cup C)$.

($\supseteq$) Let $y \in (A \cup B) \cap (A \cup C)$. Then $y \in A \cup B$ and $y \in A \cup C$. There are again two cases:

(a) If $y \in A$, then we are done, for then $y \in A \cup (B \cap C)$.

(b) If $y \notin A$, then $y \in B$ and $y \in C$. Hence $y \in B \cap C$. In particular $y \in A \cup (B \cap C)$.

In both cases $y \in A \cup (B \cap C)$.

■

Reading Quiz

3.3.1 Let $\mathcal{U} = \mathbb{Z}$, $A = 2\mathbb{Z}$, $B = \{1, 3, 5\}$. Which of the following statements are true?

(a) $B \subseteq A^{\mathsf{C}}$.

(b) A and B are not disjoint.

(c) $A \cup B = \mathbb{Z}$.

(d) $\mathbb{Z} \setminus B$ is finite.

(e) $A^{\mathsf{C}} = 2\mathbb{Z} + 1$

3.3.2 True or False: if A and B are sets, then $B \subseteq A \cup B$.

3.3.3 For sets A and B, the result that $(A \cup B)^{\mathsf{C}} = A^{\mathsf{C}} \cap B^{\mathsf{C}}$ is most similar to which of the following laws of logic?

(a) Law of double negation.

(b) Law of absorption.

(c) De Morgan's laws.

(d) Law of associativity.

3.3.4 For sets A and B, which of the following are true?

(a) $A \cap B \subseteq A \setminus B$.

(b) $B = (A \cap B) \cup (B \setminus A)$.

(c) $A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A)$.

(d) $A \setminus B = B \setminus A$.

Practice Problems

3.3.1 Let $a, b, c, d \in \mathbb{R}$. Show

$$(a, b) \cap (c, d) = (\max\{a, c\}, \min\{b, d\})$$

where we take the convention that $(\alpha, \beta) = \emptyset$ if $\beta < \alpha$.

Video Solution

3.3.2 Let $\mathcal{U}$ be a universal set and A and B sets. Prove that $(A \setminus B)^{\mathsf{C}} = A^{\mathsf{C}} \cup B$.

Video Solution

3.3.3 Let A be a set. Prove that if $A \cup B \subseteq B$ for every set B, then $A = \emptyset$.

Video Solution

Exercises

3.3.1 Describe each of the following sets in as simple a manner as you can: e.g.,

$$\{x \in \mathbb{R} : (x^2 > 4 \text{ and } x^3 < 27) \text{ or } x^2 = 15\} = (-\infty, -2) \cup (2, 3) \cup \{\sqrt{15}\}.$$

(a) $\{x \in \mathbb{R} : x^2 \neq x\}$

(b) $\{x \in \mathbb{R} : x^3 - 2x^2 - 3x \leq 0 \text{ or } x^2 = 4\}$

(c) $\{x^2 \in \mathbb{R} : x \neq 1\}$

(d) $\{z \in \mathbb{Z} : z^2 \text{ is even and } z^3 \text{ is odd}\}$

3.3.2 (a) Let A and B be sets. Use logical connectives to rewrite the following propositions

i. $x \in A \cap B$

ii. $x \in A \cup B$

iii. $x \in A \setminus B$‘

iv. $x \in (A \cup B) \setminus (A \cap B)$

in terms of the statements P: ‘$x \in A$′ and Q: ‘$x \in B$′.

(b) For $A = \{1, 3, 5, 7, 9, 11\}$ and $B = \{1, 4, 7, 10, 13\}$, compute the following sets:

i. $A \cap B$

ii. $A \cup B$

iii. $A \setminus B$

iv. $(A \cup B) \setminus (A \cap B)$.

3.3.3 Let $A \subseteq \mathbb{R}$, and let $x \in \mathbb{R}$. We say that the point x is *far away* from the set A if and only if:

$$\exists d > 0\colon \text{ No element of } A \text{ belongs to the set } [x - d, x].$$

Equivalently, $A \cap [x - d, x] = \emptyset$. If this does not happen, we say that x is *close to* A.

(a) Draw a picture of a set A and an element x such that is *far away* from A.

(b) Draw a picture of a set A and an element x such that x is *close* to A.

(c) Compute the definition of "x is close to A". [So negate "x is far away from A".]

(d) Let $A = \{1, 2, 3\}$. Show that $x = 4$ is *far away* from A, by using definitions.

(e) Let $A = \{1, 2, 3\}$. Show that $x = 1$ is *close* to A, by using definitions.

(f) Show that if $x \in A$, then x is *close* to A.

(g) Let A be the open interval (a, b). Is the end-point a *far away* from A? What about the end-point b?

3.3.4 Consider Theorems 3.8 and 3.10. In all seven results, replace the symbols in the first row of the following table with those in the second. Which of the results seem familar? Which are false?

$\emptyset$	A, B, C sets	$\cup$	$\cap$	$\subseteq$
0	$A, B, C \in \mathbb{N}_0$	$+$	$\cdot$	$\leq$

3.3.5 Prove that $B \setminus A = B \iff A \cap B = \emptyset$.

3.3.6 Practice your proof skills by giving formal proofs of the following results from Theorems 3.8 and 3.9. With practice you should be able to prove *all* of parts of these theorems (and of Theorem 3.10) these *without* looking at the arguments in the notes!

(a) $\emptyset \cap A = \emptyset$.

(b) $A \cap (B \cap C) = (A \cap B) \cap C$.

(c) $(A^{\mathsf{C}})^{\mathsf{C}} = A$.

(d) $A \subseteq B \iff B^{\mathsf{C}} \subseteq A^{\mathsf{C}}$.

3.3.7 Write out a formal proof of the set identity

$$A = (A \setminus B) \cup (A \cap B)$$

by showing that each side is a subset of the other. Now repeat your argument using only results from set algebra (Theorems 3.9 and 3.10).

3.3.8 Let $\mathcal{U}$ be a universal set and A and B be sets. Prove the following:

(a) $A \cap B = A \setminus (A \setminus B)$

(b) $A \cup (A \cap B) = A$

(c) $A \setminus B = B^{\mathsf{C}} \setminus A^{\mathsf{C}}$

3.3.9 (a) Let A be a set. Prove A is empty if and only if there exists a set B such that $A \subseteq B \setminus A$.

(b) Let A, B and C be sets. Prove or disprove: if $A \cap C \subseteq A \cap B$, then $C \subseteq A \cup B$.

3.3.10 Let A and B be sets.

(a) Let A and B be sets. Show that $A \cup B$ is the smallest set containing both A and B, in the sense that if $A \subseteq C$ and $B \subseteq C$, then $A \cup B \subseteq C$.

(b) Show that $A \cap B$ is the largest set contained in both A and B, in the sense that if $C \subseteq A$ and $C \subseteq B$, then $C \subseteq A \cap B$.

3.3.11 Let $\mathcal{U}$ be a universal set and A and B be sets. Define the *symmetric difference* of A and B to be the set

$$A \Delta B = (A \cup B) \setminus (A \cap B).$$

(a) Draw a Venn diagram of A and B and shade in the part of the diagram that comprises $A \Delta B$.

(b) Prove $A \Delta B = (A \setminus B) \cup (B \setminus A)$.

(c) Say we wish to write $A \Delta B$ in set-builder notation as

$$A \Delta B = \{x \in \mathcal{U} : (x \in A) \oplus (x \in B)\}.$$

Construct the truth table for $\oplus$.

3.3.12 Let A and B be finite sets. Find necessary and sufficient conditions on A and B such that $|A \cap B| = |A|$. In other words, fill in the blank in the following statement

$$|A \cap B| = |A| \text{ if and only if } \underline{\hspace{4cm}}.$$

Prove this statement is true.

3.3.13 Let A and B be finite sets.

(a) Find and example of A and B for which $|A \cup B| \neq |A| + |B|$.

(b) Find an example of A and B for which $|A \cup B| = |A| + |B|$. What do you notice about $A \cap B$ in this example?

(c) The *inclusion-exclusion principle* says that, in general, we have

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Looking at a Venn diagram, why does this make sense?

(d) Consider a calculus class with 100 students. Suppose that 85 of the students are either math majors or engineering majors, 78 are only majoring in engineering, and 3 are double majoring in both math and engineering. How many math majors are in the class?

(e) Formulate a similar expression for the cardinality of the union of three sets $A \cup B \cup C$ (you do not have to prove your assertion is correct).

(f) Using your answer to the previous part, find the number of integers between 1 and 100 which are not divisible by 2, 5, or 7.

3.4 Introduction to Functions

You have been using functions for a long time. A formal definition in terms of relations will be given in Section 7.2. For the present, we will just use the following.

Definition 3.11. Let A and B be sets. A **function from** A **to** B is a rule f that assigns one (and only one) element of B to each element of A.

The **domain** of f, written $\mathrm{dom}(f)$, is the set A. The **codomain** of f is the set B.

If f is a function from A to B we write $f : A \to B$. If $a \in A$, we write $b = f(a)$ for the the element of B assigned to a by the function f. We can also write $f : a \mapsto b$, which is read "f maps a to b".

You can think of the domain of f as the set of all inputs for the function and the codomain is the set of all *potential* output values the function may take (not all of the values in the codomain are necessarily achieved).

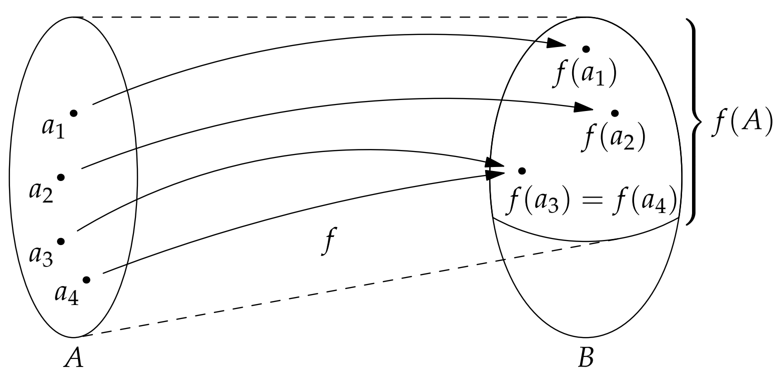
Definition 3.12. If $f : A \to B$ is a function and U is a subset of A. Then the **image of** U is the following subset of B,

$$f(U) = \{f(u) \in B : u \in U\}.$$

The image of A is called the **range** or **image of** f,

$$f(A) = \mathrm{range}(f) = \mathrm{Im}(f) = \{f(a) \in B : a \in A\}.$$

So the codomain of f is the set of all potential outputs, the range or image of f is the subset of the codomain consisting of all actual outputs of f.

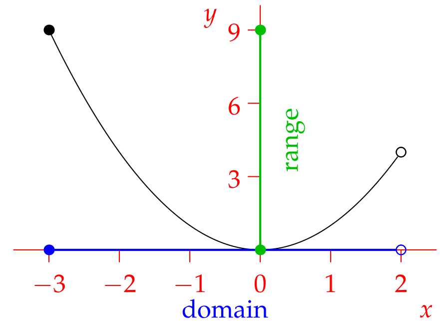


For simple real-valued functions, the domain and range are easily seen in a graph. For instance if $f : [-3, 2) \to \mathbb{R}$ is the square function

$$f : x \mapsto x^2,$$

then we have $\operatorname{dom}(f) = [-3, 2)$ and $\operatorname{range}(f) = [0, 9]$, as seen in the picture. We could also calculate other images, for example,

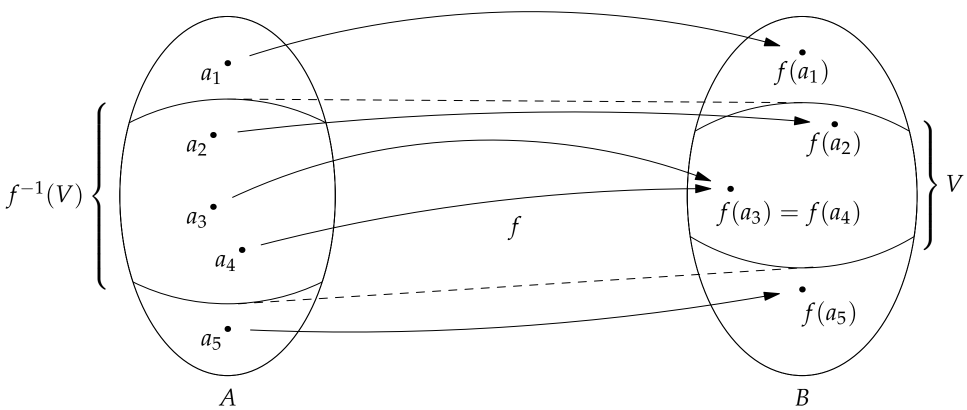
$$f\big([-1, 2)\big) = [0, 4).$$



There is a dual construction to the image, where we start with a subset of the codomain, and look at the set of inputs which get mapped into this set.

Definition 3.13. Let $f : A \to B$ be a function and V a subset of B. Then the **preimage of** V (also called the **inverse image of** V is the following subset of A,

$$f^{-1}(V) = \{a \in A : f(a) \in V\}.$$



For most functions we will not be able to sketch a graph. Here are several examples where a graph is either unhelpful, or simply impossible to draw!

Examples. 3.4.1 Define $f : \mathbb{Z} \to \{0, 1, 2\}$ by $f : n \mapsto r$, where r is the remainder of n^2 upon division by 3. Here $\operatorname{dom}(f) = \mathbb{Z}$ and the codomain is $\{0, 1, 2\}$, but what is the range? Trying a few examples, we see the following:

n	0	1	2	3	4	5	6	7	8	9	10
$f(n)$	0	1	1	0	1	1	0	1	1	0	1

It looks like the range is simply $\{0, 1\}$. In fact, we have already proved this fact in Theorem 2.27. In other words, $f(\mathbb{Z}) = \operatorname{range}(f) = \Im(f) = \{0, 1\}$. Notice that the range is a *proper* subset of the codomain: nothing gets mapped to 2!

3.4.2 Let $A = \{0, 1, 2, \ldots, 9\}$ and define $f : A \to A$ by in the following way:

n	0	1	2	3	4	5	6	7	8	9
$f(n)$	0	3	6	9	2	5	8	1	4	7

It should be obvious that $f(A) = \text{range}(f) = A$. Let $U = \{0, 1, 2, 3, 4\}$. Then

$$f(U) = \{f(u) : u \in U\} = \{f(u) : u \in \{0, 1, 2, 3, 4\}\} = \{f(0), f(1), f(2), f(3), f(4)\} = \{0, 3, 6, 9, 2\}.$$

Let $V = \{0, 1, 2, 3, 4\}$. Then

$$f^{-1}(V) = \{a \in A : f(a) \in V\} = \{a \in A : f(a) \in \{0, 1, 2, 3, 4\}\} = \{0, 1, 4, 7, 8\}.$$

3.4.3 With the same notation as the previous example, let $g : A \to A$ be given by the following table:

n	0	1	2	3	4	5	6	7	8	9
$g(n)$	0	4	8	2	6	0	4	8	2	6

with $\text{range}(g) = \{0, 2, 4, 6, 8\}$.

3.4.4 Let $A = \{1, 2, 3, 4, 5\}$ and let $B = \{\text{two-element subsets of } A\}$. We define

$$f : A \to B : a \mapsto \begin{cases} \{a, a+1\} & \text{if } a \neq 5, \\ \{5, 1\} & \text{if } a = 5. \end{cases}$$

This is tricky to read, since B is a set of sets. You should be able to convince yourself that

$$\text{range}(f) = \big\{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{5, 1\}\big\}$$

and, for example, that

$$f\left(\{1, 4\}\right) = \{f(1), f(4)\} = \big\{\{1, 2\}, \{4, 5\}\big\}.$$

If $V = \big\{\{1, 2\}, \{5, 1\}, \{2, 2\}, \{3, 3\}\big\}$. Then $f^{-1}(V) = \{1, 5\}$.

Injections, surjections and bijections

Definition 3.14. A function $f : A \to B$ is **one-to-one**, **injective**, or an **injection** if it never takes the same value twice. Equivalently,[a]

$$\forall a_1, a_2 \in A,\ f(a_1) = f(a_2) \implies a_1 = a_2.$$

$f : A \to B$ is **onto**, **surjective**, or a **surjection** if it takes every value in the codomain: i.e., $B = \text{range}(f)$. Equivalently,[b]

$$\forall b \in B,\ \exists a \in A \text{ such that } f(a) = b.$$

$f : A \to B$ is **invertible**, **bijective**, or a **bijection** if it is both injective and surjective.

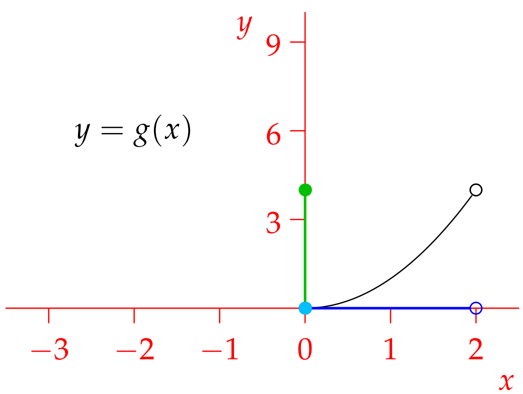
[a] This is the contrapositive: if f never takes the same value twice, then $\forall a_1, a_2 \in A$ we have $a_1 \neq a_2 \implies f(a_1) \neq f(a_2)$.
[b] This is the statement $B \subseteq \text{range}(f)$. The opposite inclusion $\text{range}(f) \subseteq B$ is true for *any* function.

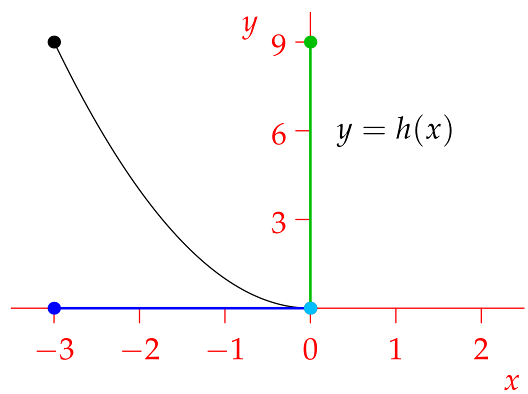
Since the definitions of injective and surjective are both 'for all' statements, to show that a function is *not injective* or *not surjective* you will need *counterexamples.* For instance, consider the quadratic function $f : [-3, 2) \to \mathbb{R} : x \mapsto x^2$ seen above. It is straightforward to see that f is neither injective nor surjective. Indeed we have the following counterexamples:

- $f(-1) = f(1)$. If f were injective, the values at 1 and -1 would have to be different.
- $81 \in \mathbb{R}$, yet there is no $x \in [-3, 2)$ such that $f(x) = 81$. Thus f is not surjective.

With a small change to either the domain or codomain, we can easily create an injective or a surjective function. For instance we can shrink the domain to obtain two injective functions:

$$g : [0, 2) \to \mathbb{R} : x \mapsto x^2 \quad \text{and} \quad h : [-3, 0] \to \mathbb{R} : x \mapsto x^2$$





To see this, note that

$$g(x_1) = g(x_2) \implies x_1^2 = x_2^2 \implies x_1 = \pm x_2 \implies x_1 = x_2$$

since both must be non-negative. The argument for h is similar.
By shrinking the codomain to equal the range we immediately create a surjective function:

$$j : [-3, 2) \to [0, 9] : x \mapsto x^2$$

Now consider the examples on page 96. The details are provided for example 1. For the others, make sure you understand why the answer is correct.

Examples. 3.4.1 $f : \mathbb{Z} \to \{0, 1, 2\} : n \mapsto n^2 \pmod 3$ is neither injective nor surjective.

- If f were injective, then we could not have $f(1) = f(2)$.
- 2 is in the codomain $\{0, 1, 2\}$ of f, yet $2 \notin \text{range}(f)$, so f is not surjective.

3.4.2 This is a bijection. Indeed f is a *permutation*, a bijection from a set onto itself. To see injectivity, note that in the table

n	0	1	2	3	4	5	6	7	8	9
$f(n)$	0	3	6	9	2	5	8	1	4	7

none of the values in the second row appears more than once. For surjectivity, observe that every element in the codomain $\{0, 1, 2, \ldots, 9\}$ appears *at least* once in the second row. Being bijective means that each element of the codomain appears *exactly once*.

3.4.3 Neither injective, nor surjective.

3.4.4 Injective, but not surjective.

Here is a more complicated example.

Example. Prove that $f : \mathbb{R} \setminus \{1\} \to \mathbb{R} \setminus \{2\}$ defined by $f(x) = 2 + \frac{1}{1-x}$ is bijective.

(Injectivity) Suppose that x_1 and x_2 are in $\mathbb{R} \setminus \{1\}$, and $f(x_1) = f(x_2)$. Then

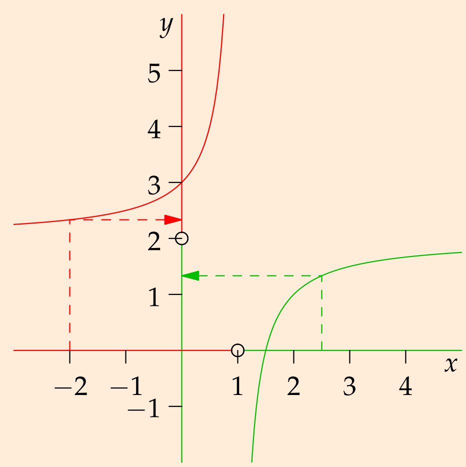
$$2 + \frac{1}{1-x_1} = 2 + \frac{1}{1-x_2}.$$

A little elementary algebra shows that $x_1 = x_2$, whence f is injective.

(Surjectivity) Let $y \in \mathbb{R} \setminus \{2\}$ and define $x = 1 - \frac{1}{y-2}$. This makes sense since $y \neq 2$. Then

$$f(x) = 2 + \frac{1}{1 - (1 - \frac{1}{y-2})} = y$$

whence f is surjective.



The graphic is colored so that you can see how the different parts of the range and domain correspond. The argument for surjectivity is sneaky: how did we know to choose $x = 1 - \frac{1}{y-2}$? The answer is scratch work: just solve $y = 2 + \frac{1}{1-x}$ for x. Essentially we've shown that f has the inverse function $f^{-1}(x) = 1 - \frac{1}{x-2}$.

Aside. Inverse Functions

The word *invertible* is a synonym for bijective because bijective functions really have inverses! Indeed, suppose that $f : A \to B$ is bijective. Since f is surjective, we know that $B = \text{range}(f)$ and so every element of B has the form $f(a)$ for some $a \in A$. Moreover, since f is injective, the a in question is unique. The upshot is that, when f is bijective, we can construct a new *function*

$$f^{-1} : B \to A : f(a) \mapsto a.$$

This may appear difficult at the moment but we will return to it in Chapter 7.

Instead, recall that in Calculus you saw that any injective function has an inverse. How does this fit with our definition? Consider, for example, $f : [0,2] \to \mathbb{R} : x \mapsto x^4$. This is injective but not surjective. To fix this, simply define a new function with the same formula but with codomain equal to the range of f. We obtain the bijective function

$$g : [0,2] \to [0,16] : x \mapsto x^4,$$

with inverse

$$g^{-1} : [0,16] \to [0,2] : x \mapsto \sqrt[4]{x}.$$

In Calculus we didn't nitpick like this and would simply go straight to $f^{-1}(x) = \sqrt[4]{x}$.

In general, if $f : A \to B$ is any injective function, then $g : A \to f(A) : x \mapsto f(x)$ is automatically bijective, since we are forcing the codomain of g to match its range.

Functions and Cardinality

Injective and surjective functions are intimately tied to the notion of cardinality. Indeed, in Chapter 8, we will use such functions to give a *definition* of cardinality for infinite sets. For the present we stick to finite sets.

Theorem 3.15. *Let A and B be finite sets. The following are equivalent:*

3.4.1 $|A| \le |B|$.

3.4.2 $\exists f : A \to B$ *injective.*

3.4.3 $\exists g : B \to A$ *surjective.*

Read the theorem carefully. It is simply saying that, of the three statements, if *any* one is true then *all* are true. Similarly, if one is false then so are the others. It might appear that we require six arguments! Instead we illustrate an important technique: when showing that multiple statements are equivalent, it is enough to prove in a circle. For instance, if we prove the three implications indicated in the picture, then 1 $\Rightarrow$ 3 will be true because *both* 1 $\Rightarrow$ 2 and 2 $\Rightarrow$ 3 are true.

$$\begin{tikzcd} & 1 \arrow[dr, Rightarrow] & \\ 3 \arrow[ur, Rightarrow] & & 2 \arrow[ll, Rightarrow] \end{tikzcd}$$

More generally, to show that n statements are equivalent, only n arguments are required.

The proof may appear very abstract, but it is motivated by two straightforward pictures. Don't be afraid to use pictures to illustrate your proofs if it's going to make them easier to follow! If $|A| = m$ and $|B| = n$, then the two functions can be displayed pictorially. Refer back to these pictures as you read through the proof.

$$\begin{array}{c} A = \{a_1, a_2, a_3, \cdots, a_m\} \\ \downarrow \;\; \downarrow \;\; \downarrow \qquad\quad \downarrow \qquad\qquad\quad \\ B = \{b_1, b_2, b_3, \cdots, b_m, \cdots, b_n\} \end{array} \qquad\qquad \begin{array}{c} A = \{a_1, a_2, a_3, \cdots, a_m\} \\ \uparrow \;\; \uparrow \;\; \uparrow \qquad\quad \uparrow \qquad\qquad\qquad\qquad \\ B = \{b_1, b_2, b_3, \cdots, b_m, \overbrace{b_{m+1}, \cdots, b_n}\} \end{array}$$

The function f $\qquad\qquad$ The function g

Proof. The proof relies crucially on the fact that A, B are finite. Suppose that $|A| = m$ and $|B| = n$ throughout and list the elements of A and B as,

$$A = \{a_1, a_2, \ldots, a_m\}, \qquad\qquad B = \{b_1, b_2, \ldots, b_n\}.$$

$(1 \Rightarrow 2)$ Assume that $m \le n$. Define $f : A \to B$ by $f(a_k) = b_k$. This is injective since the elements $b_1, \ldots, b_m$ are distinct.

$(2 \Rightarrow 3)$ Suppose that $f : A \to B$ is injective. Without loss of generality we may assume that the elements of A and B are labeled such that $f(a_k) = b_k$. Now define $g : B \to A$ by

$$g(b_k) = \begin{cases} a_k & \text{if } k \le m, \\ a_1 & \text{if } k > m. \end{cases}$$

Then g is surjective since every element a_k is in the image of g.

$(3 \Rightarrow 1)$ Finally suppose that $g : B \to A$ is surjective. Without loss of generality we may assume that $a_k = g(b_k)$ for $1 \le k \le m$. Thus $n \ge m$. ∎

It is worth noting in the proof of $(3 \Rightarrow 1)$ that the elements $b_{m+1}, \ldots, b_n$ may be mapped *anywhere*, not just to a_1 as suggested in the picture above.

If you read the proof carefully, it should be clear that when $m = n$, the function f is actually a *bijection* (with inverse $f^{-1} = g$).

Corollary 3.16. *If A, B are finite sets, then $|A| = |B| \iff \exists f : A \to B$ bijective.*

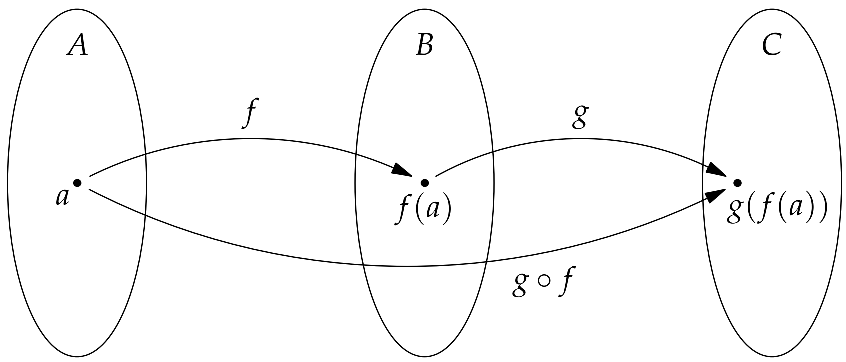
Proof. Suppose that $m = n$. The argument $1 \Rightarrow 2$ creates an injective function $f : A \to B$. However every element $b_k \in B$ is in the image of f, so this function is also surjective. Hence f is a bijection. Conversely, if $f : A \to B$ is a bijection, then it is injective, whence $m \le n$. It is also surjective, from which $n \le m$. Therefore $m = n$. ∎

Composition of functions

Finally, we consider composing function and, more particularly, how injectivity and surjectivity interact with composition.

Definition 3.17. Suppose that $f : A \to B$ and $g : B \to C$ are functions. The **composition** $g \circ f : A \to C$ is the function defined by $(g \circ f)(a) = g(f(a))$.

Note the order: to compute $(g \circ f)(x)$, you apply f first, then g.



Example. If $f(x) = x^2$ and $g(x) = \frac{1}{x-1}$, then

$$(g \circ f)(x) = \frac{1}{x^2 - 1}, \quad \text{and} \quad (f \circ g)(x) = \frac{1}{(x-1)^2}.$$

You should be extra careful of ranges and domains when composing functions. The domain and range are not always explicitly mentioned, and at times some restriction of the domain is implied. In this example, you might assume that $\operatorname{dom}(f) = \mathbb{R}$ and $\operatorname{dom}(g) = \mathbb{R} \setminus \{1\}$. This is perfectly good if we are considering f and g separately. However, it should be clear from the formulæ that the implied domains of the compositions are,

$$\operatorname{dom}(g \circ f) = \mathbb{R} \setminus \{\pm 1\}, \quad \text{and} \quad \operatorname{dom}(f \circ g) = \mathbb{R} \setminus \{1\}.$$

Our first two results on composing injective and surjective functions is easy to remember.

Theorem 3.18. *Let $f : A \to B$ and $g : B \to C$ be functions. Then:*

3.4.1 If f and g are injective, then $g \circ f$ is injective.

3.4.2 If f and g are surjective, then $g \circ f$ is surjective.

It follows that the composition of bijective functions is also bijective.

Proof. 3.4.1 Suppose that f and g are injective and let $a_1, a_2 \in A$ satisfy $(g \circ f)(a_1) = (g \circ f)(a_2)$. We are required to show that $a_1 = a_2$. However,

$$\begin{aligned} (g \circ f)(a_1) = (g \circ f)(a_2) &\implies g\big(f(a_1)\big) = g\big(f(a_2)\big) \\ &\implies f(a_1) = f(a_2) && \text{(since } g \text{ is injective)} \\ &\implies a_1 = a_2 && \text{(since } f \text{ is injective)} \end{aligned}$$

■

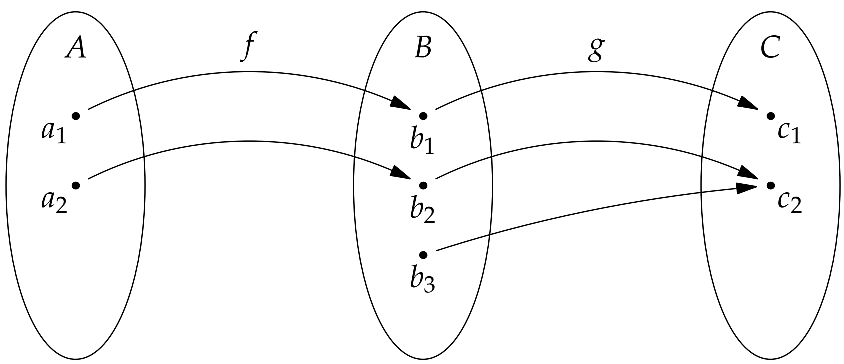
Part 2 is in the Exercises. It is interesting to observe that the converse of this theorem is *false.* Assuming that a composition is injective or surjective only forces *one* of the original functions to be so.

Theorem 3.19. *Suppose that $f : A \to B$ and $g : B \to C$ are functions.*

3.4.1 If $g \circ f$ is injective, then f is injective.

3.4.2 If $g \circ f$ is surjective, then g is surjective.

Before showing the proof, consider the following representation of two functions f and g which simultaneously illustrate both parts of the theorem. It should be clear that $g \circ f$ is *bijective,* f is *only injective,* and g is *only surjective.*



Here is a formulaic example of the same thing. Make sure you're comfortable with the definitions and draw pictures or graphs to help make sense of what's going on.

$f : [0,2] \to [-4,4] : x \mapsto x^2$ (injective only)

$g : [-4,4] \to [0,16] : x \mapsto x^2$ (surjective only)

$g \circ f : [0,2] \to [0,16] : x \mapsto x^4$ (bijective!)

This time we leave part 1 of the proof for the Exercises.

Proof. 3.4.2 Let $c \in C$ and assume that $g \circ f$ is surjective. We wish to prove that $\exists b \in B$ such that $g(b) = c$.
Since $g \circ f$ is surjective, $\exists a \in A$ such that $(g \circ f)(a) = c$. But this says that

$$g(f(a)) = c.$$

Hence $b = f(a)$ is an element of B for which $g(b) = c$. Thus g is surjective. ∎

Reading Quiz

3.4.1 The range of a function $f : A \to B$ is (select all that apply)

(a) a subset of the domain.
(b) a subset of the codomain.
(c) always equal to the codomain.
(d) also called the image of the function.
(e) equal to $f(A)$.

3.4.2 Suppose $f : A \to B$ and $g : B \to C$ are functions. If $g \circ f$ is bijective, which of the following *must* be true?

(a) f is injective.
(b) g is injective.
(c) f is surjective.
(d) g is surjective.

3.4.3 True or False: We can always make a function surjective by making its domain smaller.

3.4.4 True or False: If $A \subseteq B$, there is an injective function $f : A \to B$.

Practice Problems

3.4.1 (a) Explain why the map $g : \{\text{all lines in the planes}\} \to \mathbb{R}$ which sends a line ℓ to the slope of ℓ is not a function.
Video Solution

(b) Let L be the set of all non-vertical lines in the plane. The map $f : L \to \mathbb{R}$ defined by $\ell \mapsto$ slope of ℓ is a well defined function. Find $f(Z)$ where Z is the subset of L consisting of the lines that intersect the line $y = 2x + 5$ at exactly one point.
Video Solution

(c) Now let $U = \{-2\}$. Describe the inverse image $f^{-1}(U)$ of U under the function f defined in part (b).
Video Solution

(d) Is the function f bijective?
Video Solution
(e) Find a subset B of L so that the function $f : B \to \mathbb{R}$ is a bijection.
Video Solution

3.4.2 Suppose $f : A \to B$ and $g : B \to C$ are functions. For each of the following, either find an example or explain why no such example exists.

(a) f surjective and g not surjective so that the composition $g \circ f$ is surjective.
(b) f not surjective and g surjective so that the composition $g \circ f$ is surjective.
(c) f surjective and g surjective so that the composition $g \circ f$ is not surjective.
(d) f injective and g not injective so that the composition $g \circ f$ is injective.
(e) f not injective and g injective so that the composition $g \circ f$ is injective.
(f) f injective and g injective so that the composition $g \circ f$ is not injective.

Video Solution (Parts (a)-(c))

3.4.3 Suppose $f : A \to B$ is a function. Prove or disprove each of the following statements:

(a) Let X and Y be subsets of A. If $X \cap Y = \emptyset$ then $f(X) \cap f(Y) = \emptyset$.
(b) Let W and Z be subsets of B. If $W \cap Z = \emptyset$ then $f^{-1}(W) \cap f^{-1}(Z) = \emptyset$.

Video Solution

Exercises

3.4.1 For each of the following functions $f : A \to B$ determine whether f is injective, surjective or bijective. Prove your assertions.

(a) $f : [0,3] \to \mathbb{R}$ where $f(x) = 2x$.
(b) $f : [3,12) \to [0,3)$ where $f(x) = \sqrt{x-3}$.
(c) $f : (-4,1] \to (-5,-3]$ where $f(x) = -\sqrt{x^2+9}$.

3.4.2 Suppose that $f : [-3,\infty) \to [-8,\infty)$ and $g : \mathbb{R} \to \mathbb{R}$ are defined by

$$f(x) = x^2 + 6x + 1, \qquad g(x) = 2x + 3.$$

Compute $g \circ f$ and show that $g \circ f$ is injective.

3.4.3 Find:

(a) A set A so that the function $f : A \to \mathbb{R} : x \mapsto \cos x$ is injective.
(b) A set B so that the function $f : \mathbb{R} \to B : x \mapsto \cos x$ is surjective.

3.4.4 (*If you did Exercise 2.2.7 you should find this easy*) Let X be a subset of $\mathbb{R}$. A function $f : X \to \mathbb{R}$ is *strictly increasing* if

$$\forall\, a,\, b \in X, \quad a < b \Longrightarrow f(a) < f(b).$$

For example, the function $f\colon [0,\, \infty) \to \mathbb{R},\, x \mapsto x^2$ is increasing because

$$\forall a,\, b \in [0,\, \infty), \quad a < b \Longrightarrow f(a) = a^2 < b^2 = f(b).$$

(a) Give another example of a function that is increasing. Draw its graph, and prove that the function is increasing.

(b) By negating the above definition, state what it means for a function *not to be strictly increasing.*

(c) Give an example of a function that is *not* strictly increasing. Draw its graph, and prove that the function is not strictly increasing.

(d) Let $f, g : \mathbb{R} \to \mathbb{R}$ be strictly increasing. Prove or disprove: The function $h = f + g$ is strictly increasing. Note that the formula for h is $h(x) = f(x) + g(x)$.

3.4.5 Let L be the set of all non-vertical lines in the plane. Let $f : L \to \mathbb{R}$ be the function which sends each line to its y-intercept. Is f injective? Is f surjective? Justify your answers.

3.4.6 You may assume that $g : [2, \infty) \to \mathbb{R} : x \mapsto \sqrt{x^3 - 8}$ is an injective function. Find a function $f : \mathbb{R} \to \mathbb{R}$ which is *not injective*, but for which the composition $f \circ g : [2, \infty) \to \mathbb{R}$ *is injective.* Justify your answer.

3.4.7 A function $f : \mathbb{R} \to \mathbb{R}$ is *even* if

$$\forall x \in \mathbb{R},\ f(-x) = f(x).$$

For example, the function $f : \mathbb{R} \to \mathbb{R},\ x \mapsto x^2$ is even because

$$\forall x \in \mathbb{R},\ f(-x) = (-x)^2 = x^2 = f(x).$$

Note that f is even if and only if the graph of f is symmetric with respect to the y axis.

(a) Give an example of a function that is even. Draw its graph, and prove that the function is even.

(b) Define what it means for a function *not to be even*, by negating the definition above.

(c) Give an example of a function that is *not* even. Draw its graph, and prove that the function is not even.

(d) Prove or disprove: for every $f, g \colon \mathbb{R} \to \mathbb{R}$ even, the composition $h = f \circ g$ is even. Here h is the function mapping x to $f(g(x))$.

3.4.8 Define $f : (-\infty, 0] \to \mathbb{R}$ and $g : [0, \infty) \to \mathbb{R}$ by

$$f(x) = x^2, \qquad g(x) = \begin{cases} \frac{x}{1-x} & x < 1, \\ 1 - x & x \geq 1. \end{cases}$$

Does $g \circ f$ map $(-\infty, 0]$ onto $\mathbb{R}$? Justify your answer.

3.4.9 Express, using quantifiers, what it means for a function to be

(a) Not injective.

(b) Not surjective.

3.4.10 Let $f : \mathbb{R} \to \mathbb{R}^+$ be the function defined by $f(x) = e^x$. Explain why the following "proof" that f is surjective is incorrect. Then, give a correct proof.

Proof. Let $e^x \in \mathbb{R}^+$ be arbitrary. Then $f(x) = e^x$. So f is surjective. ∎

3.4.11 Prove that the composition of two surjective functions is surjective.

3.4.12 Suppose that $g \circ f$ is injective. Prove that f is injective.

3.4.13 In the proof of Theorem 3.15 we twice invoked *without loss of generality*. In both cases explain why the phrase applies.

3.4.14 Let $f : A \to B$ be a function. Let $X_1, X_2 \subseteq A$. Prove or disprove the following:

(a) $X_1 \subseteq X_2$ implies $f(X_1) \subseteq f(X_2)$.
(b) $f(X_1 \cup X_2) = f(X_1) \cup f(X_2)$.
(c) $f(X_1 \cap X_2) \subseteq f(X_1) \cap f(X_2)$.
(d) $f(X_1) \cap f(X_2) \subseteq f(X_1 \cap X_2)$.

3.4.15 Let $f : A \to B$ be a function. Suppose that $f(X_1 \cap X_2) = f(X_1) \cap f(X_2)$ for **all** $X_1, X_2 \subseteq A$. Show f is injective.

3.4.16 (a) Let $A = \{a, b, c\}$ and $B = \{1, 2, 3, 4\}$ and $f : A \to B$ be the function given by $f(a) = f(c) = 1$ and $f(b) = 3$. Compute $f^{-1}(\{1\})$, $f^{-1}(\{3\})$, $f^{-1}(\{1, 3\})$, and $f^{-1}(\{2, 4\})$.
(b) Let $g : [-1, \infty) \to \mathbb{R}$ be $g(x) = x^2 + 2x + 1$. Compute $g^{-1}((0, 2))$.
(c) Let $h : \mathbb{R} \to \mathbb{R}$ be $h(x) = \sin x$. Find $h^{-1}(\{-1, 1\})$.

3.4.17 Let $f : A \to B$ be a function and $Y_1, Y_2 \subseteq B$.

(a) Prove $f^{-1}(Y_1 \cup Y_2) = f^{-1}(Y_1) \cup f^{-1}(Y_2)$.
(b) Prove $f^{-1}(Y_1 \cap Y_2) = f^{-1}(Y_1) \cap f^{-1}(Y_2)$.

3.4.18 Let $f : A \to B$ be a function and let $X \subseteq A$. Fill in the details in the following to give a proof of the following two facts:

(a) $X \subseteq f^{-1}(f(X))$.
(b) If f is injective, $X = f^{-1}(f(X))$.

3.4.19 Let $f : A \to B$ be a function and let $Y \subseteq B$. Prove the following two facts:

(a) $f(f^{-1}(Y)) \subseteq Y$.
(b) If f is surjective, $f(f^{-1}(Y)) = Y$.

3.4.20 Let A, B, C, and D be sets and $f : A \to B$, $g : B \to C$, and $h : C \to D$ be functions. Show that for all $a \in A$, we have

$$(f \circ (g \circ h))(a) = ((f \circ g) \circ h)(a).$$

3.4.21 (Uses calculus) This exercise will give an example of how to use calculus to prove some properties of (certain) functions. Let $f : (-\pi/2, \pi/2) \to \mathbb{R}$ be defined by $f(x) = \tan x$. Recall that f is differentiable, and hence continuous, on its domain.

(a) Compute $\lim_{x \to \frac{\pi}{2}^-} f(x)$ and $\lim_{x \to \frac{-\pi}{2}^+} f(x)$.

(b) Recall the Intermediate Value Theorem: if $g : [a, b] \to \mathbb{R}$ is a continuous function, then for any y between $g(a)$ and $g(b)$, there is $x \in [a, b]$ such that $g(x) = y$. Use the Intermediate Value Theorem and the results of part 1 to prove f is surjective.

(c) Why is a strictly increasing function (see Exercise 4.4.4) injective?

(d) Compute $\frac{d}{dx} f(x)$ and use this to show f is strictly increasing, and therefore injective by part 3.

3.4.22 Show there is a bijection between $\mathbb{Z}$ and $2\mathbb{Z}$.

3.4.23 Let S be the set of all circles in the plane which are centered at the origin. Find a bijection between S and $\mathbb{R}^+$.

3.4.24 Let A and B be **finite** sets. If $A \subsetneq B$, is it possible for there to be a bijection between A and B?

4 Divisibility and the Euclidean Algorithm

In this section we introduce the notion of *congruence*: a generalization of the idea of separating all integers into 'even' and 'odd.' At its most basic it involves going back to elementary school when you first learned division and would write something similar to

$33 \div 5 = 6\,\mathrm{r}\,3$ and read '6 remainder 3.'

The study of congruence is of fundamental importance to Number Theory, and provides some of the most straightforward examples of Groups and Rings. We will cover the basics in this section—enough to compute with—then return later for more formal observations.

4.1 Remainders and Congruence

Definition 4.1. Let m and n be integers, with $n \neq 0$. We say that n **divides** m and write $n \mid m$ if m is divisible by n: that is if there exists some integer k such that $m = kn$. Equivalently, we say that n is a **divisor** of m, or that m is a **multiple** of n.

Examples. Since $20 = 4 \cdot 5$ we may write $4 \mid 20$. Similarly $17 \mid 51$. We may also use the symbol $\nmid$ for 'does not divide.' Thus $12 \nmid 8$ and $7 \nmid 9$.

When an integer does not divide another, there is a remainder left over.

Theorem 4.2 (The Division Algorithm)**.** *Let m be an integer and n a positive integer. Then there exist unique integers q (the* quotient*) and r (the* remainder*) which satisfy the following conditions:*

4.1.1 $0 \leq r < n$.

4.1.2 $m = qn + r$.

The theorem should be read as saying that n goes q times into m with r left over.

Examples. 4.1.1 7 goes into 23 three times with 2 left over: an elementary school student would write '$23 \div 7 = 3$ remainder 2.' In the language of the Division Algorithm, we have $m = 23$ and $n = 7$. We look for the smallest integer $r \geq 0$ so that $23 - r$ is divisible by 7: since $7 \mid 21$ we choose $r = 2$. The quotient is $q = 3$ and we write

$$23 = 3 \cdot 7 + 2$$

4.1.2 Similarly, if $m = -11$ and $n = 3$, then $q = -4$ and $r = 1$, since

$$-11 = (-4) \cdot 3 + 1$$

For practice, find a formula for all the integers that have remainder 4 *after division by* 6.

The proof of the Division Algorithm relies on the development of induction, to which we will return in Chapter 5. For our purposes, the point of the division algorithm is that every integer m has a nicely-defined remainder r when divided by n. This allows us to construct an alternative form of arithmetic.

Definition 4.3. Let a and b be integers, and n a positive integer. We say that a **is congruent to** b **modulo** n and write

$$a \equiv b \pmod{n}$$

if a and b have the *same remainder* upon dividing by n. The integer n is called the **modulus**. When the modulus is unambiguous we tend simply to write $a \equiv b$.

Examples. We write $7 \equiv 10 \pmod{3}$, since both 7 and 10 have the same remainder ($r = 1$) on division by 3.
Since 6 and 10 do not have the same remainder on division by 3, we would write $6 \not\equiv 10 \pmod{3}$.

Can you find a formula for *all* the integers that are congruent to 10 modulo 3?

For a little practice with the notation, consider the following conjectures, where a is any integer. Are they true or false?

Conjecture 4.4. $a \equiv 8 \pmod{6} \implies a \equiv 2 \pmod{3}$.

Conjecture 4.5. $a \equiv 2 \pmod{3} \implies a \equiv 8 \pmod{6}$.

The first conjecture is true. Indeed, if $a \equiv 8 \pmod{6}$, we can write $a = 6k + 8$ for some integer k. Then

$$a = 6k + 8 = 6k + 6 + 2 = 3(2k + 2) + 2$$

and so a has remainder 2 upon division by 3, showing that a is congruent to 2 modulo 3.
On the other hand, the second conjecture is false. All we need is a counterexample. Consider $a = 5$: clearly a is congruent to 2 modulo 3. However a has remainder 5 on division by 6, whereas 8 has remainder 2. Therefore a and 8 do not have the same remainder and are not congruent modulo 6.

Reasoning and calculating in the above fashion is tedious. What is useful is to tie the concept of congruence to that of divisibility. The following theorem is crucial, and provides an equivalent definition of congruence.

Theorem 4.6. *Let a and b be integers and n a positive integer. Then $a \equiv b \pmod{n} \iff n \mid (b-a)$.*

Proof. There are two separate theorems here, although both rely on the Division Algorithm (Theorem 4.2) to divide both a and b by n. Given a, b, n, the Division Algorithm shows that there exist unique quotients q_1, q_2 and remainders r_1, r_2 which satisfy

$$a = q_1 n + r_1, \qquad b = q_2 n + r_2, \qquad 0 \le r_1, r_2 < n. \tag{$*$}$$

Now we perform both directions of the proof.

($\Rightarrow$) Suppose that $a \equiv b \pmod{n}$. By definition, this means that a and b have the same remainder when divided by n. That is, $r_1 = r_2$. Subtracting a from b gives us

$$b - a = (q_2 - q_1)n + (r_2 - r_1) = (q_2 - q_1)n,$$

which is divisible by n. Therefore $n \mid (b-a)$.

($\Leftarrow$) This direction is a more subtle. We assume that $b - a$ is divisible by n. Thus $b - a = kn$ for some integer k. Invoking $(*)$, we see that

$$\begin{aligned} r_2 - r_1 &= (b - q_2 n) - (a - q_1 n) = (b-a) - (q_2 - q_1)n \\ &= (k - q_2 + q_1)n \end{aligned}$$

is also a multiple of n. Now consider the condition on the remainders in $(*)$: since $0 \le r_1, r_2 < n$, we quickly see that

$$\begin{cases} 0 \le r_2 < n \\ -n < -r_1 \le 0 \end{cases} \implies -n < r_2 - r_1 < n.$$

This says that $r_2 - r_1$ is a multiple of n lying strictly between $\pm n$. The only possibility is that $r_2 - r_1 = 0$. Otherwise said, $r_2 = r_1$, whence a and b have the same remainder, and so $a \equiv b \pmod{n}$. ■

If you are having trouble with the final step, think about an example. Suppose that $n = 26$ and that and that $x = r_2 - r_1$ is an *integer* satisfying the two conditions:

$$\begin{cases} x \text{ is divisible by } 26 \\ -26 < x < 26 \end{cases}$$

The strict inequalities should make it obvious that $x = 0$.

To gain some familiarity with congruence, try using Theorem 4.6 to show that

$$a \equiv b \pmod{n} \iff b \equiv a \pmod{n}.$$

Note that this expression and the theorem both contain a hidden quantifier ($\forall a, b \in \mathbb{Z}$), as discussed in Section 2.2. Moreover, combining the theorem with Definition 4.1 leads to the observation that

$$\begin{aligned} a \equiv b \pmod{n} &\iff \exists k \in \mathbb{Z} \text{ such that } b - a = kn \\ &\iff b = a + kn \text{ for some integer } k \end{aligned}$$

Congruence and Divisibility

The previous two theorems may appear a little abstract, so it's a good idea to recap the relationship between congruence and divisibility. The following observations should be immediate to you.

Let a be any integer and let n be a positive integer. Then

- a is congruent to *exactly one* of the integers $0, 1, 2, \ldots, n-1$ modulo n.
- a is divisible by n if and only if $a \equiv 0 \pmod{n}$.
- a is *not* divisible by n if and only if $a \equiv 1,\ 2,\ 3, \ldots,$ or $n-1$ modulo n.

To test your level of comfort with the definition of congruence, and review some proof techniques, prove the following theorem.

Theorem 4.7. *Suppose that n is an integer. Then*

$$n^2 \not\equiv n \pmod 3 \iff n \equiv 2 \pmod 3.$$

If you don't know how to start, try completing the following table before writing a formal proof:

n	n^2	Is $n^2 \equiv n \pmod 3$?
0	0	Yes
1		
2		

That the congruence sign $\equiv$ appears similar to the equals sign $=$ is no accident. In many ways it behaves exactly the same. In Section 7.3 we shall see that congruence is an important example of an *equivalence relation:* these generalize the notion of equality. Indeed, two integers are congruent if and only if something about them is equal, namely their remainders.

Modular Arithmetic

The arithmetic of remainders is almost exactly the same as the more familiar arithmetic of real numbers, but comes with all manner of fun additional applications, most importantly cryptography and data security: cell-phones and computers perform millions of these calculations every day! Here we spell out the basic rules of congruence arithmetic.[11]

Theorem 4.8. *Suppose that a, b, c, d are integers, and that all congruences are modulo the same integer n.*

4.1.1 $a \equiv b$ *and* $c \equiv d \implies ac \equiv bd$

4.1.2 $a \equiv b$ *and* $c \equiv d \implies a \pm c \equiv b \pm d$

[11] The usual associative, commutative and distributive laws of arithmetic

$$a+(b+c) \equiv (a+b)+c, \qquad a(bc) \equiv (ab)c, \qquad a+b \equiv b+a, \qquad ab \equiv ba, \qquad a(b+c) \equiv ab+ac$$

all follow because $x = y \implies x \equiv y \pmod{n}$, regardless of n: equal numbers have the same remainder after all!

What the theorem says is that the operations of 'take the remainder' and 'add' (or 'multiply') can be performed in any order or combination, the result will be the same.

Example. Consider $a = 29$, $b = 14$ and $n = 6$. We could add a and b then take the remainder when dividing by n:

$$29 + 14 = 43 = 6 \cdot 7 + 1 \implies 29 + 14 \equiv 1 \pmod 6.$$

Alternatively we could take the remainders of a and b modulo n and then add these:

$$5 + 2 = 7, \quad \text{which has the } \textit{same remainder } 1 \text{ modulo } 6.$$

Either way, we may write the result as a congruence,

$$29 + 14 \equiv 1 \pmod 6.$$

Proof of Theorem 4.8. Suppose that $a \equiv b$ and $c \equiv d$. By Theorem 4.6 we have $a - b = kn$ and $c - d = ln$ for some integers k, l. It follows that

$$\begin{aligned} ac &= (b + kn)(d + ln) = bd + n(bl + kd + kln) \\ \implies ac - bd &= n(bl + kd + kln) \end{aligned}$$

which is divisible by n. Hence $ac \equiv bd$.
Try the second argument yourself. ∎

The ability to take remainders *before* adding and multiplying is remarkably powerful, and allows us to perform some surprising calculations.

Examples. 4.1.1 What is the remainder when 39^{23} is divided by 10? At the outset this question appears impossible to answer. Ask your calculator and it will tell you that $39^{23} \approx 3.93 \times 10^{36}$, which is of no assistance; we need to discover the *units* digit of 39^{23}, whereas your calculator reports only a few of the significant digits at the other end of the number.
Instead of relying on a calculator, we think about the rules of arithmetic modulo 10. Since $39 \equiv 9 \equiv -1 \pmod{10}$, we quickly notice that

$$39 \cdot 39 \equiv (-1) \cdot (-1) \equiv 1 \pmod{10},$$

whence $39^2 \equiv 1 \pmod{10}$. Since positive integer exponents signify repeated multiplication, we can repeat the exercise to obtain

$$39^{23} \equiv \underbrace{(-1) \cdot (-1) \cdots (-1)}_{23 \text{ times}} = (-1)^{23} \equiv -1 \equiv 9 \pmod{10}$$

Therefore 39^{23} has remainder 9 when divided by 10. Otherwise said, the last digit of 39^{23} is a 9. If you ask a computer for all the digits you can check this yourself.

4.1.2 Now that we understand powers, more complex examples become easy. Here we compute modulo $n = 6$.

$$7^9 + 14^3 \equiv 1^9 + 2^3 \equiv 1 + 8 \equiv 9 \equiv 3 \pmod 6.$$

Hence $7^9 + 14^3 = 40356351$ has remainder 3 when divided by 6.

4.1.3 Find the remainder when $124^{12} \cdot 65^{49}$ is divided by 11. This time we need to perform multiple calculations to reduce these large numbers to something manageable. Since $124 = 11^2 + 3$ and $65 = 11 \cdot 6 - 1$, we write

$$\begin{aligned} 124^{12} \cdot 65^{49} &\equiv 3^{12} \cdot (-1)^{49} \equiv 27^4 \cdot (-1) \equiv 5^4 \cdot (-1) \\ &\equiv -(25^2) \equiv -(3^2) \equiv 2 \pmod{11} \end{aligned}$$

The remainder is therefore 2. There is no way to do this on a pocket calculator, since the original number $124^{12} \cdot 65^{49} \approx 9 \times 10^{113}$ is far too large to work with!

There are two points to stress when performing these calculations:

4.1.1 You are trying to replace each integer with something which has the same remainder and is *small*: thus $124 \equiv 3 \pmod{11}$ is more helpful than $124 \equiv -8 \pmod{11}$, since powers of 3 are easier to work with than powers of 8.

4.1.2 You may only reduce the *base* of an exponential expression modulo n, not the exponent! It is correct to write $17^{23} \equiv 3^{23} \pmod 7$, but you *cannot* claim that this is congruent to 3^2.

Division and Congruence The primary difference between modular and normal arithmetic is, perhaps unsurprisingly, with regard to *division.*

Theorem 4.9. *Suppose a and b are integers and k and n are positive integers. If $ka \equiv kb \pmod{kn}$ then $a \equiv b \pmod n$.*

The modulus is divided by k as well as the terms, so the meaning of $\equiv$ changes. In Exercise 4.1.14 you will prove this theorem, and observe that, in general, we do not expect $a \equiv b \pmod{kn}$.

Reading Questions

4.1.1 Which of the following connectives makes the following true for any $a, b \in \mathbb{Z}$ and $n \in \mathbb{N}$?

$$a \equiv b \pmod n \quad \underline{\qquad} \quad a = b.$$

(a) $\Longrightarrow$

(b) $\Longleftarrow$

(c) $\Longleftrightarrow$

(d) $\wedge$

4.1.2 Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. Is it possible that there are multiple pairs of integers q and r such that $m = qn + r$ and $0 \le r < n$?

(a) It is never possible.
(b) It is sometimes possible, depending on what m and n are.
(c) It is always possible.

4.1.3 Which of the following are true statements for $a, b \in \mathbb{Z}$ and $n \in \mathbb{N}$? Select all that apply.

(a) a is congruent to exactly one of $0, 1, \ldots, n-1$ modulo n.
(b) a can be congruent to more than one of $0, 1, \ldots, n-1$ modulo n.
(c) a is divisible by n if and only if $a \equiv 0 \pmod{n}$.
(d) $n \equiv 0 \pmod{n}$.

Practice Problems

4.1.1 Use the Division Algorithm to show that any prime number $p \geq 5$ must have remainder 1 or 5 upon division by 6. Use this to show that $p^2 + 2$ is composite for all such primes p.
Video Solution

4.1.2 Find the remainder of $57^{33} + 42^{100}$ upon division by 6.
Video Solution

4.1.3 Prove that $n^2 \equiv 0 \pmod{4}$ or $n^2 \equiv 1 \pmod{4}$ for all $n \in \mathbb{Z}$.
Video Solution

Exercises

4.1.1 Check explicitly that $3^{23} \not\equiv 3^2 \pmod{7}$.

4.1.2 Find the remainder when $22^9 + 29^{24}$ is divided by 10.

4.1.3 Compute the remainder when 43^{10} is divided by 13.

4.1.4 Find all integers x which satisfy the congruence equation $5x \equiv 2 \mod 8$.

4.1.5 Find the remainder when $17^{251} \cdot 23^{12} - 19^{41}$ is divided by 5. *Hint:* $17 \equiv 2$ *and* $2^2 \equiv -1 \pmod{5}$.

4.1.6 Find the remainder when $12^{10} + 2^{36} \cdot 18^{12}$ is divided by 141. *Hint: what nice number is close to 141? Use a calculator to help with some of the sums.*

4.1.7 Is the following statement identical to Theorem 4.7? Why/why not?

$$n^2 \equiv n \pmod{3} \iff n \equiv 0 \pmod{3} \text{ or } n \equiv 1 \pmod{3},$$

4.1.8 Prove the first part of Theorem 4.8: that if $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $a + c \equiv b + d \pmod{n}$.

4.1.9 Let $a, b, c \in \mathbb{Z}$ and $n \in \mathbb{N}$. Prove

(a) $a \equiv a \pmod{n}$

(b) if $a \equiv b \pmod{n}$ then $b \equiv a \pmod{n}$

(c) if $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$ then $a \equiv c \pmod{n}$.

4.1.10 Prove that if $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$ then $3a - c^2 \equiv 3b - d^2 \pmod{n}$.

4.1.11 Find a natural number n and integers a, b such that $a^2 \equiv b^2 \pmod{n}$ but $a \not\equiv b \pmod{n}$.

4.1.12 (a) Let n be a positive integer. Prove that n is congruent to the sum of its digits modulo 9. *Hint: first consider an example such as* $345 = 3 \cdot 10^2 + 4 \cdot 10 + 5 \ldots$

(b) Is the integer 123456789 divisible by 9?

4.1.13 Let p be a prime number greater than or equal to 3. Show that if $p \equiv 1 \pmod{3}$, then $p \equiv 1 \pmod{6}$. *Hint: p is odd.*

4.1.14 Suppose that $7x \equiv 28 \pmod{42}$. By Theorem 4.9, it follows that $x \equiv 4 \pmod{6}$.

(a) Check this explicitly using Theorem 4.6.

(b) If $7x \equiv 28 \pmod{42}$, is it possible that $x \equiv 4 \pmod{42}$?

(c) Is it always the case that $7x \equiv 28 \pmod{42} \implies x \equiv 4 \pmod{42}$? Why/why not?

(d) Prove Theorem 4.9.

4.1.15 If $a \mid b$ and $b \mid c$, prove that $a \mid c$.

4.1.16 Suppose $a, b, c \in \mathbb{Z}$ and $a \mid b$ and $a \mid c$. Prove that for any $x, y \in \mathbb{Z}$, we have $a \mid (bx + cy)$.

4.1.17 Let a, b be positive integers. Prove that $a = b \iff a \mid b$ and $b \mid a$.

4.1.18 Decide whether each conjecture is true or false and prove/disprove your assertions.

Conjecture 1: $a \mid b$ and $a \mid c \implies a \mid bc$.

Conjecture 2: $a \mid c$ and $b \mid c \implies ab \mid c$.

4.1.19 Fermat's Little Theorem (to distinguish it from his 'Last') states that if p is prime and $a \not\equiv 0 \mod p$, then $a^{p-1} \equiv 1 \pmod{p}$.

(a) Use Fermat's Little Theorem to prove that $b^p \equiv b \pmod{p}$ for *any* integer b.

(b) Prove that if p is prime then $p \mid (2^p - 2)$.

(c) Find a counterexample to the converse: some non-prime n such that $n \mid 2^n - 2$.

4.1.20 Abraham Lincoln was born on February 12$^{\text{th}}$ 1809. On what day of the week was this? More generally, describe how to find the weekday given any date (in the Gregorian calendar).

4.1.21 For $n \in \mathbb{N}$, show

$$\frac{n(n+1)(2n+1)}{6}$$

is an integer.

4.1.22 Consider numbers of the form

$$\underbrace{11\cdots 11}_{n \text{ times}}$$

for $n \geq 2$.

(a) Prove every such number can be written as $4k+3$ for some $k \in \mathbb{Z}$. [For example, $11 = 4(2)+3$ and $111 = 4(27)+3$.]

(b) Use part (a) to show that no such number is a square.

4.1.23 Prove that $3 \mid (4^n - 1)$ for all $n \geq 1$.

4.1.24 Prove that for any integer n, one of $n, n+2, n+4$ is divisible by 3.

4.1.25 Let $n \in \mathbb{Z}$.

(a) Find all possible remainders of n^2 upon division by 7.

(b) Find all possible remainders of n^3 upon division by 7.

(c) Use parts (a) and (b) to show that any number which is simultaneously a square and a cube must be of the form $7k$ or $7k+1$ for some integer k.

4.1.26 Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. The Division Algorithm states that there exist unique integers q and r such that $m = qn + r$ and $0 \leq r < n$. While we will wait until the next chapter to see a full proof, we can give a proof of the *uniqueness* part now. Fill in the blanks in the following proof of the uniqueness part of the theorem.

Proof. Replace absolute values with more intuitive approach. The standard proof technique for uniqueness proofs is to assume there are two objects satisfying the conditions of the statement under question and proceeding to show that these objects are the same. Towards this goal, suppose that there exist two pairs of integers q, r and q', r' satisfying the conclusion of the Division Algorithm:

$$m = qn + r, \qquad m = q'n + r'$$

and $0 \leq r, r' < n$. We show $q = q'$ and $r = r'$.

Then

$$r - r' = \underline{\hspace{3cm}}.$$

Taking absolute values of both sides, we have

$$\underline{\hspace{2.5cm}} = \underline{\hspace{3cm}}.$$

Since $0 \leq r' < n$, we have $-n < \underline{\hspace{1.2cm}} \leq 0$. Adding this inequality to $0 \leq r < n$, we get $-n < \underline{\hspace{2.5cm}} < n$. In other words, $|r - r'| < n$. Hence $n|q' - q| < n$ as well. Dividing by

n, we have $|q' - q| < 1$. But since $|q' - q|$ is a positive integer, this means that $|q' - q| =$ ______.
Thus $q = q'$. It then follows that ____________ as well. ∎

4.1.27 Let $n \in \mathbb{N}$. Prove that $\sqrt{4n+6}$ is not an integer. Hint. You may use the following lemma without proof: $\forall k \in \mathbb{Z},\ k^2 \equiv 0 \text{ or } 1 \pmod 4$.

4.2 Greatest Common Divisors and the Euclidean Algorithm

At its most basic, Number Theory involves finding *integer* solutions to equations. Here are two simple-sounding questions:

4.2.1 The equation $9x - 21y = 6$ represents a straight line in the plane. Are there any *integer points* on this line? That is, can you find integers x, y satisfying $9x - 21y = 6$?

4.2.2 What about on the line $4x + 6y = 1$?

Before you do anything else, try sketching both lines (lined graph paper will help) and try to decide if there are any integer points. If there are integer points, how many are there? Can you find them all?

In this section we will see how to answer these questions in general: for which lines $ax + by = c$ with $a, b, c \in \mathbb{Z}$, are there integer solutions, and how can we find them all? The method introduces the appropriately named *Euclidean algorithm,* a famous procedure dating at least as far back as Euclid's *Elements* (c. 300 BCE.).

Definition 4.10. Let m and n be integers, not both zero. Their **greatest common divisor** $\gcd(m, n)$ is the largest (positive) divisor of both m and n. We say that m and n are **relatively prime** (or **coprime**) if $\gcd(m, n) = 1$.

Example. Let $m = 60$ and $n = 90$. The positive divisors of the two integers are listed in the table:

m	1	2	3	4	5	6	10	12	15	20	<u>30</u>	60
n	1	2	3	5	6	9	10	15	18	<u>30</u>	45	90

The greatest common divisor is the largest number common to both rows: clearly $\gcd(60, 90) = 30$.

Finding the greatest common divisor of two integers by listing all the positive divisors of both numbers is extremely inefficient, especially when the integers are large. This is where Euclid rides to the rescue.

Euclidean Algorithm. To find $\gcd(m, n)$ for two positive integers $m > n$:

(i) Use the Division Algorithm (Theorem 4.2) to write $m = q_1 n + r_1$ with $0 \le r_1 < n$.

(ii) If $r_1 = 0$, then n divides m and so $\gcd(m, n) = n$. Otherwise, repeat:
If $r_1 > 0$, divide n by r_1 to obtain $n = q_2 r_1 + r_2$ with $0 \le r_2 < r_1$.

(iii) If $r_2 = 0$, then $\gcd(m, n) = r_1$. Otherwise, repeat:
If $r_2 > 0$, divide r_1 by r_2 to obtain $r_1 = q_3 r_2 + r_3$ with $0 \le r_3 < r_2$.

(iv) Repeat the process, obtaining a decreasing sequence of non-negative integers

$$r_1 > r_2 > r_3 > \ldots \ge 0$$

Theorem 4.11. *The Algorithm eventually produces a remainder of zero: there exists p such that $r_{p+1} = 0$. The greatest common divisor of m and n is then the last non-zero remainder:* $\gcd(m,n) = r_p$.

The proof is in the exercises. If m and n are not both positive, take absolute values first and apply the algorithm. For instance $\gcd(-6, 45) = 3$.

Example. We compute $\gcd(1260, 750)$ using the Euclidean Algorithm. Since each line of the algorithm is a single case of the Division Algorithm $m = qn + r$, you might find it easier to create a table and observe each remainder moving diagonally left and down at each successive step.

$$1260 = 1 \times 750 + 510$$
$$750 = 1 \times 510 + 240$$
$$510 = 2 \times 240 + 30$$
$$240 = 8 \times 30 + 0$$

m	q	n	r
1260	1	750	510
750	1	510	240
510	2	240	30
240	8	30	0

Theorem 4.11 says that $\gcd(1260, 750) = 30$, the last non-zero remainder.

As you can see, the Euclidean Algorithm is very efficient.

Reversing the Algorithm: Integer Points on Lines

To apply the Euclidean Algorithm to the problem of finding integer points on lines, we must reverse it. We start with the penultimate line of the algorithm and substitute the remainders from the previous lines one at a time: the result is an expression of the form $\gcd(m,n) = mx + ny$ for some integers x, y. This is easiest to demonstrate by continuing our example.

Example (continued)**.** We find integers x, y such that $1260x + 750y = 30$.

Solve for 30 (the gcd of 1260 and 750) using the third step of the algorithm:

$$30 = 510 - 2 \times 240.$$

Now use the second line of the algorithm to solve for 240 and substitute:

$$30 = 510 - 2 \times (750 - 510) = 3 \times 510 - 2 \times 750.$$

Finally, substitute for 510 using the first line:

$$30 = 3 \times (1260 - 750) - 2 \times 750 = 3 \times 1260 - 5 \times 750.$$

Rearranging this, we see that the integers $x = 3$ and $y = -5$ satisfy the equation $1260x + 750y = 30$. Otherwise said, the integer point $(3, -5)$ lies on the line with equation $1260x + 750y = 30$.

Note how the process for finding an integer point (x, y) is twofold: first we compute $\gcd(m,n)$ using the Euclidean Algorithm, then we perform a series of back-substitutions to recover x and y.

This process of reversing the algorithm works in general, and we have the following corollary of Theorem 4.11.

Corollary 4.12 (Bézout's Identity)**.** *Given integers m, n, not both zero, there exist integers x, y such that*

$$\gcd(m,n) = mx + ny.$$

We are now in a position to solve our motivating problem: finding all integer points on the line $ax + by = c$ where a, b, c are integers. Again we appeal first to our example.

Example (take III)**.** We have already found a single integer solution $(x, y) = (3, -5)$ to the equation $1260x + 750y = 30$. Notice that the equation is equivalent to dividing through by the greatest common divisor $30 = \gcd(1260, 750)$:

$$42x + 25y = 1$$

Since 42 and 25 have no common factors, it seems that the only way to alter x and y while keeping the equation in balance is to increase x by a multiple of 25 and decrease y by the same multiple of 42. For example $(x, y) = (3 + 25, -5 - 42) = (28, -47)$ is another solution. Indeed, all integer solutions are given by

$$(x, y) = (3, -5) + (25, -42)t, \quad \text{where } t \text{ is any integer.}$$

In general, we have the following result.

Theorem 4.13. *Let a, b, c be integers where a, b are non-zero, and let $d = \gcd(a, b)$. Then the equation $ax + by = c$ has an integer solution (x, y) if and only if $d \mid c$.*
In such a case, suppose that (x_0, y_0) is some fixed solution. Then all integer solutions are given by

$$x = x_0 + \frac{b}{d}t, \qquad y = y_0 - \frac{a}{d}t, \tag{$*$}$$

where t is any integer.

The general approach is to use the Euclidean Algorithm to find the initial solution (x_0, y_0), then to apply $(*)$ to obtain all solutions.[12] The proof is again in the exercises.

Warning! If $c \neq \gcd(a, b)$, you will need to modify the integers obtained in Bézout's Identity in order to find the initial solution (x_0, y_0). For example, since $1260 \times 3 + 750 \times (-5) = 30$ we multiply by 3 to see that $(x_0, y_0) = (9, -15)$ is an initial solution to $1260x + 750y = 90$. All integer points on this line therefore have the form

$$(x, y) = (9 + 25t, -15 - 42t), \text{ where } t \in \mathbb{Z}$$

[12]The astute observer should recognize the similarity between this and the complementary function/particular integral method for linear differential equations: (x_0, y_0) is a 'particular solution' to the full equation $ax + by = c$, while $(\frac{b}{d}t, -\frac{a}{d}t)$ comprises all solutions to the 'homogeneous equation' $ax + by = 0$.

Examples. 4.2.1 Consider the line $570x - 123y = 7$. We calculate the greatest common divisor using the Euclidean algorithm: note that the negative sign is irrelevant.

$$\left.\begin{aligned} 570 &= 4\times 123 + 78 \\ 123 &= 1\times 78 + 45 \\ 78 &= 1\times 45 + 33 \\ 45 &= 1\times 33 + 12 \\ 33 &= 2\times 12 + 9 \\ 12 &= 1\times 9 + 3 \\ 9 &= 3\times 3 + 0 \end{aligned}\right\} \implies \gcd(570,123) = 3.$$

Since $3 \nmid 7$, we conclude that the line $570x - 123y = 7$ contains no integer points.

4.2.2 Applied to the line with equation $570x - 123y = -6$, we reverse the algorithm to obtain

$$\begin{aligned} 3 &= 12 - 9 = 12 - (33 - 2\times 12) \\ &= 3\times 12 - 33 = 3(45 - 33) - 33 \\ &= 3\times 45 - 4\times 33 = 3\times 45 - 4(78 - 45) \\ &= 7\times 45 - 4\times 78 = 7(123 - 78) - 4\times 78 \\ &= 7\times 123 - 11\times 78 = 7\times 123 - 11(570 - 4\times 123) \\ &= 570\times(-11) - 123\times(-51) \end{aligned}$$

Multiplying by -2 so that our solution conforms to the desired equation, it follows that $(x_0, y_0) = (22, 102)$ is an initial solution. The general solution is then

$$(x,y) = (22,102) + \left(-\frac{123}{3}, -\frac{570}{3}\right)t = (22 - 41t, 102 - 190t)$$

Reading Quiz

4.2.1 True or False: $\gcd(-21,-12) = -3$.

4.2.2 Suppose that $a \neq 0$. Then $\gcd(a, 0)$ is equal to which number?

(a) 0
(b) 1
(c) a
(d) $|a|$

4.2.3 The sequence of remainders produced by the Euclidean Algorithm when computing $\gcd(m, n)$ (select all that apply)

(a) is decreasing
(b) is increasing

(c) has all non-negative terms

(d) is infinite

4.2.4 True or False: If a and b are relatively prime then the equation $ax + by = 1$ has an integer solution (x, y).

Practice Problems

4.2.1 Use the Euclidean Algorithm to compute $\gcd(260, 816)$. Then find integers x, y such that $260x + 816y = \gcd(260, 816)$.

Video Solution

4.2.2 Find solutions to the congruence $5x \equiv 1 \pmod 6$.

Video Solution

4.2.3 Find all integer points on the line $225x + 120y = 15$.

Video Solution

4.2.4 Suppose $a, b, c \in \mathbb{Z}$ are such that a and b are relatively prime, $a \mid c$, and $b \mid c$. Show $ab \mid c$. [Sketch proof and redo].

Video Solution

Exercises

4.2.1 Use the Euclidean Algorithm to compute the greatest common divisors indicated.

(a) $\gcd(20, 12)$ (b) $\gcd(100, 36)$ (c) $\gcd(207, 496)$

4.2.2 For each part of Question 4.2.1, find integers x, y which satisfy Bézout's Identity $\gcd(m, n) = mx + ny$.

4.2.3 (a) Answer our motivating problems from the beginning of the section using the above process.

(i) Find all integer points on the line $9x - 21y = 6$.

(ii) Show that there are no integer points on the line $4x + 6y = 1$.

(b) Can you give an elementary proof as to why there are no integer points on the line $4x + 6y = 1$?

4.2.4 Find all the integer points on the following lines, or show that none exist.

(a) $16x - 33y = 2$.

(b) $122x + 36y = 3$.

(c) $303x + 204y = 6$.

(d) $324x - 204y = -12$.

4.2.5 Show that there exists no integer x such that $3x \equiv 5 \pmod 6$.

4.2.6 Find all solutions x to the congruence equation $12x \equiv 1 \pmod{17}$

4.2.7 Five people each take the same number of candies from a jar. Then a group of seven people does the same: in so doing they empty the jar. If the jar originally contained 239 candies. Can you be sure how much candies each person took?

4.2.8 Here we sketch a proof that the Euclidean Algorithm (Theorem 4.11) terminates with $r_p = \gcd(m, n)$. Note that you *cannot* use Bézout's Identity in to prove any of what follows, since it is a corollary of the algorithm.

(a) Suppose you have a decreasing sequence

$$m > n > r_1 > r_2 > \cdots \geq 0 \tag{$*$}$$

of positive integers. Explain why the sequence can only have *finitely many* terms. This shows that the Euclidean Algorithm eventually terminates with some $r_{p+1} = 0$.

(b) Suppose that $m = qn + r$ for some integers m, n, q, r. Prove that $\gcd(m, n) \,|\, r$.

(c) Explain why $\gcd(m, n) \,|\, r_p$.

(d) Explain why r_p divides all of the integers in the sequence $(*)$, in particular that $r_p \,|\, m$ and $r_p \,|\, n$.

(e) Explain why $r_p \leq \gcd(m, n)$. Why does this force us to conclude that $r_p = \gcd(m, n)$?

4.2.9 Suppose that $d \,|\, m$ and $d \,|\, n$. Prove that $d \,|\, \gcd(m, n)$.

4.2.10 Prove the following:

$$\gcd(m, n) = 1 \iff \exists x, y \in \mathbb{Z} \text{ such that } mx + ny = 1.$$

One direction can be done by applying Bézout's Identity, but the other direction requires an argument.

4.2.11 Let $a, b, c \in \mathbb{Z}$.

(a) Suppose $a \mid bc$ and $\gcd(a, b) = 1$. Show $a \mid c$.

(b) Use part (a) to show that if p is a prime and $p \mid ab$, then either $p \mid a$ or $p \mid b$.

(c) Show that if $n > 1$ is a number with the property that for any a, b, $n \mid ab$ implies $n \mid a$ or $n \mid b$, then n is a prime number.

4.2.12 Show that if a is relatively prime to b, and a is relatively prime to c, then a is relatively prime to bc.

4.2.13 In this question we prove the Theorem 4.13 on integer solutions to linear equations. Let $a, b, c \in \mathbb{Z}$. Suppose that (x_0, y_0) and (x_1, y_1) are two integer solutions to the linear Diophantine equation $ax + by = c$.

(a) Show that $(x_0 - x_1, y_0 - y_1)$ satisfies the equation $ax + by = 0$.

(b) Suppose that $\gcd(a, b) = d$. Prove that $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$. (*Use Question 4.2.10)*

(c) Find all integer solutions (x, y) to $ax + by = 0$ (*Don't use the Theorem, it's what you're trying to prove! Think about part (b) and divide through by d first.*).

(d) Use (a) and (b) to conclude that (x, y) is an integer solution to $ax + by = c$ if and only if

$$x = x_0 + \frac{b}{d}t \qquad y = y_0 - \frac{a}{d}t, \qquad \text{where } t \in \mathbb{Z}.$$

4.2.14 Show that $\gcd(5n + 2, 12n + 5) = 1$ for every integer n. *There are two ways to approach this: you can try to use the Euclidean algorithm abstractly, or you can use the result of Exercise 4.2.10.*

4.2.15 Use the Euclidean Algorithm to show that for any $k \in \mathbb{N}$, we have $\gcd(ka, kb) = k \gcd(a, b)$.

4.2.16 Let n be a positive integer. Complete the table

n	1	2	3	4	5	6
$\gcd(2n, n+1)$						

Now make a conjecture for the value of $\gcd(2n, n + 1)$ and prove it.

4.2.17 For nonzero integers a and b, the **least common multiple** $\operatorname{lcm}(a, b)$ is defined to be the least positive integer m which is a multiple of both a and b.

(a) If $m = \operatorname{lcm}(a, b)$, $a \mid c$, and $b \mid c$, show $m \mid c$.

(b) If a and b are both positive, show $\gcd(a, b) \operatorname{lcm}(a, b) = ab$.

4.2.18 The set of remainders $\mathbb{Z}_n = \{0, 1, 2, \ldots, n - 1\}$ is called a *ring* when equipped with addition and multiplication modulo n. For example $5 + 6 \equiv 3 \pmod 8$. We say that $b \in \mathbb{Z}_n$ is an *inverse* of $a \in \mathbb{Z}_n$ if

$$ab \equiv 1 \pmod n.$$

(a) Show that 2 has no inverse modulo 6.

(b) Show that if $n = n_1 n_2$ is composite ($\exists$ integers $n_1, n_2 \geq 2$) then there exist elements of the ring $\mathbb{Z}_n$ which have no inverses.

(c) Prove that a has an inverse modulo n if and only if $\gcd(a, n) = 1$. Conclude that the only sets $\mathbb{Z}_n$ for which all non-zero elements have inverses are those for which n is prime.
You will find Exercise 4.2.10 helpful.

5 Mathematical Induction and Well-ordering

In Section 2.3 we discussed three methods of proof: direct, contrapositive, and contradiction. The fourth standard method of proof, *induction,* has a very different flavor. In practice it formalizes the idea of spotting a pattern. Before we give the formal definition of induction, we consider where induction fits into the investigative process.

5.1 Investigating Recursive Processes

In applications of mathematics, one often has a simple recurrence relation but no general formula. For instance, a process might be described by an expression of the form

$$x_{n+1} = f(x_n),$$

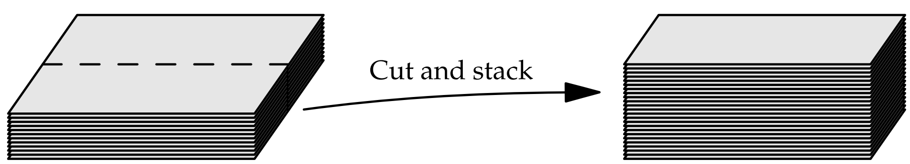
where some initial value x_1 is given. While investigating such recurrences, you might hypothesize a *general formula*

$$x_n = g(n).$$

Induction is a method of proof that allows us to *prove* the correctness of such general formulæ. Here is a simple example of the process.

Stacking Paper

Consider the operation whereby you take a stack of paper, cut all sheets in half, then stack both halves together.



If a single sheet of paper has thickness 0.1 mm, how many times would you have to repeat the process until the stack of paper reached to the sun? ($\approx$ 150 million kilometers).

The example is describing a recurrence relation. If h_n is the height of the stack after n operations, then we have a sequence $(h_n)_{n=0}^{\infty}$ satisfying

$$\begin{cases} h_{n+1} = 2h_n \\ h_0 = 0.1\,\text{mm}. \end{cases}$$

It is easy to compute the first few terms of the sequence:

n	0	1	2	3	4	5	6	7	8	$\cdots$
h_n (mm)	0.1	0.2	0.4	0.8	1.6	3.2	6.4	12.8	25.6	$\cdots$

It is not hard to hypothesize that, after n such operations, the stack of paper will have height

$$h_n = 2^n \times 0.1\,\text{mm}.$$

All we have done is to spot a pattern. We can reassure ourselves by checking that the first few terms of the sequence satisfy the formula: certainly $h_0 = 2^0 \times 0.1\,\text{mm}$ and $h_1 = 2^1 \times 0.1\,\text{mm}$, etc. Unfortunately the sequence has *infinitely many* terms, so we need a trick which confirms *all of them at once.* Unless we can *prove* that our formula is correct *for all* $n \in \mathbb{N}_0$ it will remain just a guess. This is where induction steps in.

The trick is called the *induction step*. We *assume* that we have already confirmed the formula for some fixed, but unspecified, value of n and then use what we know (the recurrence relation $h_{n+1} = 2h_n$) to confirm the formula for the *next value* $n+1$. Here it goes:

Induction Step Suppose that $h_n = 2^n \times 0.1$ mm, for some fixed $n \in \mathbb{N}_0$. Then

$$h_{n+1} = 2h_n = 2(2^n \times 0.1) = 2^{n+1} \times 0.1 \text{ mm}.$$

This is exactly the expression we hoped to find for the $(n+1)$th term of the sequence. Think about what the induction step is doing. By leaving n unspecified, we have proved an *infinite collection of implications at once!* Each implication has the form

$$h_n = 2^n \times 0.1 \implies h_{n+1} = 2^{n+1} \times 0.1.$$

Since the implications have been proved for all $n \in \mathbb{N}_0$, we can string them together:

$$h_0 = 2^0 \times 0.1 \implies h_1 = 2^1 \times 0.1 \implies h_2 = 2^2 \times 0.1 \implies h_3 = 2^3 \times 0.1 \implies \cdots$$

We have already checked that the first formula $h_0 = 2^0 \times 0.1$ in the implication chain is true. By the induction step, the *entire infinite collection of formulæ must be true.* We have therefore *proved* that

$$h_n = 2^n \times 0.1\,\text{mm} = 2^n \times 10^{-4}\,\text{m}, \quad \forall n \geq 0.$$

Now that we've proved the formula for every h_n, finishing the original problem is easy: we need to find $n \in \mathbb{N}_0$ such that

$$h_n = 2^n \times 10^{-4} \geq 150 \times 10^9\,\text{m} \iff 2^n \geq 15 \times 10^{14}.$$

Since logorithms are increasing functions, they preserve inequalities and we may easily solve to see that

$$n \geq \log_2(15 \times 10^{14}) = \log_2 15 + 14 \log_2 10 \approx 50.4.$$

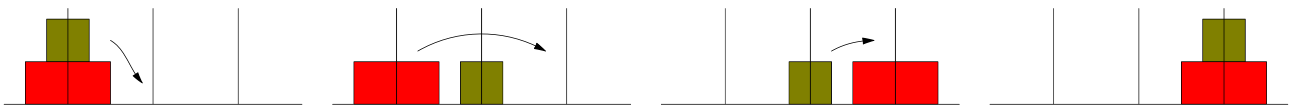
Thus 51 iterations of the cut-and-stack process are sufficient for the pile of paper to reach the sun!

We will formalize the discussion of induction in the next section so that you will never have to write as much as we've just done. However, it is important to remember how induction fits into a practical investigation. It is the missing piece of logic that turns a *guess* into a justified formula. Before we do so, here is a famous and slightly more complicated problem.

The Tower of Hanoi

The *Tower of Hanoi* is a game involving circular disks of decreasing radii stacked on three pegs. A 'move' consists of transferring the top disk in any stack onto a larger disk or an empty peg. If we start with n disks on the first peg, how many moves are required to transfer all the disks to one of the other pegs?

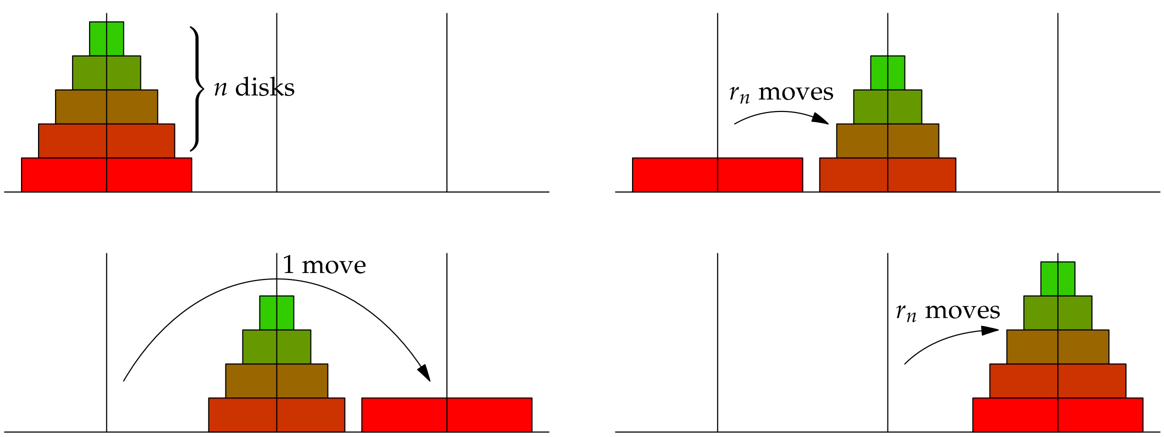
The challenge here is that we have no formula to play with, only the variable n for the number of disks. The first thing to do is to play the game. If the variable r_n represents the number of moves required when there are n disks, then it should be immediately clear that $r_1 = 1$: one disk only requires one move! The picture below shows that $r_2 = 3$.



With more disks you can keep experimenting and find that $r_3 = 7$, etc. At this point you may be ready to hypothesize a general formula.

Conjecture 5.1. *The Tower of Hanoi with n disks requires $r_n = 2^n - 1$ moves.*

Certainly the conjecture is true for $n = 1, 2$ and 3. To see that it is true in general, we need to think about how to move a stack of $n + 1$ disks. Since the largest disk can only be moved onto an empty peg, it follows that the n smaller disks must already be stacked on a single peg *before* the $(n + 1)$th disk can move. From the starting position this requires r_n moves.



The largest disk can now be moved to the final peg, before the original n disks are moved on top of it. In total this requires $r_n + 1 + r_n$ moves, as illustrated in the picture. We therefore have a recurrence relation for r_n:

$$\begin{cases} r_{n+1} = 2r_n + 1 \\ r_1 = 1. \end{cases}$$

We are now in a position to prove our conjecture. We know that the conjecture is true for $n = 1$ and we assume that the formula $r_n = 2^n - 1$ is true for some fixed but unspecified n. Now we use the recurrence relation to prove that $r_{n+1} = 2^{n+1} - 1$.

Induction Step Suppose that $r_n = 2^n - 1$ for some fixed $n \in \mathbb{N}$. Then

$$\begin{aligned} r_{n+1} &= 2r_n + 1 = 2(2^n - 1) + 1 && \text{(since we are } \textit{assuming } r_n = 2^n - 1\text{)} \\ &= 2^{n+1} - 2 + 1 = 2^{n+1} - 1 \end{aligned}$$

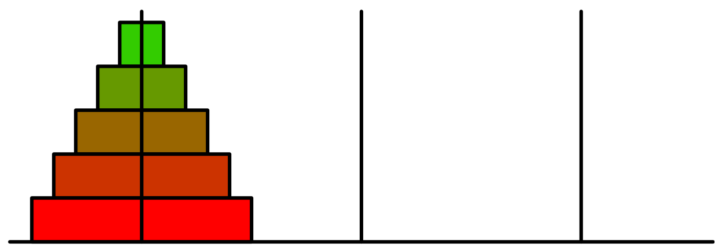
Exactly as in the paper-stacking example, we have simultaneously proved an *infinite collection of implications*:

$$r_1 = 2^1 - 1 \implies r_2 = 2^2 - 1 \implies r_3 = 2^3 - 1 \implies r_4 = 2^4 - 1 \implies \cdots$$

Since the first of these statements is true, it follows that *all of the others are true.* Hence Conjecture 5.1 is true, and becomes a theorem.

As an illustration of how ridiculously time-consuming the Tower becomes, the following table gives the time taken to complete the Tower if you were able to move one disk per second.

Disks	Time
5	31sec
10	17min 3sec
15	9hr 6min 7sec
20	12days 3hrs 16min 15sec
25	$\sim$ 1yr 23days
30	$\sim$ 34yrs 9days



Animation of five disks (click)

Exercises

5.1.1 A room contains n people. Everybody wants to shake everyone else's hand (but not their own).

(a) Suppose that n people require h_n handshakes. If an $(n+1)$th person enters the room, how many *additional* handshakes are required? Obtain a recurrence relation for h_{n+1} in terms of h_n.

(b) Hypothesize a general formula for h_n, and prove it using the method in this section.

5.1.2 Skippy the Kangaroo is playing jump rope, but he tires as the day goes on. The heights h_n (inches) of successive jumps are related by the recurrence

$$h_{n+1} = \frac{8}{9}h_n + 1.$$

(a) Suppose that Skippy's initial jump has height $h_1 = 100$ in. Show that Skippy fails to jump above 10in for the first time on the 40th jump.

(b) Find the *total* height jumped by Skippy in the first n jumps.

You may find it useful to define $H_n = h_n - 9$ and think about the recurrence for H_n. Now guess and prove a general formula for H_n. Finally, remind yourself about geometric series.)

5.2 Proof by Induction

The previous section motivated the need for induction and helped us see where induction fits into a logical investigation. In this section we formally lay out several induction proofs.

Induction is the mathematical equivalent of a domino rally; toppling the nth domino causes the $(n+1)$th domino to fall, hence to knock all the dominos over it is enough merely to topple the first. Instead of dominoes, in mathematics we consider a sequence of *propositions:* $P(1)$, $P(2)$, $P(3)$, etc. Induction demonstrates the truth of *every* proposition $P(n)$ by doing two things:

5.2.1 Proving that $P(1)$ is true (*Base Case*)

5.2.2 Proving that $\forall n \in \mathbb{N}\ P(n) \implies P(n+1)$ is true (*Induction Step*)

You could think of the base case as knocking over the first domino, and the induction step as the nth domino knocking over the $(n+1)$th, *for all n.* Both of the examples in the previous section followed this pattern.[13] Unpacking the induction step gives an infinite chain of implications:

$$P(1) \implies P(2) \implies P(3) \implies P(4) \implies P(5) \implies \cdots.$$

The base case says that $P(1)$ is true, and so *all* of the remaining propositions $P(2)$, $P(3)$, $P(4)$, $P(5)$, ... are also true.

All induction proofs have the same formal structure:

(*Set-up*) Define the propositional function $P(n)$, set-up notation and orient the reader as to what you are about to prove.

(*Base Case*) Prove that $P(1)$ is true.

(*Induction Step*) Let $n \in \mathbb{N}$ be fixed and assume that $P(n)$ is true. This assumption is the **induction hypothesis**. Perform calculations or other reasoning to conclude that $P(n+1)$ is true.

(*Conclusion*) Remind the reader what it is that you have proved.

As you read more mathematics, you will find that the induction step is typically the most involved part of the proof. The *set-up* stage is often no more than a sentence: 'We prove by induction,' and the explicit definition of $P(n)$ is commonly omitted. These are the only shortcuts that it is sensible to take until you are extremely comfortable with induction. Practice making it completely clear what you are doing at each juncture.

Here is a straightforward theorem, where we write the proof in the above language.

Theorem 5.2. *The sum of the first n positive integers is given by the formula*

$$\sum_{i=1}^{n} i = \frac{1}{2}n(n+1).$$

[13] In the cut-and-stack example, the initial proposition would be labelled $P(0)$ rather than $P(1)$.

Proof. (*Set-up*) We prove by induction. For each $n \in \mathbb{N}$, let $P(n)$ be the proposition

$$\sum_{i=1}^{n} i = \frac{1}{2}n(n+1).$$

(*Base Case*) Clearly $\sum_{i=1}^{1} i = 1 = \frac{1}{2}1(1+1)$, and so $P(1)$ is true.

(*Induction Step*) Assume that $P(n)$ is true for some fixed $n \geq 1$. We compute the sum of the first $n+1$ positive integers using our induction hypothesis $P(n)$ to simplify:

$$\begin{aligned}\sum_{i=1}^{n+1} i &= (n+1) + \sum_{i=1}^{n} i = (n+1) + \frac{1}{2}n(n+1) && \text{(by assumption of } P(n)\text{)}\\ &= \left(1 + \frac{1}{2}n\right)(n+1) = \frac{1}{2}(n+2)(n+1)\\ &= \frac{1}{2}(n+1)\big[(n+1)+1\big].\end{aligned}$$

This last says that $P(n+1)$ is true.

(*Conclusion*) By mathematical induction, we conclude that $P(n)$ is true for all $n \in \mathbb{N}$. That is

$$\forall n \in \mathbb{N}, \quad \sum_{i=1}^{n} i = \frac{1}{2}n(n+1).$$

■

Note how we grouped $\frac{1}{2}(n+1)\big[(n+1)+1\big]$ so that it is obviously the right hand side of $P(n+1)$.

Here is another example in the same vein, but done a little faster.

Theorem 5.3. *Prove that $n(n+1)(2n+1)$ is divisible by 6 for all natural numbers n.*

Proof. We prove by induction. For each $n \in \mathbb{N}$, let $P(n)$ be the proposition

$n(n+1)(2n+1)$ is divisible by 6.

(*Base Case*) Clearly $1 \cdot (1+1) \cdot (2 \cdot 1 + 1) = 6$ is divisible by 6, hence $P(1)$ is true.

(*Induction Step*) Assume that $P(n)$ is true for some fixed $n \in \mathbb{N}$. Then

$$n(n+1)(2n+1) = 6k$$

for some $k \in \mathbb{Z}$. But now we have

$$\begin{aligned}(n+1)(n+2)\big[2(n+1)+1\big] - n(n+1)(2n+1) &= (n+1)\big[(n+2)(2n+3) - n(2n+1)\big]\\ &= (n+1)(2n^2+7n+6-2n^2-n)\\ &= 6(n+1)^2.\end{aligned}$$

By the induction hypothesis, we have that

$$(n+1)(n+2)\left[2(n+1)+1\right] = n(n+1)(2n+1) + 6(n+1)^2 = 6(k + (n+1)^2)$$

is divisible by 6. Thus $P(n+1)$ is true. By mathematical induction, $P(n)$ is true for all $n \in \mathbb{N}$. ■

Theorem 5.3 is also true for $n = 0$, and indeed for *all* integers n. As we shall see in the next section, induction works perfectly well with any base case (say $n = 0$): you are not tied to $n = 1$. We could even modify the argument to prove the same result when n is a negative integer!

After reading the proof, you are possibly thinking, 'How would I know to do that calculation?' The answer is that you wouldn't, at least not without *experience reading proofs.* It is better to think on how much scratch work was done before the originator stumbled on exactly this argument. Read more proofs and practice writing them, and you'll soon find that strategies like these will suggest themselves!

Here is another example, written in a more advanced style: we don't explicitly name the propositions $P(n)$, and the reader is expected to be familiar enough with induction to realize when we are covering the base case and the induction step. If you find reading this proof a challenge, you should rewrite it in the same style as we used previously. Some assistance in this regard is given below.

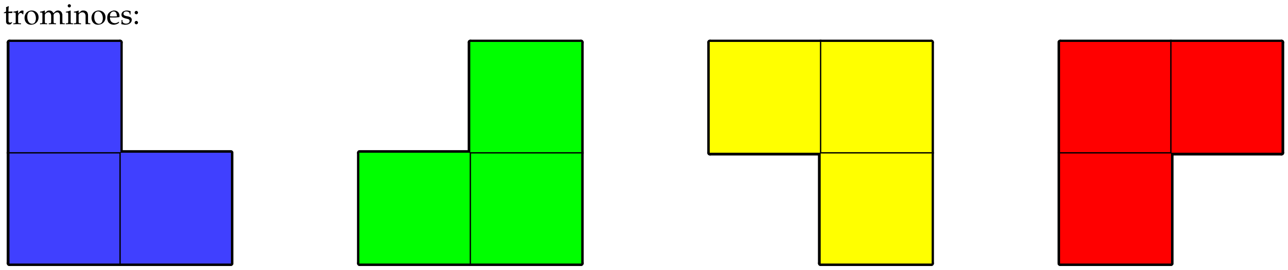
Theorem 5.4. *For all $n \in \mathbb{N}$, $2 + 5 + 8 + \cdots + (3n-1) = \frac{1}{2}n(3n+1)$.*

Proof. For $n = 1$ we have $2 = 2$, hence the proposition holds. Now suppose that the proposition holds for some fixed $n \in \mathbb{N}$. Then

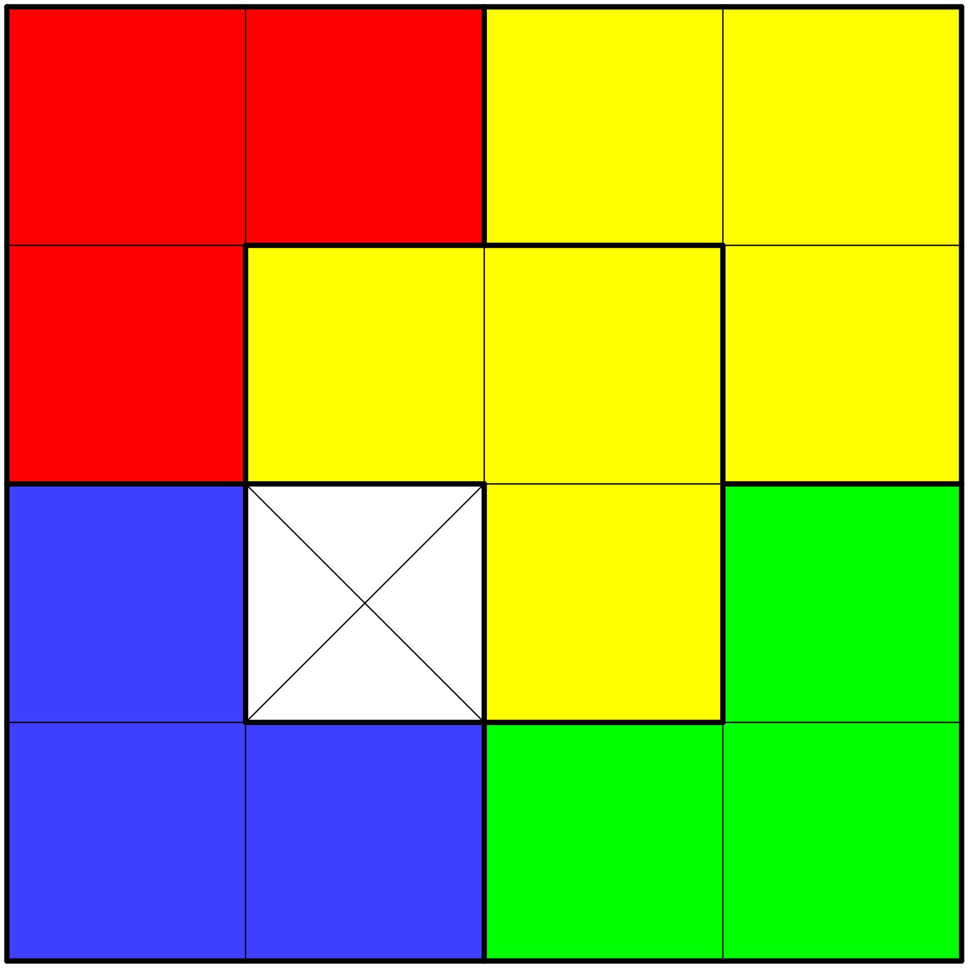
$$\begin{aligned} 2 + 5 + \cdots + [3(n+1)-1] &= [2 + 5 + \cdots + (3n-1)] + 3n + 2 \\ &= \frac{1}{2}n(3n+1) + 3n + 2 = \frac{1}{2}(3n^2 + 7n + 4) \\ &= \frac{1}{2}(n+1)(3n+4) = \frac{1}{2}(n+1)\left[3(n+1)+1\right] \end{aligned}$$

which says that the proposition holds for $n+1$. By mathematical induction the proposition holds for all $n \in \mathbb{N}$. ■

This last example has a different flavor than the ones we have seen so far. The example concerns a $2^n \times 2^n$ board of squares. By an **L-shaped tromino**, we mean three squares arranged in an "L" shape (though possibly flipped or rotated). For example, the following are examples of L-shaped trominoes:



The result is that if one takes any $2^n \times 2^n$ board and removes any one square, the rest of the board may be tiled by L-shaped triominoes. Here is an example of the 4×4 case:



Theorem 5.5. *Let $n \in \mathbb{N}$. Then any $2^n \times 2^n$ board of squares may be tiled by L-shaped trominoes after removing any square.*

Proof. We proceed by induction on n. For $n = 1$, we look at a 2×2 board. It should be clear that no matter which of the four squares we choose to exclude, the remaining three squares form an L-shaped tromino (if you are unsure, try to draw the board).

Now fix $n \in \mathbb{N}$ and suppose that given a $2^n \times 2^n$ board and *any* choice of one of the squares to exclude, we can tile the rest of the board with trominoes. Now take a $2^{n+1} \times 2^{n+1}$ board and pick an arbitrary square to remove. Divide the board into four quadrants Q_1, Q_2, Q_3, Q_4, where each of which are boards of size $2^n \times 2^n$. The removed square must lie in one of these quadrants, without loss of generality say it is in quadrant Q_1. By the induction hypothesis, it is possible to tile Q_1 minus the removed square by trominoes. Now consider the other three quadrants and choose one corner of each. Again by the induction hypothesis, it is possible to tile each of these quandrants with the chosen corner removed by trominoes. Finally, rotate the three remaining quadrants so that their chosen corners lie adjacent in the center of the board. These removed corners now form an L-shaped tromino, and can thus be covered by one more tromino. ■

Scratch work is your friend! Once you are comfortable with the structure of an induction proof, the challenge is often in finding a clear argument for the induction step. Don't dive straight into the proof! First try some scratch calculations. Be creative, since the same approach will not work for all proofs.

One of the benefits of explicitly stating $P(n)$ is that it helps you to isolate what you know and to identify your goal. When stuck, write down both expressions $P(n)$ and $P(n+1)$ and you will often see how to proceed. Consider, for example, the proof of Theorem 5.4. We have:

$$\begin{aligned} P(n): &\quad 2+5+8+\cdots+(3n-1) = \frac{1}{2}n(3n+1). \\ P(n+1): &\quad 2+5+8+\cdots+[3(n+1)-1] = \frac{1}{2}(n+1)\big[3(n+1)+1\big] \end{aligned}$$

Simply by writing these down, we know that our goal is to somehow convert the left hand side of $P(n+1)$ into the right hand side, using $P(n)$. In this situation it is clear how to proceed, for almost all of the left hand side of $P(n+1)$ can be substituted for that of $P(n)$.

As a final comment on scratch work, remember that such is *very unlikely* to constitute a proof. Here is a typical attempt at a proof of Theorem 5.4 by someone who is new to induction.

False Proof.

$$\begin{aligned} P(n+1) = \quad & \underbrace{2+5+\cdots+(3n-1)}_{=\frac{1}{2}n(3n+1)\text{ by }P(n)} + [3(n+1)-1] = \frac{1}{2}(n+1)\big[3(n+1)+1\big] \\ & = \frac{1}{2}(n+1)(3n+4) \\ \implies \quad & \frac{3}{2}n^2 + \frac{1}{2}n + 3n + 3 - 1 = \frac{1}{2}(3n^2+7n+4) \\ \implies \quad & \frac{3}{2}n^2 + \frac{7}{2}n + 2 = \frac{3}{2}n^2 + \frac{7}{2}n + 2 \end{aligned}$$

■

Such an approach is likely to score very poorly in an exam! Here are some of the reasons why.

- $P(n+1)$ is the *goal*, the conclusion of the induction step. You cannot prove $P(n) \implies P(n+1)$ by *starting* with $P(n+1)$!
- $P(n+1)$ is a *proposition* and $2+5+\cdots+(3n-1)+[3(n+1)-1]$ is a *number*, thus it makes no sense to write that they are equal! Use words or another symbol to disambiguate the two.
- More subtly: the false proof's argument says that something we don't know ($P(n) \wedge P(n+1)$) implies something true (the trivial final line). Since the implications $T \implies T$ and $F \implies T$ are both true (Definition 2.3), this tells us *nothing* about whether $P(n+1)$ is true.
- Reversing the arrows and turning the false proof upside down would be a start. However there is no explanation as to *why* the calculation is being done. The induction step is only part of an induction proof and it needs to be placed and explained in context. More concretely:

– There is no set-up. $P(n)$ has not been defined, neither indeed has n. You cannot use the expression $P(n)$ (or any other symbols) in a proof unless it has been properly defined.
– The base case is missing.
– There is no conclusion. Indeed the word *induction* isn't mentioned: is the reader supposed to guess that we're doing induction?!

For all this negativity, there are some good things here. If you remove the $\Longrightarrow$ symbols, you are left with an excellent piece of scratch work. By simplifying both sides of your goal you can more easily see how to calculate.

Your scratch work may make perfect sense to you, but if a reader cannot follow it without your assistance, then it isn't a proof. The moral of the story is to do your scratch work for the induction step *then* lay out the structure of the proof (set-up, base case, etc.) before incorporating your calculation into a coherent and convincing argument.

Reading Questions

5.2.1 In an induction proof of the fact that $P(n)$ is true for all $n \in \mathbb{N}$, the base case consists of proving that

(a) $P(1)$ is false.
(b) $P(1)$ is true.
(c) for all n, $P(n) \implies P(n+1)$.
(d) $P(1) \implies P(2)$.

5.2.2 In an induction proof of the fact that $P(n)$ is true for all $n \in \mathbb{N}$, the induction hypothesis is the assumption that

(a) $P(1)$ is true.
(b) for all n, $P(n) \implies P(n+1)$.
(c) $P(n)$ is true for some fixed $n \in \mathbb{N}$.
(d) $P(n)$ is true for all $n \in \mathbb{N}$.

5.2.3 True or False: in our formal proofs, it is acceptable to write

$$P(n) = \sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$$

as shorthand for "$P(n)$ is the proposition $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$".

Practice Problems

5.2.1 (a) Prove by induction that $\forall n \in \mathbb{N}$ we have $3 \mid (2^n + 2^{n+1})$.
(b) Give a direct proof that $3 \mid (2^n + 2^{n+1})$ for all integers $n \geq 1$ *and* for $n = 0$.
(c) Look carefully at your proof for part (a). If you had started with the base case $n = 0$ instead of $n = 1$, would your proof still be valid?
Video Solution

Exercises

5.2.1 (a) Complete Gauss' direct proof of Theorem 5.2.

(b) Give a direct proof of Theorem 5.3.

(c) In Theorem 5.3, what is the proposition $P(n+1)$?

(d) In the Induction Step of Theorem 5.3, explain why it would be incorrect to write

$$\begin{aligned} P(n+1) - P(n) &= (n+1)\big[(n+2)(2n+3) - n(2n+1)\big] \\ &= (n+1)(2n^2 + 7n + 6 - 2n^2 - n) \\ &= 6(n+1)^2. \end{aligned}$$

5.2.2 Prove by induction that for each natural number n, we have $\sum_{j=0}^{n} 2^j = 2^{n+1} - 1$.

5.2.3 Consider the following Theorem: If n is a natural number, then

$$\sum_{k=1}^{n} k^3 = \frac{1}{4} n^2 (n+1)^2$$

(a) What explicitly is the meaning of $\sum_{k=1}^{4} k^3$?

(b) What would be meant by the expression $\sum_{k=1}^{n} n^3$, and why is it different to $\sum_{k=1}^{n} k^3$?

(c) If the Theorem is written in the form $\forall n \in \mathbb{N},\ P(n)$, what is the proposition $P(n)$?

(d) Give as many reasons as you can as to why the following 'proof' of the induction step is incorrect.

$$\begin{aligned} P(n+1) = \sum_{k=1}^{n+1} k^3 &= \frac{1}{4}(n+1)^2((n+1)+1)^2 \\ = \sum_{k=1}^{n} k^3 + (n+1)^3 &= \frac{1}{4}(n+1)^2(n+2)^2 \\ = \frac{1}{4}n^2(n+1)^2 + (n+1)^3 &= \frac{1}{4}(n+1)^2(n+2)^2 \\ = \frac{1}{4}(n+1)^2\left[n^2 + 4(n+1)\right] &= \frac{1}{4}(n+1)^2(n+2)^2 \\ = \frac{1}{4}(n+1)^2(n+2)^2 &= \frac{1}{4}(n+1)^2(n+2)^2 \end{aligned}$$

(e) Give a correct proof of the Theorem by induction.

5.2.4 Show by induction that for every $n \in \mathbb{N}$ we have: $n \equiv 5 \pmod 3$ or $n \equiv 6 \pmod 3$ or $n \equiv 7 \pmod 3$.

5.2.5 Prove by induction that, for all $n \in \mathbb{N}$,

$$1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n(n+1) = \frac{1}{3} n(n+1)(n+2)$$

5.2.6 (a) Show, by induction, that for all $n \in \mathbb{N}$, the number 4 divides the integer $11^n - 7^n$.

(b) More generally, use induction to prove that $(a-b) \mid (a^n - b^n)$ for any positive integers a, b, n.

5.2.7 Prove that for all $k \in \mathbb{N}$, we have that $8^k - 1$ is a multiple of 7.

5.2.8 (a) Find a formula for the sum of the first n odd natural numbers. Prove your assertion by induction.

(b) Give an alternative direct proof of your formula from part (a). You may use results such as $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$.

5.2.9 We mimic the previous question for the sum of the squares of the first n natural numbers.

(a) Use the fact that $\sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1)$ to compute directly an expression for the sum of the squares of the first n *odd* natural numbers.

Hint: $\sum_{i=1}^{n}(2i-1)^2 = \sum_{i=1}^{2n} i^2 - \sum_{i=1}^{n}(2i)^2 \dots$

(b) Prove the truth of your formula by induction.

5.2.10 Find the error in the following "proof" by induction of the statement "all cats have the same color fur".

> *Proof.* We let $P(n)$ be the proposition "any set of n cats have the same color fur". The result will follow if we prove that $P(n)$ holds for all $n \in \mathbb{N}$. We proceed by induction on n.
>
> It is clear that the base case $n = 1$ holds as any cat has the same color fur as itself. For the induction step, fix $n \in \mathbb{N}$ and assume $P(n)$ holds. Take any set $S = \{C_1, C_2, \dots, C_{n+1}\}$ of $n+1$ cats. Select one cat, say C_1, and put it aside. Now we have a set $S \setminus \{C_1\}$ of n cats and by the induction hypothesis they must all have the same color fur. Now put C_1 back select a different cat C_2. Again by the induction hypothesis, all cats in $S \setminus \{C_2\}$ must have the same color fur. But combining this with the previous sentence means that all cats in S must have the same color fur. Since S was an arbitrary set of $n+1$ cats, this shows $P(n+1)$ holds. We conclude that $P(n)$ is true for all $n \in \mathbb{N}$ by induction.
>
> ■

5.2.11 Using only the product rule and the fact that $\frac{d}{dx}x = 1$, prove the power rule from calculus: for all $n \geq 1$,

$$\frac{d}{dx}(x^n) = nx^{n-1}.$$

5.2.12 Recall that a polynomial is a function $\mathbb{R} \to \mathbb{R}$ of the form $p(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_2 x^2 + a_1 x + a_0$. The numbers a_i are called the *coefficients* and the *degree* of p is the largest d such that the coefficient $a_d \neq 0$.

(a) Prove that for all $n \in \mathbb{N}$,

$$\frac{d^n}{dx^n} e^{x^2} = p(x) e^{x^2}$$

where $p(x)$ is a polynomial.

(b) Strengthen the result in part (a) by proving that for all $n \in \mathbb{N}$,

$$\frac{d^n}{dx^n} e^{x^2} = p(x) e^{x^2} = p_n(x) e^{x^2}$$

where $p_n(x)$ is a polynomial of degree n.

5.2.13 (Hard) Let $p(x)$ be a polynomial of degree $d \geq 1$. Show p has at most d roots. [Hint: induct on the degree d.]

5.2.14 Let $a, b \in \mathbb{Z}$.

(a) Prove that for all $n \in \mathbb{N}$, if a and b are relatively prime, then a^{2^n} and b^{2^n} are relatively prime.

(b) Use part (a) to show that for all $k \in \mathbb{N}$, if a and b are relatively prime, then a^k and b^k are relatively prime.

5.2.15 Consider the following scratch work. Determine what result is being proved, then convert the scratch work into a formal proof of that result.

$$\begin{aligned}
(1+x)^{n+1} &= (1+x)^n(1+x) \\
&\geq (1+nx)(1+x) \\
&= 1 + x + nx + nx^2 \\
&= 1 + (n+1)x + nx^2 \\
&\geq 1 + (n+1)x
\end{aligned}$$

5.2.16 Prove that for any $n \geq 1$,

$$\sum_{i=1}^{n} \frac{1}{i^2} < 2.$$

[Hint: prove the stronger fact that $\sum_{i=1}^{n} \frac{1}{i^2} < 2 - \frac{1}{n}$ for all $n \geq 1$.]

5.3 Well-ordering and the Principle of Mathematical Induction

Before seeing more examples, it is worth thinking more carefully about the logic behind induction. The fact that induction really works depends on a fundamental property of the natural numbers.

Definition 5.6. A set of real numbers A is **well-ordered** if every non-empty subset of A has a minimum element.

The definition is delicate: to test if a set A is well-ordered, we need to check *all* of its non-empty subsets. The definition could be written as follows:

$\forall B \subseteq A$ such that $B \neq \emptyset$, we have that $\min(B)$ exists.

Consequently, to show that a set A is *not* well-ordered, we need only exhibit a non-empty subset B which has *no minimum.*

Examples. 5.3.1 $A = \{4, -7, \pi, 19, \ln 2\}$ is a well-ordered set. There are 31 non-empty subsets of A, each of which has a minimum element. Can you justify this fact *without* listing the subsets?

5.3.2 The interval $[3, 10)$ is not well-ordered. Indeed $(3, 4)$ is a non-empty subset which has no minimum element (see the exercises).

5.3.3 The integers $\mathbb{Z}$ are not well-ordered. For instance, $\mathbb{Z}$ is a non-empty subset of itself, and there is no minimum integer.

More generally, every finite set of numbers is well-ordered, while intervals are not. Are there any *infinite* sets which are also well-ordered? The answer is *yes.* Indeed it is part of the standard definition (Peano's Axioms) of the natural numbers that $\mathbb{N}$ is such a set.

Axiom. $\mathbb{N}$ is well-ordered.

Any set that 'looks like' $\mathbb{N}$ is automatically well-ordered.[14] For example

$$B = \left\{0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots\right\} = \left\{\frac{n}{n+1} : n \in \mathbb{N}\right\}$$

Armed with this axiom, we can justify the method of proof by induction.

Theorem 5.7 (Principle of Mathematical Induction)**.** *Let $P(n)$ be a proposition for each $n \in \mathbb{N}$. Suppose:*

(a) $P(1)$ is true.

(b) $\forall n \in \mathbb{N}\, P(n) \implies P(n+1)$.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

[14] When the elements are written in increasing order, the set has the form $B = \{b_1, b_2, b_3, b_4, \ldots\}$.

Proof. We argue by contradiction. Assume that conditions (a) and (b) hold and that $\exists n \in \mathbb{N}$ such that $P(n)$ is *false.* Then the set

$$S := \{k \in \mathbb{N} : P(k) \text{ is false}\}$$

is a non-empty subset of the well-ordered set $\mathbb{N}$. It follows that S has a minimum element

$$m := \min(S)$$

Note that $P(m)$ is *false.*
By condition (a), $P(1)$ is true, and so $m \neq 1$. Therefore $m \geq 2$ from which we see that $m - 1 \in \mathbb{N}$.
Since $m = \min(S)$ it follows that $m - 1 \notin S$ and so $P(m-1)$ must be *true.*
However, by condition (b), we see that $P(m-1) \implies P(m)$, whence $P(m)$ is *true.*
This is a contradiction. In addition to properties (a) and (b), our only assumption was that *at least one* proposition $P(n)$ was false, therefore this is what we have contradicted. We conclude that conclude that $P(n)$ is true for all $n \in \mathbb{N}$.

∎

Different Base Cases for Induction

An induction argument need not begin with the case $n = 1$. By proving Theorem 5.7 it should be clear where we used the well-ordering of $\mathbb{N}$ in order to justify induction. Now fix an integer m (positive, negative or zero) and consider the set

$$\mathbb{Z}_{\geq m} = \{n \in \mathbb{Z} : n \geq m\} = \{m, m+1, m+2, m+3, \ldots\}.$$

This set is well-ordered, whence the following modification of the induction principle is immediate.

Corollary 5.8. *Let $m \in \mathbb{Z}$ be some fixed integer. Let $P(n)$ be a proposition for each integer $n \geq m$. Suppose:*

(a) $P(m)$ is true.

(b) $\forall n \geq m\, P(n) \implies P(n+1)$.

Then $P(n)$ is true for all $n \geq m$.

We are simply changing the base case. The induction concept is exactly the same as before:

$$P(m) \implies P(m+1) \implies P(m+2) \implies P(m+3) \implies \cdots$$

As long as you explicitly prove the first claim in the sequence, and you show the induction step, then all the propositions are true.

Here is an example where the induction argument begins with $m = 4$.

Theorem 5.9. *For all integers $n \geq 4$, we have $3^n > n^3$.*

Proof. (*Base Case*) If $n = 4$, we have $3^n = 81 > 64 = n^3$. The proposition is therefore true for $n = 4$.
(*Induction Step*) Fix $n \in \mathbb{Z}_{\geq 4}$ and suppose that $3^n > n^3$. Then

$$3^{n+1} = 3 \cdot 3^n > 3n^3.$$

To finish the proof, we want to see that this right hand side is at least $(n+1)^3$. Now

$$3n^3 \geq (n+1)^3 \iff 3 \geq \left(1 + \frac{1}{n}\right)^3$$

This is true for $n = 3$ and, since the right hand side is decreasing as n increases, it is certainly true when $n \geq 4$. We therefore conclude, for $n \geq 4$, that

$$3^n > n^3 \implies 3^{n+1} > (n+1)^3$$

which is the induction step. By induction, we have shown that $3^n > n^3$ whenever $n \in \mathbb{Z}_{\geq 4}$. ∎

Our next example is reminiscent of sequences and series from elementary calculus. If you follow a textbook derivation of such a formula, you'll probably see liberal use of ellipsis dots (...). When you see these, it is often because the author is hiding an induction argument.

Theorem 5.10. *For all integers $n \geq 3$, we have*

$$\sum_{i=3}^{n} \frac{1}{i(i-2)} = \frac{3}{4} - \frac{2n-1}{2n(n-1)}. \tag{$*$}$$

Proof. (*Base Case*) When $n = 3$, $(*)$ reads $\sum\limits_{i=3}^{3} \frac{1}{i(i-2)} = \frac{3}{4} - \frac{5}{12}$. Both sides equal $\frac{1}{3}$, whence $(*)$ is true.
(*Induction Step*) Assume that $(*)$ is true for some fixed $n \geq 3$. Then

$$\begin{aligned}
\sum_{i=3}^{n+1} \frac{1}{i(i-2)} &= \sum_{i=3}^{n} \frac{1}{i(i-2)} + \frac{1}{(n+1)(n-1)} \\
&= \frac{3}{4} - \frac{2n-1}{2n(n-1)} + \frac{1}{(n+1)(n-1)} && \text{(by the induction hypothesis)} \\
&= \frac{3}{4} - \left[\frac{(2n-1)(n+1) - 2n}{2(n+1)n(n-1)}\right] = \frac{3}{4} - \left[\frac{1+n-2n^2}{2(n+1)n(n-1)}\right] \\
&= \frac{3}{4} + \frac{(2n+1)(1-n)}{2(n+1)n(n-1)} = \frac{3}{4} - \frac{2n+1}{2(n+1)n}
\end{aligned}$$

which is exactly $(*)$ when n is replaced by $n+1$.
By induction $(*)$ holds for all integers $n \geq 3$. ∎

A calculus discussion would finish by taking the limit as $n \to \infty$ to conclude that $\sum\limits_{i=3}^{\infty} \frac{1}{i(i-2)} = \frac{3}{4}$.

Our final example involves a little abstraction.

Theorem 5.11. *The interior angles of an n-gon (n-sided polygon) sum to $180(n-2)$ degrees.*

We will take the initial case ($n = 3$) that the angles of a triangle sum to 180° as given (can you prove it?) and merely prove the induction step. The main logical difficulty is that we must consider *all* n-gons simultaneously. If we were to write the induction step in the form

$$\forall n \in \mathbb{Z}_{\geq 3},\ P(n) \implies P(n+1),$$

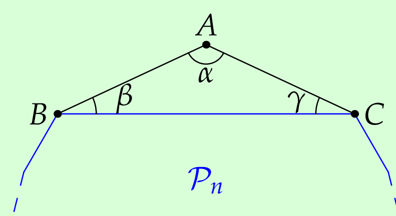
then the proposition $P(n)$ would be

$$P(n): \quad \forall n\text{-gons } \mathcal{P}_n, \text{ the sum of the interior angles of } \mathcal{P}_n \text{ is } 180(n-2)°.$$

To prove our induction step for a *fixed* integer n, we must show that *all* $(n+1)$-gons have the correct sum of interior angles. We therefore assume that we are given some $(n+1)$-gon $\mathcal{P}_{n+1}$ and proceed to compute its interior angles in terms of a related n-gon.

Proof. Fix an integer $n \geq 3$, and suppose that *all* n-gons have interior angles summing to $180(n-2)°$. Suppose we are given an $(n+1)$-gon $\mathcal{P}_{n+1}$. Select any vertex A and label the adjacent vertices B and C. Delete A, and join B and C with a straight edge. The result is an n-gon $\mathcal{P}_n$. There are two cases to consider.[15]

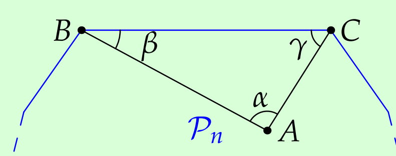
Case 1: The deleted point A is *outside* $\mathcal{P}_n$. The sum of the interior angles of $\mathcal{P}_{n+1}$ exceeds those of P_n by $\alpha + \beta + \gamma = 180°$. Therefore $\mathcal{P}_{n+1}$ has interior angles summing to $180(n-2)° + 180° = 180[(n+1)-2]°$.



Case 1: A outside $\mathcal{P}_n$

Case 2: The deleted point A is *inside* $\mathcal{P}_n$. To obtain the sum of the interior angles of $\mathcal{P}_{n+1}$, we take the sum of the interior angles of $\mathcal{P}_n$ and do three things:

- *Subtract* β
- *Subtract* γ
- *Add the reflex angle* $360°-\alpha$ *at* A



Case 2: A inside $\mathcal{P}_n$

We are therefore adding an additional

$$-\beta - \gamma + (360° - \alpha) = 360° - (\alpha + \beta + \gamma) = 180°$$

$\mathcal{P}_{n+1}$ again has interior angles summing to $180[(n+1)-2]°$. ∎

[15] We are obscuring two subtleties here. It is a fact, though not an obvious one, that it is always possible to choose a vertex A so that the new polygon $\mathcal{P}_n$ doesn't cross itself. Read about 'ears' and 'mouths' of polygons and triangulation if you're interested. There are also two other, less likely, cases which we didn't consider: when deleting a point from an $(n+1)$-gon it is possible to obtain an $(n-1)$-gon, or even an $(n-2)$-gon. To think it out, try drawing a 12-gon in the shape of a Star of David. Deleting one of the outer corners creates a 9-gon! Dealing with these cases strictly requires strong induction, so we return to them later.

Optional: Density of the Rationals

In our last example, we offer a more direct application of $\mathbb{N}$ being well-ordered. One of the key properties of the rational numbers $\mathbb{Q}$ is their density in the real line. Intuitively, the idea is that no matter how close you "zoom in" on the real line, you can always locate a rational number. We formalize this with the following definition.

Definition 5.12. We say a set $A \subseteq \mathbb{R}$ is **dense** (in $\mathbb{R}$) if for any real numbers x and y such that $x < y$, there is $a \in A$ such that $x < a < y$.

So if you take two real numbers, you can always find an element from A in between them, no matter how close the two real numbers are from each other. Our goal will be to prove that the rational numbers $\mathbb{Q}$ are dense in $\mathbb{R}$. For this, we will use the well-orderedness of $\mathbb{N}$ along with the following:

Axiom. The real numbers $\mathbb{R}$ have the **Archimedean property**, that is, for any real numbers $x, y > 0$, there is $n \in \mathbb{N}$ such that $nx > y$.

It is not really necessary to take this as an axiom as the Archimedean property of $\mathbb{R}$ can be proved from more basic principles. However, this requires some knowledge about how to construct the real numbers which lies beyond the scope of this course. Back to our goal, we need the following lemma which states that if two real numbers differ by more than 1, then there must be an integer between them.

Lemma 5.13. *Suppose we have $x, y \in \mathbb{R}$ with $y - x > 1$. Then there exists $k \in \mathbb{Z}$ such that $x < k < y$.*

Proof. The idea is to take k to be the least integer greater than x. We will show such an integer exists using the fact that $\mathbb{N}$ is well-ordered. Let $A = \{n \in \mathbb{Z} : n > x\}$. Then $A \neq \emptyset$ by the Archimedean property (why?). Let $m \in \mathbb{Z}$ be a number such that $m < x$ (this is another application of the Archimedean property), and thus $m < n$ for all $n \in A$, by definition of A. Let

$$S = \{n - m + 1 : n \in A\}.$$

So $S \subseteq \mathbb{N}$ and since $A \neq \emptyset$, we have $S \neq \emptyset$. Since $\mathbb{N}$ is well-ordered, S has a minimum element s. Then $k = s + m - 1$ is the minimum element of A (why?).

By definition $x < k$. But by minimality of k, $k - 1 \notin A$, i.e., $x \geq k - 1$. Thus $x < k \leq x + 1$. Finally, since $y - x > 1$, we have $x + 1 < y$. All together, $x < k \leq x + 1 < y$. So k is as required. ∎

Now we can prove our main result.

Theorem 5.14. *The rational numbers $\mathbb{Q}$ are dense in $\mathbb{R}$.*

Proof. Let $x, y \in \mathbb{R}$ with $x < y$ be arbitrary. We need to find $r \in \mathbb{Q}$ with $x < r < y$. Then $y - x > 0$. By the Archimedean property, there is $n \in \mathbb{N}$ such that $n(y - x) > 1$. Since $ny - nx > 1$, we can apply Lemma 5.13 to get $k \in \mathbb{Z}$ such that $nx < k < ny$. As $n \geq 1 > 0$, dividing yields $x < \frac{k}{n} < y$. Take $r = \frac{k}{n} \in \mathbb{Q}$. ■

Aside. Well-ordering more generally

Well-ordering is a fundamental concept whose implications are far beyond what we're discussing here. Informally speaking, *well-ordering* a set A involves listing the elements of A in some order so that every non-empty subset of A has a first element *with respect to that order.*

Consider, for example, the set of negative integers $\mathbb{Z}^-$. For the purposes of these notes we will always consider the standard ordering:

$$\cdots < -4 < -3 < -2 < -1.$$

Written in the standard order, $\mathbb{Z}^- = \{\ldots, -4, -3, -2, -1\}$ is *not* a well-ordered set. In a more advanced discussion, one could consider alternative orderings, and the definition of well-ordered would change accordingly. If we choose the ordering $\prec$ where

$$-1 \prec -2 \prec -3 \prec \cdots, \qquad (*)$$

then $\mathbb{Z}^-$ would be well-ordered using $\prec$ as the order: if $B \subseteq \mathbb{Z}^-$ is non-empty and has its elements listed in the same order as $(*)$, then B has a minimum element (with respect to $\prec$). With a little thinking, we could modify the proof of the principle of mathematical induction to allow us to prove theorems of the form $\forall n \in \mathbb{Z}^-,\ P(n)$, by induction. The base case is $n = -1$ and the induction step justifies the chain

$$P(-1) \implies P(-2) \implies P(-3) \implies \cdots$$

An extremely important theorem in advanced set theory states that it is possible to well-order *every* set. With a slight modification of the process, this massively increases the applicability of induction. In these notes we keep things simple: well-ordering is always in the sense of Definition 5.6, where we list the elements of a set in the usual increasing order. For a more esoteric example of a well-ordered set, see the final Exercise below.

Reading Quiz

5.3.1 Which of the following statements are true? Select all that apply.

- (a) Every well-ordered set of real numbers has a minimum element.
- (b) If a set of real numbers has a minimum element, then it is well-ordered.
- (c) Any finite set of real numbers is well-ordered.
- (d) Induction proofs must have a base case of 0 or 1.

5.3.2 The fact that $\mathbb{N}$ is well-ordered is considered a(n)

(a) theorem
(b) opinion
(c) axiom
(d) proof

5.3.3 True or False: a finite set can be dense in $\mathbb{R}$.

Practice Problems

5.3.1 Prove that $n! > 2^n$ for all $n \geq 4$.
Video Solution

5.3.2 Fill in the details in the proof of Lemma 5.13.
Video Solution

Exercises

5.3.1 Prove by contradiction that the interval $(3, 4)$ has no minimum element.

5.3.2 (a) Suppose that $n \geq 3$. Prove that $\left(\frac{n+1}{n}\right)^2 < 2$.
(b) Hence or otherwise, prove that $n^2 < 2^n$ for all natural numbers $n \geq 5$.

5.3.3 Consider the following result. For every natural number $n \geq 2$,

$$\left(1 - \frac{1}{4}\right)\left(1 - \frac{1}{9}\right)\left(1 - \frac{1}{16}\right)\cdots\left(1 - \frac{1}{n^2}\right) = \frac{n+1}{2n}$$

(a) If the statement is written in the form $\forall n \in \mathbb{N}_{\geq 2},\ P(n)$, what is the proposition $P(n)$?
(b) Π-notation is used for products in the same way as Σ-notation for sums: for example

$$\prod_{k=1}^{5}(k+1)^k = 2^1 \cdot 3^2 \cdot 4^3 \cdot 5^4 \cdot 6^5$$

Rewrite the statement using Π-notation.
(c) Prove the result by induction (you may use whatever notation you wish).

5.3.4 Show that for any $n \geq 3$, there is a set A consisting of n natural numbers such that the sum of the numbers in A is divisible by every element of A.

5.3.5 Recall the geometric series formula from calculus: if $r \neq 1$ is constant, and $n \in \mathbb{N}_0$, then

$$\sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r} \tag{$*$}$$

(a) Here is an incorrect proof by induction. Explain why it is incorrect.

Proof. Let $P(n) = \sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r}$.

(*Base Case* $n = 0$) $P(0) = \sum_{k=0}^{0} r^k = r^0 = 1 = \frac{1-r^{0+1}}{1-r}$ is true.

(*Induction Step*) Fix $n \in \mathbb{N}_0$ and assume that $P(n)$ is true. Then

$$\begin{aligned} P(n+1) &= \sum_{k=0}^{n+1} r^k = \sum_{k=0}^{n} r^k + r^{n+1} = \frac{1-r^{n+1}}{1-r} + r^{n+1} \\ &= \frac{1-r^{n+1}}{1-r} + \frac{r^{n+1}-r^{n+2}}{1-r} = \frac{1-r^{n+2}}{1-r}, \text{ is true.} \end{aligned}$$

By induction, $(*)$ is true for all $n \in \mathbb{N}_0$. ■

(b) Give a correct proof of $(*)$.

5.3.6 Here is an argument attempting to justify $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1) + 7$. What is wrong with it?

Proof. Assume that the statement is true for some fixed n. Then

$$\sum_{i=1}^{n+1} i = \sum_{i=1}^{n} i + (n+1) = \frac{1}{2}n(n+1) + 7 + (n+1) = \frac{1}{2}(n+1)[(n+1)+1] + 7,$$

hence the statement is true for $n+1$ and, by induction, for all $n \in \mathbb{N}$. ■

5.3.7 Let $P(n)$ and $Q(n)$ be propositions for each $n \in \mathbb{N}$.

(a) Assume that m is the smallest natural number such that $P(m)$ is false. Let

$$A = \{n \in \mathbb{N} : n < m\}.$$

What can you say about the elements in the set A, with respect to the property P?

(b) Assume that a is the smallest natural number such that $P(a) \vee Q(a)$ is false. Let

$$B = \{n \in \mathbb{N} : n < a\}.$$

What can you say about the elements in the set B, with respect to the properties P and Q?

(c) Assume that u is the smallest natural number such that $P(u) \wedge Q(u)$ is false. Let

$$C = \{n \in \mathbb{N} : n < u\}.$$

What can you say about the elements in the set C, with respect to the properties P and Q?

(d) Assume that $P(1)$ is true, but that '$\forall n \in \mathbb{N},\ P(n)$' is false. Show that there exists a natural number k such that the implication $P(k) \implies P(k+1)$ is false.

5.3.8 Prove that if $A \subseteq \mathbb{R}$ is a *finite* set, then A is well-ordered.

5.3.9 Show $\mathbb{Q}$ is not well-ordered.

5.3.10 In this question we use the fact that $\mathbb{N}_0$ is well-ordered to prove the Division Algorithm (Theorem 4.2).

Theorem: *If $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, then $\exists$ unique $q, r \in \mathbb{Z}$ such that $m = qn + r$ and $0 \leq r < n$.*

Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ be given, and define $S = \{k \in \mathbb{N}_0 : k = m - qn \text{ for some } q \in \mathbb{Z}\}$.

(a) Show that S is a *non-empty* subset of $\mathbb{N}_0$.

(b) $\mathbb{N}_0$ is well-ordered. By part (a), S has a minimal element r. Prove that $0 \leq r < n$.

(c) Suppose that there are two pairs of integers (q_1, r_1) and (q_2, r_2) which satisfy $m = q_i n + r_i$. Prove that $r_1 = r_2$ and, consequently, that the division algorithm is true.

5.3.11 In the text, we show that the principle of mathematical induction can be proved using the axiom that $\mathbb{N}$ is well-ordered. In fact we show in this exercise that one can go the other way. That is, one can take the principle of mathematical induction as an axiom, and derive that $\mathbb{N}$ is well-ordered.

(a) Assume that the principle of mathematical induction is true. We aim to show that $\mathbb{N}$ is well-ordered. Write out what this means.

(b) Explain why it is enough to show that if $A \subseteq \mathbb{N}$ has no minimum element, then $A = \emptyset$.

(c) Let $B = \{n \in \mathbb{N} : 1, 2, \ldots, n \notin S\}$. Use induction to show $B = \mathbb{N}$. Conclude that $A = \emptyset$.

5.3.12 We consider Peano's five axioms for the natural numbers:

Initial element: $1 \in \mathbb{N}$

Successor elements: There is a *successor function* $f : \mathbb{N} \to \mathbb{N}$. For each $n \in \mathbb{N}$, the successor $f(n)$ is also a natural number.

No predecessor of the initial element: $\forall n \in \mathbb{N}$, $f(n) = 1$ is false.

Unique predecessor: f is injective: $f(n) = f(m) \implies m = n$.

Induction: If $A \subseteq \mathbb{N}$ has the following properties:

- $1 \in A$,
- $\forall a \in A, f(a) \in A$,

then $A = \mathbb{N}$.

The successor function f is simply 'plus one' in disguise: $f(n) = n + 1$. Moreover, if you think carefully about the proof of Theorem 5.7, you should be convinced that the *induction* axiom is equivalent to the axiom that $\mathbb{N}$ is well-ordered, at least in the presence of the other four axioms.

(a) Suppose you replace $\mathbb{N}$ with $\mathbb{Z}$ in each of the above axioms. Which axioms are still true and which are false?

(b) Let (m,n) represent an ordered pair of natural numbers. Let T be the set of all pairs

$$T = \{(m,n) : m,n \in \mathbb{N}\}.$$

Let $f : T \to T$ be the function $f(m,n) = (m+1,n)$. Letting the pair $(1,1)$ play the role of '1' in Peano's axioms, and f be the successor function, decide which of the above axioms are satisfied by the set T.

(c) (Hard!) With the same set T as in part (b), take the successor function $f : T \to T$ to be

$$f(m,n) = \begin{cases} (m-1,n+1) & \text{if } m \geq 2, \\ (m+n,1) & \text{if } m = 1. \end{cases}$$

Which of the above axioms are satisfied by T and f?

5.3.13 (*Ignore this question if you haven't studied matrices*) Suppose that $A = \left(\begin{smallmatrix} 7 & 12 \\ -2 & -3 \end{smallmatrix}\right)$. We prove that

$$\forall n \in \mathbb{Z}, \quad A^n = \begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix} + 3^n \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}. \tag{†}$$

Here $A^{-n} = (A^n)^{-1}$ is the inverse of A^n, and we follow the convention that $A^0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is the identity matrix.

(a) Prove by induction that (†) holds $\forall n \in \mathbb{N}_0$.

(b) Modify your argument in part (a) to prove that (†) holds $\forall n \in \mathbb{Z}_0^-$. (*Use the fact that, when written in reverse order,* $\mathbb{Z}_0^- = \{0, -1, -2, -3, -4, \ldots\}$ *is a well-ordered set.*)

(c) Using what you know about matrix inverses, give a direct proof that (†) holds $\forall n \in \mathbb{Z}_0^-$. (*If C and D are 2×2 matrices such that* $CD = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, *then* $D = C^{-1}$.)

(d) Diagonalize the matrix A and thereby give a direct proof of (†) for all integers n.

5.3.14 (Hard!) You might assume from our earlier discussion that all well-ordered sets must look like the natural numbers. To disabuse you of this error, consider the set

$$B = \left\{0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots, 1, \frac{3}{2}, \frac{5}{3}, \frac{7}{4}, \frac{9}{5}, \ldots\right\} = \left\{\frac{n}{n+1} : n \in \mathbb{N}\right\} \cup \left\{\frac{2n-1}{n} : n \in \mathbb{N}\right\}$$

Prove that B is well-ordered.[16]

Hint: If $C \subseteq B$ is non-empty, consider the cases where $\exists c < 1$ and when all $c \geq 1$ separately.

5.3.15 (a) If $r \in \mathbb{Q}$ and $r \neq 0$ and $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, show $\alpha r \in \mathbb{R} \setminus \mathbb{Q}$.

(b) Use part (a), along with the density of $\mathbb{Q}$, to show that the irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ is also dense in $\mathbb{R}$.

5.3.16 Show that if $x \geq 0$ and $x < 1/n$ for all $n \in \mathbb{N}$, then $x = 0$.

---

[16] The principle of mathematical induction does not apply to propositions indexed by this set. The reason is that '1' is not a successor element in B: there is no element $b \in B$ such that 1 is 'the element after b.' Happily, there is a more general notion of *transfinite induction* which extends induction to propositions indexed by well-ordered sets like B. Transfinite induction proofs require an additional step in order to deal with *limit elements* like $1 \in B$.

5.4 Strong Induction

The principle of mathematical induction as stated in Theorem 5.7 is sometimes known as *weak* induction. In weak induction, we require only that one proposition $P(n)$ be true in order to demonstrate the truth of the succeeding proposition $P(n+1)$. By contrast, the induction step in *strong* induction additionally requires that more, perhaps *all*, of the propositions coming before $P(n)$ are also true.

Theorem 5.15 (Principle of Strong Induction)**.** *Let m be an integer and suppose that $P(n)$ is a proposition for each $n \in \mathbb{Z}_{\geq m}$. Also fix an integer $l \geq m$. Suppose:*

(a) $P(m), P(m+1), \ldots, P(l)$ *are true.*

(b) $\forall n \geq l,\ (P(m) \wedge P(m+1) \wedge \cdots \wedge P(n)) \implies P(n+1)$.

Then $P(n)$ is true for all $n \in \mathbb{Z}_{\geq m}$.

The statement is a little complicated: we show in the Exercises that it is equivalent to the earlier Principle of Mathematical Induction. What matters is that $\mathbb{Z}_{\geq m}$ is a well-ordered set. In the simplest examples, we have $m = 1$ and $\mathbb{Z}_{\geq 1} = \mathbb{N}$. The challenge in strong induction is identifying how much you need to assume in order to effect the induction step (b), and then how many base cases $l - m + 1$ are required.
It is much easier to learn strong induction by seeing it in action. Consider the Fibonacci numbers, an excellent source of strong induction examples.

Definition 5.16. The **Fibonacci numbers** are the sequence $(f_n)_{n=1}^{\infty} = (1, 1, 2, 3, 5, 8, 13, 21, \ldots)$ defined by the recurrence relation

$$\begin{cases} f_{n+1} = f_n + f_{n-1} & \text{if } n \geq 2 \\ f_1 = f_2 = 1 \end{cases}$$

Theorem 5.17. *For all natural numbers n we have $f_n < 2^n$.*

Proof. For each natural number n, let $P(n)$ be the proposition $f_n < 2^n$.
(*Base cases* $n = 1, 2$) $f_1 = 1 < 2^1$ and $f_2 = 1 < 2^2$, whence $P(1)$ and $P(2)$ are true.
(*Induction step*) Fix $n \geq 2$ and suppose that $P(1), \ldots, P(n)$ are true. Then

$$f_{n+1} = f_n + f_{n-1} < 2^n + 2^{n-1} < 2^n + 2^n = 2^{n+1}$$

which says that $P(n+1)$ is true.
By strong induction $P(n)$ is true for all $n \in \mathbb{N}$, and so $f_n < 2^n$. ∎

In terms of Theorem 5.15, we have $m = 1$ and $l = 2$ with $l - m + 1 = 2$ base cases. The reason we need $m = 1$ is because the first claim in the Theorem is about the integer 1, namely $f_1 < 2^1$. We need two base cases because the recurrence relation defining the Fibonacci numbers requires the previous *two* terms of the sequence in order to construct the next.

To help us understand strong induction, it is instructive to see why a proof by weak induction would fail in this setting.

Wrong Proof A. We show, by weak induction, that $\forall n \in \mathbb{N},\ f_n < 2^n$.
(*Base Case* $n = 1$) By definition, $f_1 = 1 < 2^1$, whence the claim is true for $n = 1$.
(*Induction Step*) Fix $n \in \mathbb{N}$ and assume that $f_n < 2^n$. We want to show that $f_{n+1} < 2^{n+1}$. By the recurrence relation, we can write

$$f_{n+1} = f_n + f_{n-1}. \tag{$*$}$$

The inductive hypothesis tells us that $f_n < 2^n$, but what can we say about f_{n-1}? Absolutely nothing! We are stuck: weak induction fails to prove the theorem. ∎

The incorrect proof tells us why we need strong induction: the recurrence relation defines each Fibonacci number (except f_1 and f_2) in terms of *the previous two.* To make use of the recurrence, our induction hypothesis must assume something about *at least* f_n *and* f_{n-1}. Assuming something about only f_n is insufficient.

From *Wrong Proof A* we learned that we needed to prove Theorem 5.17 by strong induction. Now suppose that we try the following, which looks almost identical to the correct proof.

Wrong Proof B. For each $n \in \mathbb{N}$, let $P(n)$ be the proposition $f_n < 2^n$. We prove that $P(n)$ is true for all $n \in \mathbb{N}$ by strong induction.
(*Base Case* $n = 1$) By definition, $f_1 = 1 < 2^1$, whence $P(1)$ is true.
(*Induction Step*) Fix $n \in \mathbb{N}$ and assume that $P(1), \ldots, P(n)$ are all true. We want to show that $f_{n+1} < 2^{n+1}$. By the recurrence relation, we can write

$$f_{n+1} = f_n + f_{n-1} < 2^n + 2^{n-1} < 2 \cdot 2^n = 2^{n+1}. \tag{†}$$

Hence $P(n)$ is true for all $n \geq 1$. ∎

Where is the problem with this second argument? The recursive formula $f_{n+1} = f_n + f_{n-1}$ *only* applies if $n \geq 2$. If we take $n = 1$, then it reads $f_2 = f_1 + f_0$, but f_0 is not defined! In the induction step of *Wrong Proof B,* we are letting n be any integer ≥ 1. When $n = 1$, step (†) is not justified, and so the proof fails. For (†) to be legitimate, we must have $n \geq 2$. This is why, in our correct proof, we had to prove $P(1)$ *and* $P(2)$ separately.

The moral here is to try the induction step as scratch work. Your attempt will tell you *if* you need strong induction and, if you do, *how many* base cases are required.

Strong Induction on Well-ordered Sets

In the next example the first term is suffixed by $n = 0$. In the language of Theorem 5.15, we have $m = 0$ and $l = 1$ with $l - m + 1 = 2$ base cases. Just like the Fibonacci example, two base cases are required because the defining recurrence relation constructs the next term in the sequence from the two previous terms.

Theorem 5.18. *A sequence of integers* $(a_n)_{n=0}^{\infty}$ *is defined by*

$$\begin{cases} a_n = 5a_{n-1} - 6a_{n-2}, & n \geq 2, \\ a_0 = 0,\ a_1 = 1. \end{cases}$$

Then $a_n = 3^n - 2^n$ *for all* $n \in \mathbb{N}_0$.

Proof. We prove by strong induction.
(*Base cases* $n = 0, 1$) The formula is true in both cases: $a_0 = 0 = 3^0 - 2^0$ and $a_1 = 1 = 3^1 - 2^1$.
(*Induction step*) Fix an integer $n \geq 1$ and suppose that $a_k = 3^k - 2^k$ for all $k \leq n$. Then

$$\begin{aligned} a_{n+1} &= 5a_n - 6a_{n-1} = 5(3^n - 2^n) - 6(3^{n-1} - 2^{n-1}) \\ &= (15-6)3^{n-1} + (10-6)2^{n-1} = 3^{n+1} - 2^{n+1}. \end{aligned}$$

By strong induction $a_n = 3^n - 2^n$ is true for all $n \in \mathbb{N}_0$. ∎

Think about why we wrote $a_{n+1} = 5a_n - 6a_{n-1}$ *in the induction step, whereas the statement in the Theorem reads* $a_n = 5a_{n-1} - 6a_{n-2}$. *Does it matter? What does it mean to say that* n *is a 'dummy variable'?*

In the two previous examples, it might seem that strong induction is something of a logical overkill. In the induction step we are assuming far more than we need. In both examples, establishing the truth of $P(n+1)$ required only the truth of $P(n)$ and $P(n-1)$. We assumed that the earlier propositions were also true, but we never used them. Depending on the proof, you might need two, three or even all of the propositions prior to $P(n+1)$ to complete the induction step. Once you are used to strong induction you may feel comfortable slimming a proof down so that you only mention precisely what you need. For the present, the way we've stated the principle is maximally safe! For some practice with this, see Exercise 5.4.2 where *three* base cases are needed, and the induction step requires the *three* previous propositions $P(n), P(n-1), P(n-2)$ in order to prove $P(n+1)$.

To see strong induction in all its glory, where the induction step requires *all* of the previous propositions, we prove part of the famous Fundamental Theorem of Arithmetic, which states that all natural numbers may be factored (uniquely) into a product of primes: for example $3564 = 2^2 \times 3^4 \times 11$. As you read the proof of the next theorem, think carefully about why *only one* base case is required.

Theorem 5.19. *Every natural number $n \geq 2$ is either prime, or a product of primes.*

First recall Definition 2.34, that $p \in \mathbb{N}_{\geq 2}$ is *prime* if its only positive divisors are itself and 1. Otherwise said, if $q \in \mathbb{N}_{\geq 2}$ is not prime, then it is said to be *composite:* $\exists a, b \in \mathbb{N}_{\geq 2}$ such that $q = ab$.

Proof. We prove by strong induction.

(*Base case* $n = 2$) The only positive divisors of 2 are itself and 1, hence 2 is prime.

(*Induction step*) Fix $n \in \mathbb{N}_{\geq 2}$ and assume that *every* natural number k satisfying $2 \leq k \leq n$ is either prime or a product of primes. There are two possibilities:

- $n + 1$ is prime. In this case we are done.
- $n + 1$ is composite. Thus $n + 1 = ab$ for some natural numbers $a, b \geq 2$. Clearly $a, b \leq n$, and so, by the induction hypothesis, *both* are prime or the product of primes. Therefore $n + 1$ is also the product of primes.

By strong induction we see that all natural numbers $n \geq 2$ are either prime, or a product of primes. ■

Reading Quiz

5.4.1 True or False: in a strong induction proof, we may have more than one base case.

5.4.2 What are some differences between strong induction and weak induction? Select all that apply.

(a) Strong induction has no induction step, but weak induction does.

(b) Both only have one base case.

(c) When proving $P(n + 1)$, strong induction allows one to assume all previous propositions are true, whereas weak induction only assumes $P(n)$ is true.

(d) Weak induction is equivalent to $\mathbb{N}$ being well-ordered, but strong induction is not equivalent.

5.4.3 True or False: there is a number which is not a product of primes.

Practice Problems

5.4.1 Let $(f_n)_{n=1}^{\infty}$ be the Fibonacci sequence. Prove that f_n is even if and only if $n \equiv 0 \pmod 3$.
Video Solution

5.4.2 Prove that a composite number a always has a prime factor p such that $p \leq \sqrt{a}$.
Video Solution

Exercises

5.4.1 Define a sequence $(b_n)_{n=1}^{\infty}$ as follows:

$$\begin{cases} b_n = b_{n-1} + b_{n-2}, & n \geq 3, \\ b_1 = 3,\ b_2 = 6. \end{cases}$$

Prove: $\forall\, n \in \mathbb{N},\ b_n$ is divisible by 3.

5.4.2 Define a sequence $(c_n)_{n=0}^{\infty}$ as follows:

$$\begin{cases} c_{n+1} = \frac{49}{8}c_n - \frac{225}{8}c_{n-2}, & n \geq 2, \\ c_0 = 0, c_1 = 2, c_2 = 16. \end{cases}$$

Prove that $c_n = 5^n - 3^n$ for all $n \in \mathbb{N}_0$. *Hint: you need three base cases!*

5.4.3 Prove that every $n \in \mathbb{N}$ can be written as

$$n = 2^{k_1} + 2^{k_2} + \cdots + 2^{k_\ell}$$

for some $\ell \in \mathbb{N}$ and $k_1, k_2, \ldots k_\ell \geq 0$ such that all of the k_i are distinct.

5.4.4 Consider the proof of Theorem 5.19.

(a) If the Theorem is written in the form $\forall n \in \mathbb{N}_{\geq 2}, P(n)$, what is the proposition $P(n)$?

(b) Explicitly carry out the induction step for the three situations $n+1 = 9$, $n+1 = 106$ and $n+1 = 45$. How many different ways can you perform the calculation for $n+1 = 45$? Explain why it is only necessary in the induction step to assume that all integers k satisfying $2 \leq k \leq \frac{n+1}{2}$ are prime or products of primes.

(c) Rewrite the proof in the style of Theorem 5.17, explicitly mentioning the propositions $P(n)$, and thus making the logical flow of strong induction absolutely clear.

5.4.5 In this question we use recall an alternative definition of prime.[17]

Definition. $p \in \mathbb{N}_{\geq 2}$ is *prime* if $\forall a, b \in \mathbb{N},\ p \,|\, ab \implies p \,|\, a$ or $p \,|\, b$.

Let p be prime, let $n \in \mathbb{N}$, and let $a_1, \ldots, a_n$ be natural numbers such that p divides the product $a_1 a_2 \cdots a_n$. Prove by induction that,

$$\exists i \in \{1, 2, \ldots, n\} \text{ such that } p \,|\, a_i.$$

Hint: you need to cover two *base cases. Why? Think about the induction step first and it will help you decide how many base cases you need.*

5.4.6 The *Fundamental Theorem of Arithmetic* states that every $n \geq 2$ can be written as a product of prime factors in a unique way (up to reordering of the prime factors). In other words,

(1) $n = p_1 p_2 \cdots p_k$ for some primes $p_1, p_2, \ldots, p_k$ and,

[17] This is the strict definition of what it means for p to be *prime,* while Definition 2.34 is what is meant by *irreducible.* In the ring of integers, *prime* and *irreducible* are synonymous. For the details, take a Number Theory course.

(2) if $n = q_1 q_2 \cdots q_\ell$ for primes $q_1, q_2, \ldots, q_\ell$, then $k = \ell$ and $p_i = q_i$ after possibly reordering the prime factors.

We proved (1) in Theorem 5.19. Supply a proof of (2). [Hint: one way would be to use Exercise 5.]

5.4.7 Prove that the nth Fibonacci number f_n is given by the formula

$$f_n = \frac{\phi^n - \hat{\phi}^n}{\sqrt{5}}, \quad \text{where} \quad \phi = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \hat{\phi} = \frac{1-\sqrt{5}}{2}.$$

ϕ is the famous Golden ratio. ϕ and $\hat{\phi}$ are the two solutions to the equation $x^2 = x + 1$.

5.4.8 Show that for every positive integer n, $(3+\sqrt{5})^n + (3-\sqrt{5})^n$ is an even integer.
Hints: Prove simultaneously that $(3+\sqrt{5})^n - (3-\sqrt{5})^n$ is an even multiple of $\sqrt{5}$.
Subtract the nth expression from the $(n+1)$th in both cases...

5.4.9 (Hard!) Return to the proof of Theorem 5.11. Can you make a watertight argument using strong induction that also covers the two missing cases? Draw a picture to illustrate each case.

5.4.10 Suppose that $\{P(n) : n \geq m\}$ are a collection of propositions as considered in the Principle of Strong Induction. For each $n \geq m$, let $Q(n)$ be the proposition

$$Q(n) \iff P(m) \wedge P(m+1) \wedge \cdots \wedge P(n)$$

Prove that the Principle of Strong Induction is equivalent to the Principle of Induction stated as follows: Suppose that

(a) $Q(l)$ is true.

(b) $\forall n \geq l,\ Q(n) \implies Q(n+1)$.

Then $Q(n)$ is true for all $n \in \mathbb{Z}_{\geq l}$.

6 Set Theory, Part II

In this chapter we return to set theory and consider several more-advanced constructions.

6.1 Cartesian Products

You have been working with Cartesian products for years, referring to a point in the plane $\mathbb{R}^2$ by its *Cartesian coordinates* (x, y). The basic idea is that each of the coordinates x and y is a member of the set $\mathbb{R}$. The same approach can be used for any two sets.

Definition 6.1. Let A and B be sets. The **Cartesian product** of A and B is the set

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

$A \times B$ is simply the set of *ordered pairs* (a, b) where $a \in A$ and $b \in B$. Two ordered pairs (a, b) and (c, d) are equal if and only if their coordinates agree: $a = c$ and $b = d$.

Examples. 6.1.1 The Cartesian product of the real line $\mathbb{R}$ with itself is the xy-plane: rather than writing $\mathbb{R} \times \mathbb{R}$ which is unwieldy, we write $\mathbb{R}^2$.

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) : x, y \in \mathbb{R}\}.$$

More generally, $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \mathbb{R}}_{n \text{ times}}$ is the set of n-tuples of real numbers:

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

6.1.2 If $A = \{1, 2, 3\}$ and $B = \{\alpha, \beta\}$, then the Cartesian product of A and B is

$$A \times B = \{(1, \alpha),\ (1, \beta),\ (2, \alpha),\ (2, \beta),\ (3, \alpha),\ (3, \beta)\}$$

Notice that this is a *different set* to the Cartesian product of B and A:

$$B \times A = \{(\alpha, 1),\ (\beta, 1),\ (\alpha, 2),\ (\beta, 2),\ (\alpha, 3),\ (\beta, 3)\}$$

6.1.3 Suppose you go to a restaurant where you have a choice of one main course and one side. The menu might be summarized set-theoretically: consider the sets

Mains = {fish, steak, eggplant, pasta}
Sides = {asparagus, salad, potatoes}

The Cartesian product Mains $\times$ Sides is the set of all possible meals made up of one main and one side. It should be obvious that there are $4 \times 3 = 12$ possible meal choices.

These last two examples illustrates the next theorem, which explains the use of the word *product.*

Theorem 6.2. *If A and B are finite sets, then $|A \times B| = |A| \cdot |B|$.*

Proof. Label the elements of each set and list the elements of $A \times B$ lexicographically. If $|A| = m$ and $|B| = n$, then we have:

$$A \times B = \left\{ \begin{array}{lllll} (a_1, b_1), & (a_1, b_2), & (a_1, b_3), & \cdots & (a_1, b_n), \\ (a_2, b_1), & (a_2, b_2), & (a_2, b_3), & \cdots & (a_2, b_n), \\ \vdots & \vdots & \vdots & & \vdots \\ (a_m, b_1), & (a_m, b_2), & (a_m, b_3), & \cdots & (a_m, b_n) \end{array} \right\}$$

It should be clear that every element of $A \times B$ is listed exactly once. There are m rows and n columns, thus $|A \times B| = mn$. ∎

Before we go any further, consider the complement of a Cartesian product $A \times B$. If you had to guess an expression for $(A \times B)^{\mathsf{C}}$, you might well try $A^{\mathsf{C}} \times B^{\mathsf{C}}$. Let us think more carefully.

$$\begin{aligned} (x, y) \in (A \times B)^{\mathsf{C}} &\iff (x, y) \notin A \times B \\ &\iff \neg((x, y) \in A \times B) \\ &\iff \neg(x \in A \text{ and } y \in B) \\ &\iff x \notin A \text{ or } y \notin B \end{aligned}$$

However $(x, y) \in A^{\mathsf{C}} \times B^{\mathsf{C}} \iff x \notin A$ and $x \notin B$. Since the definition of Cartesian product involves *and*, its negation, by De Morgan's laws, involves *or*. It follows that the complement of a Cartesian product is *not a Cartesian product!* For more on this, see Exercise 6.1.6.

As an example of a basic set relationship involving Cartesian products, we prove a theorem.

Theorem 6.3. *Let A, B, C, D be any sets. Then $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.*

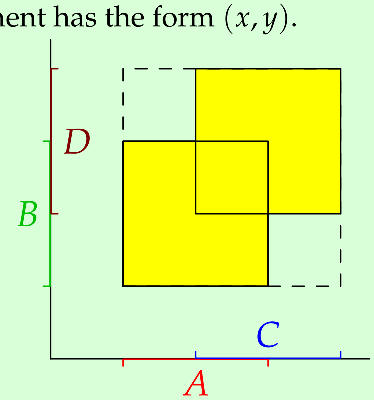
Proof. Since we are dealing with Cartesian products, the general element has the form (x, y).
Let $(x, y) \in (A \times B) \cup (C \times D)$. Then

$$(x, y) \in A \times B \qquad \text{or} \qquad (x, y) \in C \times D.$$

But then

$$(x \in A \text{ and } y \in B) \qquad \text{or} \qquad (x \in C \text{ and } y \in D).$$

Clearly $x \in A$ or $x \in C$, so $x \in A \cup C$.
Similarly $y \in B$ or $y \in D$, so $y \in B \cup D$.
Therefore $(x, y) \in (A \cup C) \times (B \cup D)$, as required. ∎



The picture is an visualization of the theorem, where we assume that the sets A, B, C and D are all intervals of real numbers. $(A \times B) \cup (C \times D)$ is the yellow shaded region, while $(A \cup C) \times (B \cup D)$ is the larger dashed square. While helpful, the picture is not a proof! The theorem is a statement about *any* sets, whereas the picture implicitly assumes that these sets are intervals.
For an application of the picture, it should be clear that if $x \in C \setminus A$ and $y \in B \setminus D$, then $(x, y) \in (A \cup C) \times (B \cup D)$ but $(x, y) \notin (A \times B) \cup (C \times D)$. We do not therefore expect these sets to be equal.

Reading Questions

6.1.1 Let A and B be sets. Let $(a, b), (c, d) \in A \times B$. Then $(a, b) = (c, d)$ if and only if

(a) $a = c$
(b) $b = d$
(c) $ad = bc$
(d) $a = c$ and $b = d$

6.1.2 True or False: $A \times B = \emptyset$ if and only if $A = B = \emptyset$.

6.1.3 Fill in the blank: If A and B are both finite nonempty sets, then $\max(|A|, |B|)$ ____ $|A \times B|$.

(a) $=$
(b) $\geq$
(c) $\leq$
(d) $\neq$

Practice Problems

6.1.1 Let A, B, C, D be sets. Prove

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D).$$

Video Solution

6.1.2 Let A and B be nonempty sets. Define a function $\pi_1 : A \times B \to A$ by $\pi_1(a, b) = a$. Show π_1 is surjective. Under what conditions is it a bijection?

Video Solution

Exercises

6.1.1 (a) Suppose that $A = \{1, 2\}$ and $B = \{3, 4, 5\}$. State the sets $A \times B$ and $B \times \emptyset$ in roster notation.

(b) Sketch both $A \times B$ and $B \times A$ using dots on the plane. What do you observe about your pictures?

(c) If A, B, C are any sets, we may define the triple Cartesian product as

$$A \times B \times C = \big\{(a, b, c) : a \in A,\ b \in B,\ c \in C\big\}$$

If $C = \{6, 7\}$ and A, B are as above, state the set $A \times B \times C$ in roster notation.

(d) For the sets A, B and C as above, is $A \times (B \times C) = A \times B \times C$?

6.1.2 Consider the following subintervals of the real line: $A = [2,5], \ B = (0,4)$.

(a) Express the set $(A \setminus B)^{\mathsf{C}}$ in interval notation, as a disjoint union of intervals.

(b) Sketch the sets $A \times B$ and $(A \times B)^{\mathsf{C}}$ on the plane $\mathbb{R}^2$. (Submit two different drawings, one for the set $A \times B$ and one for its complement.)

(c) Sketch the set $(A \setminus B)^{\mathsf{C}} \times (B \setminus A)$ on the plane $\mathbb{R}^2$.

6.1.3 Rewrite the condition

$$(x,y) \in (A^{\mathsf{C}} \cup B) \times (C \setminus D)$$

in terms of (some of) the following propositions:

$$x \in A, \quad x \notin A, \quad x \in B, \quad x \notin B, \quad y \in C, \quad y \notin C, \quad y \in D, \quad y \notin D.$$

6.1.4 Let $A = [1,3]$, $B = [2,4]$ and $C = [2,3]$. *Prove or disprove* that

$$(A \times B) \cap (B \times A) = C \times C.$$

Hint: Draw the sets $A \times B$, $B \times A$ and $C \times C$ in the Cartesian plane. The picture will give you a hint on whether or not the statement is true, but it does not constitute a proof.

6.1.5 A straight line subset of the plane $\mathbb{R}^2$ is a subset of the form

$$A_{a,b,c} = \{(x,y) : ax + by = c\}, \quad \text{for some constants } a, b, c, \text{ with } ab \neq 0.$$

(a) Draw the set $A_{1,2,3}$. Is it a Cartesian product?

(b) Which straight line subsets in the plane $\mathbb{R}^2$ are Cartesian products? Otherwise said, find a condition on the constants a, b, c for which the set $A_{a,b,c}$ is a Cartesian product.

6.1.6 Draw a picture, similar to that in Theorem 6.3, which illustrates the fact that

$$(A \times B)^{\mathsf{C}} \neq A^{\mathsf{C}} \times B^{\mathsf{C}}.$$

Using your picture, write the set $(A \times B)^{\mathsf{C}}$ in the form

$$(C_1 \times D_1) \cup (C_2 \times D_2) \cup \cdots$$

where each of the unions are *disjoint:* that is $i \neq j \implies (C_i \times D_i) \cap (C_j \times D_j) = \emptyset$. You don't have to prove your assertion.

6.1.7 Prove that $A \cap B = \emptyset \iff (A \times B) \cap (B \times A) = \emptyset$.

6.1.8 Let A, B, C be sets. Prove

(a) $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

(b) $A \times (B \cap C) = (A \times B) \cap (A \times C)$.

(c) $A \times (B \setminus C) = (A \times B) \setminus (A \times C)$.

6.1.9 (a) Give an explicit example of sets A, B, C, D such that $(A \times B) \cup (C \times D) \neq (A \cup C) \times (B \cup D)$.

(b) For sets A, B, C, D, prove that

$$(A \cup C) \times (B \cup D) = (A \times B) \cup (A \times D) \cup (C \times B) \cup (C \times D).$$

6.1.10 Let A and B be sets. Prove

$$(A \times B)^{\mathsf{C}} = (A^{\mathsf{C}} \times B^{\mathsf{C}}) \cup (A^{\mathsf{C}} \times B) \cup (A \times B^{\mathsf{C}}).$$

6.1.11 (a) Suppose that $|A| = 3$, and $|B| = 4$. What are the minimum and maximum values for the cardinalities $|(A \times B) \cap (B \times A)|$ and $|(A \times B) \cup (B \times A)|$?

(b) More generally, suppose that $|A| = m$, $|B| = n$ and $|A \cap B| = c$. What are the above cardinalities?

6.1.12 Prove the following by induction. For all $n \in \mathbb{N}$, if $A_1, \ldots, A_n$ are finite sets, then

$$|A_1 \times \cdots \times A_n| = |A_1| \cdots |A_n|$$

6.1.13 Let $E \subseteq \mathbb{N} \times \mathbb{N}$ be the smallest subset which satisfies the following conditions:

- Base case: $(1, 1) \in E$
- Generating Rule I: If $(a, b) \in E$ then $(a, a + b) \in E$
- Generating Rule II: If $(a, b) \in E$ then $(b, a) \in E$

(a) Show in detail that $(4, 3) \in E$.

(b) Show by induction that for every $n \in \mathbb{N}$, $(1, n) \in E$.

(c) (Very hard!!!) Show that $E = \{(a, b) \in \mathbb{N} \times \mathbb{N} : \gcd(a, b) = 1\}$. *Think carefully about how the Euclidean algorithm works, and what the generating rules might have to do with it...*

6.1.14 A strict set-theoretic definition requires you to build the ordered pair (a, b) as a set: typically $(a, b) = \{a, \{a, b\}\}$. One then proves that $(a, b) = (c, d) \iff a = c$ and $b = d$.

(a) One of the axioms of set theory (*regularity*) says that there is no set a for which $a \in a$. Use this to prove that the cardinality of $(a, b) = \{a, \{a, b\}\}$ is two.

(b) Prove that $(a, b) = (c, d) \implies \begin{cases} a = c \text{ and } b = d, \\ \qquad \textit{or} \\ a = \{c, d\} \text{ and } c = \{a, b\}. \end{cases}$

(c) In the second case, prove that there exists a set S such that $a \in S \in a$. The axiom of regularity also says that this is illegal. Conclude that $(a, b) = (c, d) \iff a = c$ and $b = d$.

6.1.15 Let A and B be nonempty sets. Define functions $\pi_1 : A \times B \to A$ and $\pi_2 : A \times B \to B$ by $\pi_1(a, b) = a$ and $\pi_2(a, b) = b$ respectively (these are called the *projection maps*).

(a) If $A = B = \mathbb{R}$ and $X = [1, 3]$, $Y = (2, 4]$, then $X \times Y \subseteq A \times B$. Compute the images $\pi_1(X \times Y)$ and $\pi_2(X \times Y)$.

(b) Let Z be any set and suppose there are functions $\rho_1 : Z \to A$ and $\rho_2 : Z \to B$. Show there is a unique function $h : Z \to A \times B$ such that $\rho_1 = \pi_1 \circ h$ and $\rho_2 = \pi_2 \circ h$.

6.2 Power Sets

Thusfar we have seen how to build new sets from old using the operations of subset, complement, union, intersection and Cartesian product. There is essentially only one further method whereby we can produce new sets; given a set A, we consider the collection of all of the subsets of A and we insist that this collection is a set.

Definition 6.4. The **power set** of A is the set $\mathcal{P}(A)$ of all subsets of A. That is,

$$\mathcal{P}(A) = \{B : B \subseteq A\}.$$

Otherwise said: $B \in \mathcal{P}(A) \iff B \subseteq A$.

Examples. 6.2.1 Let $A = \{1,3,7\}$. Then A has the following subsets, listed by how many elements are in each subset.

0-elements:	$\emptyset$
1-element:	$\{1\}$, $\{3\}$, $\{7\}$
2-elements:	$\{1,3\}$, $\{1,7\}$, $\{3,7\}$
3-elements:	$\{1,3,7\}$

Gathering these together, we have the power set:

$$\mathcal{P}(A) = \Big\{\emptyset, \{1\}, \{3\}, \{7\}, \{1,3\}, \{1,7\}, \{3,7\}, \{1,3,7\}\Big\}.$$

6.2.2 Consider $B = \Big\{1, \big\{\{2\},3\big\}\Big\}$. It is essential that you use different size set brackets to prevent confusion. B has only *two* elements, namely 1 and $\big\{\{2\},3\big\}$. We can gather the subsets of B in a table.

0-elements:	$\emptyset$
1-element:	$\{1\}$, $\Big\{\big\{\{2\},3\big\}\Big\}$
2-elements:	$\Big\{1, \big\{\{2\},3\big\}\Big\}$

In the second line, remember that to make a subset out of a single element you must surround the element with set brackets. Thus $1 \in B \implies \{1\} \subseteq B$ and

$$\big\{\{2\},3\big\} \in B \implies \Big\{\big\{\{2\},3\big\}\Big\} \subseteq B.$$

The power set of B is therefore

$$\mathcal{P}(B) = \bigg\{\emptyset,\ \{1\},\ \Big\{\big\{\{2\},3\big\}\Big\},\ \Big\{1, \big\{\{2\},3\big\}\Big\}\bigg\}.$$

Notation Be absolutely certain that you understand the difference between $\in$ and $\subseteq$. It is easy to become confused when considering power sets. In the context of the previous examples, here are eight propositions. Which are true and which are false?[18]

(a) $1 \in A$	(b) $1 \in \mathcal{P}(A)$	(c) $\{1\} \in A$	(d) $\{1\} \in \mathcal{P}(A)$
(e) $1 \subseteq A$	(f) $1 \subseteq \mathcal{P}(A)$	(g) $\{1\} \subseteq A$	(h) $\{1\} \subseteq \mathcal{P}(A)$

As a further exercise in being careful with notation, consider the following theorem.

Theorem 6.5. *If $A \subseteq B$, then $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.*

Proof. Suppose that $A \subseteq B$ and let $C \in \mathcal{P}(A)$. We must show that $C \in \mathcal{P}(B)$.
By definition, $C \in \mathcal{P}(A) \implies C \subseteq A$. Since subset inclusion is transitive (Theorem 3.5), we have

$$C \subseteq A \subseteq B \implies C \subseteq B.$$

This says that $C \in \mathcal{P}(B)$. Therefore $\mathcal{P}(A) \subseteq \mathcal{P}(B)$. ∎

It is very easy to get confused by the proof of this theorem. Exercises 6.2.4 and 6.2.5 discuss things further.

Cardinality and Power Sets

Let's investigate how the cardinality of a set and its power set are related. Consider a few basic examples where we list all of the subsets, grouped by cardinality.

Set A	0-elements	1-element	2-elements	3-elements	$\lvert\mathcal{P}(A)\rvert$
$\emptyset$	$\emptyset$				1
$\{a\}$	$\emptyset$	$\{a\}$			$1+1=2$
$\{a,b\}$	$\emptyset$	$\{a\},\{b\}$	$\{a,b\}$		$1+2+1=4$
$\{a,b,c\}$	$\emptyset$	$\{a\},\{b\},\{c\}$	$\{a,b\},\{a,c\},\{b,c\}$	$\{a,b,c\}$	$1+3+3+1=8$

You should have seen this pattern before: we are looking at the first few lines of Pascal's Triangle.[19] It should be no surprise that if $|A| = 4$, then $|\mathcal{P}(A)| = 1+4+6+4+1 = 16$. The progression $1, 2, 4, 8, 16, \ldots$ in the final column immediately suggests the following theorem.

Theorem 6.6. *Suppose that A is a finite set. Then $|\mathcal{P}(A)| = 2^{|A|}$.*

Conjuring up a proof may seem daunting given how little we know about A! In fact we have only one thing to work with: the *cardinality* of A. Indeed you might find it helpful to rephrase the theorem as follows:

$$\forall n \in \mathbb{N}_0, \ |A| = n \implies |\mathcal{P}(A)| = 2^n$$

[18] Only (a), (d), and (g) are true. Make sure you understand why!

[19] If you know a little about combinations from probability, it should be clear that a set A with n elements has precisely ${}^nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$ distinct r-element subsets.

Viewed this way, we see that we want to prove an infinite collection of propositions, indexed by the set $\mathbb{N}_0$: induction seems like the way forward. What might the induction step look like? The basic idea is that every set with $n+1$ elements is the disjoint union of a set with n elements and a single-element set. The induction step is essentially the observation that any $n+1$-element set B has *twice* the number of subsets of some n-element set A. It is instructive to see an example of this before writing the proof.

Example. Let $B = \{1,2,3\}$. Now choose the element $3 \in B$ and delete it to create the smaller set

$$A = \{1,2\} = B \setminus \{3\}.$$

We can split the subsets of B into two groups: those which contain 3 and those which do not. In the following table we list all of the subsets of B. In the first column are those subsets X which do not contain 3. These are exactly the subsets of A. In the second column are the subsets $Y = X \cup \{3\}$ of B which do contain 3.

X	$X \cup \{3\}$
$\emptyset$	$\{3\}$
$\{1\}$	$\{1,3\}$
$\{2\}$	$\{2,3\}$
$\{1,2\}$	$\{1,2,3\}$

It is clear that B has twice the number of subsets of A.

This method of pairing is exactly mirrored in the proof.

Proof. We prove by induction on the cardinality of A. For each $n \in \mathbb{N}_0$, we consider the proposition

$$|A| = n \implies |\mathcal{P}(A)| = 2^n. \qquad (*)$$

(*Base Case*) If $n = 0$, then $A = \emptyset$ (Theorem 3.5). But then $\mathcal{P}(A) = \{\emptyset\}$, whence $|\mathcal{P}(A)| = 1 = 2^0$.
(*Induction Step*) Fix $n \in \mathbb{N}_0$ and assume that $(*)$ is true for this n. That is, we assume that any set with n elements has 2^n subsets. Now let B be *any* set with $n+1$ elements. Choose one of the elements $b \in B$ and define $A = B \setminus \{b\}$. The subsets of B can then be separated into the following two types:

6.2.1 Subsets $X \subseteq B$ which do not contain b.

6.2.2 Subsets $Y \subseteq B$ which contain b.

In the first case, X is really a subset of A.
In the second case we can write $Y = X \cup \{b\}$, where X is again a subset of A.
Each subset $X \subseteq A$ therefore corresponds to precisely two subsets X and $X \cup \{b\}$ of B. Since $|A| = n$, the induction hypothesis tells us that there are 2^n subsets $X \subseteq A$, whence

$$|\mathcal{P}(B)| = 2\,|\mathcal{P}(A)| = 2^{n+1}.$$

By induction, $(*)$ is true for all $n \in \mathbb{N}_0$. ∎

Once you understand the proof, you should compare it to the proof of Theorem 5.11 on the interior angles of a polygon: the idea is very similar. Exercise 6.2.11 gives an alternative proof of this result.

As a final example, we consider the interaction of power sets and Cartesian products.

Example. Suppose that $A = \{a\}$ and $B = \{b, c\}$. Then

$$A \times B = \{(a, b), (a, c)\}.$$

The power set $\mathcal{P}(A \times B)$ therefore contains $2^2 = 4$ elements: indeed

$$\mathcal{P}(A \times B) = \Big\{\emptyset,\ \{(a,b)\},\ \{(a,c)\},\ \{(a,b),(a,c)\}\Big\}.$$

The power sets of A and B have 2 and 4 elements respectively:

$$\mathcal{P}(A) = \{\emptyset, \{a\}\}, \qquad \mathcal{P}(B) = \{\emptyset, \{b\}, \{c\}, \{b, c\}\}.$$

The Cartesian product of the power sets therefore has $2 \times 4 = 8$ elements:

$$\mathcal{P}(A) \times \mathcal{P}(B) = \Big\{(\emptyset,\emptyset),\ (\emptyset,\{b\}),\ (\emptyset,\{c\}),\ (\emptyset,\{b,c\}),$$
$$(\{a\},\emptyset),\ (\{a\},\{b\}),\ (\{a\},\{c\}),\ (\{a\},\{b,c\})\Big\}.$$

It should be clear from this example not only that $\mathcal{P}(A \times B) \neq \mathcal{P}(A) \times \mathcal{P}(B)$, but that the elements of the two sets are completely different. The elements of $\mathcal{P}(A \times B)$ are *sets of ordered pairs*, while the elements of $\mathcal{P}(A) \times \mathcal{P}(B)$ are *ordered pairs of sets*.

Reading Questions

6.2.1 Which of the following are true statements. Select all that apply.

(a) $[0, 1) \in \mathcal{P}(\mathbb{R})$
(b) $7 \in \mathcal{P}(\mathbb{N})$
(c) $\{(3, 5), (2, 9)\} \subseteq \mathcal{P}(\mathbb{N} \times \mathbb{N})$
(d) $\{4, \pi\} \in \mathcal{P}(\mathbb{R})$

6.2.2 Let $A = \{(1, 2), 3, (4, \{5\})\}$. What is $|\mathcal{P}(A)|$?

(a) 3
(b) 8
(c) 16
(d) 32

Practice Problems

6.2.1 Let $A = \{\emptyset, 1, \{a\}\}$. List the elements of $\mathcal{P}(A)$, compute its cardinality. Then answer True or False for the following:

(a) $\emptyset \in A$

(b) $\emptyset \subseteq A$

(c) $\emptyset \in \mathcal{P}(A)$

(d) $\emptyset \subseteq \mathcal{P}(A)$

(e) $\{\{a\}\} \subseteq \mathcal{P}(A)$

(f) $\{\{\emptyset, 1\}, \{\emptyset\}, \emptyset\} \subseteq \mathcal{P}(A)$

(g) $A \in \mathcal{P}(A)$

(h) $A \subseteq \mathcal{P}(A)$

Video Solution

6.2.2 Prove that $A \subseteq B$ if and only if $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

Video Solution

Exercises

6.2.1 Write the following sets in roster notation:

(a) $\mathcal{P}(A)$ for $A = \{1, 2\}$.

(b) $\mathcal{P}(A)$ for $A = \{1, 2, 3\}$.

(c) $\mathcal{P}(A)$ for $A = \{(1, 2), (2, 3)\}$.

(d) $\mathcal{P}(A)$ for $A = \{\emptyset, 3, \{4\}\}$.

(e) $\mathcal{P}(\mathcal{P}(A))$ for $A = \{3, 5\}$.

(f) $\{X \in \mathcal{P}(\{1, 2, 3, 4\}) : |X| = 1\}$.

6.2.2 Let $A = \{1, 3\}$ and $B = \{2, 4\}$.

(a) Draw a picture of the set $A \times B$.

(b) Compute $\mathcal{P}(A \times B)$.

(c) What is the cardinality of $\mathcal{P}(A) \times \mathcal{P}(B)$? *Don't compute the set!*

6.2.3 Determine whether the following statements are true or false (*in (b), the symbol $\subsetneq$ means 'proper subset'*). Justify your answers.

(a) If $\{7\} \in \mathcal{P}(A)$, then $7 \in A$ and $\{7\} \notin A$.

(b) Suppose that A, B and C are sets such that $A \subsetneq \mathcal{P}(B) \subsetneq C$ and $|A| = 2$. Then $|C|$ can be 5, but $|C|$ cannot be 4.

(c) If a set B has one more element than a set A, then $\mathcal{P}(B)$ has at least two more elements than $\mathcal{P}(A)$.

(d) Suppose that the sets A, B, C and D are all subsets of $\{1, 2, 3\}$ with cardinality two. Then at least two of these sets are equal.

6.2.4 Here are three incorrect proofs of Theorem 6.5. Explain why each fails.

(a) Let $x \in \mathcal{P}(A)$. Then $x \in A$. Since $A \subseteq B$, we have $x \in B$. Therefore $x \in \mathcal{P}(B)$, and so $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

(b) Let $A = \{1,2\}$ and $B = \{1,2,3\}$. Then $\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, A\}$, and $\mathcal{P}(B) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, B\}$. Thus $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

(c) Let $x \in A$. Since $A \subseteq B$, we have $x \in B$. Since $x \in A$ and $x \in B$, we have $\{x\} \in \mathcal{P}(A)$, and $\{x\} \in \mathcal{P}(B)$.

6.2.5 Consider the converse of Theorem 6.5. Is it true or false? Prove or disprove your conjecture.

6.2.6 (a) Prove that $\mathcal{P}(A) \cup \mathcal{P}(B) \subseteq \mathcal{P}(A \cup B)$. Provide a counter-example to show that we do not expect equality.

(b) Does anything change if you replace $\cup$ with $\cap$ in part (a)? Justify your answer.

6.2.7 Let A and B be sets. Prove or disprove: $A \subseteq B \implies \mathcal{P}(A) \subseteq \mathcal{P}(B)$.

6.2.8 (a) For any set A, show there is an injection $\iota : A \to \mathcal{P}(A)$. (Explicitly construct a map, and show that it is one-to-one.)

(b) Is there any set A such that $A \cap \mathcal{P}(A) \neq \emptyset$?

6.2.9 If we define an ordered pair (a,b) as $\{\{a\},\{a,b\}\}$, show that $A \times B \subseteq \mathcal{P}(\mathcal{P}(A \cup B))$.

6.2.10 Consider the proof of Theorem 6.6. Let B be a set with $n+1$ elements, let $b \in B$ and let $A = B \setminus \{b\}$. Prove that the function $f : \mathcal{P}(A) \times \{1,2\} \to \mathcal{P}(B)$ defined by

$$f(X,1) = X, \qquad f(X,2) = X \cup \{b\}$$

is a bijection, and that consequently, by Theorem 3.15, $|\mathcal{P}(A) \times \{1,2\}| = |\mathcal{P}(B)|$.

6.2.11 We use the following notation for the binomial coefficient: $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. This symbol denotes the number of distinct ways one can choose r objects from a set of n objects.

(a) Use the definition of the binomial coefficient to prove the following:

$$\text{If } 1 \le r \le n, \text{ then } \binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}.$$

(b) Prove by induction that $\forall n \in \mathbb{N}_0$, $\sum\limits_{r=0}^{n} \binom{n}{r} = 2^n$.

Hint: Use part (a) in the induction step. Note that the smallest n for which it applies is $n = 1 \ldots$

(c) Explain why part (b) provides an alternative proof of Theorem 6.6.

If you found this easy, try proving the binomial theorem: $\forall n \in \mathbb{N}$, $(x+y)^n = \sum\limits_{r=0}^{n} \binom{n}{r} x^r y^{n-r}$.

6.2.12 Let A and B be nonempty sets. We use the notation A^B to denote the set of all functions from B to A.

(a) If $A = \{0,1\}$ and $B = \{a,b,c\}$, list all elements of A^B. What is $|A^B|$?

(b) If A and B are finite sets, show $|A^B| = |A|^{|B|}$.

(c) Let B be a set and $Y \subseteq B$. Define $\chi_Y : B \to \{0,1\}$ by

$$\chi_Y(x) = \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{if } x \notin Y. \end{cases}$$

We call χ_Y the *characteristic function* of Y. By definition, $\chi_Y \in \{0,1\}^B$ for any $Y \subseteq B$. Show every element of $\{0,1\}^B$ is the characteristic function of some subset of B. In other words, prove that for all $f \in \{0,1\}^B$, there exists $Y \subseteq B$ such that $f = \chi_Y$.

(d) Let B be a set. Define $\Phi : \mathcal{P}(B) \to \{0,1\}^B$ by $\Phi(Y) = \chi_Y$. Show that Φ is a bijection.

(e) If B is finite, conclude that $|\mathcal{P}(B)| = |\{0,1\}^B| = 2^{|B|}$.

6.2.13 Let A, B, C, D be nonempty sets. Suppose that there is a bijection $f : A \to B$ and a bijection $g : C \to D$. Show there is a bijection between C^A and D^B.

6.2.14 Let X be an infinite set. A collection of sets $\mathcal{F} \subseteq \mathcal{P}(X)$ is called a *filter* if the following conditions are satisfied:

(1) $\emptyset \notin \mathcal{F}$ and $X \in \mathcal{F}$,

(2) if $A \subseteq B \subseteq X$ and $A \in \mathcal{F}$, then $B \in \mathcal{F}$,

(3) if $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$.

Filters are meant to capture a notion of largeness for sets.

(a) Show that $\{A : X \setminus A \text{ is finite}\}$ is a filter (this is called the *cofinite* or *Frechét filter*).

(b) A filter $\mathcal{U} \subseteq \mathcal{P}(X)$ is called an *ultrafilter* if it is a filter and for any $A \in \mathcal{P}(X)$, we have either $A \in \mathcal{U}$ or $X \setminus A \in \mathcal{U}$. Show that the cofinite filter is not an ultrafilter.

(c) Show that a filter $\mathcal{F}$ is an ultrafilter if and only if for any $A_1, \ldots, A_n \in \mathcal{P}(X)$ such that $A_1 \cup \cdots \cup A_n \in \mathcal{F}$, there is $1 \leq i \leq n$ such that $A_i \in \mathcal{F}$.

(d) Let $s \in X$, and define $\mathcal{U}_s = \{A \in \mathcal{P}(X) : s \in A\}$. Show $\mathcal{U}_s$ is an ultrafilter, called the *principal ultrafilter generated by s*.

(e) An ultrafilter $\mathcal{U}$ is *nonprincipal* if it is not equal to $\mathcal{U}_s$ for any $s \in X$. Show an ultrafilter $\mathcal{U}$ is nonprincipal if and only if it contains the cofinite filter (as a subset).

6.3 Indexed Collections of Sets

In this section we consider collections of sets A_n, where each n lies in some *indexing set* I. It is often the case that $I = \mathbb{N}$ or $\mathbb{Z}$. If I is some other set, for example the real numbers $\mathbb{R}$, the label for the index may be chosen accordingly: e.g. A_x.

Definition 6.7. Given a family of indexed sets $\{A_n : n \in I\}$, we may form the **union** and **intersection** of the collection:

$$\bigcup_{n \in I} A_n = \{x : x \in A_n \text{ for some } n \in I\},$$
$$\bigcap_{n \in I} A_n = \{x : x \in A_n \text{ for all } n \in I\}.$$

Otherwise said,

$$x \in \bigcup_{n \in I} A_n \iff \exists n \in I \text{ such that } x \in A_n$$
$$x \in \bigcap_{n \in I} A_n \iff \forall n \in I \text{ we have } x \in A_n$$

A indexed collection $\{A_n : n \in I\}$ is **pairwise disjoint** if $A_m \cap A_n = \emptyset$ whenever $m \neq n$.

When the indexing set is $\mathbb{N}$, it is common to use the notations $\bigcup\limits_{n=1}^{\infty} A_n$ and $\bigcap\limits_{n=1}^{\infty} A_n$.

Example. Let the indexing set be $I = \{\alpha, \beta, \gamma\}$, and let

$$A_\alpha = \{1, 3, 5\}, \qquad A_\beta = \{2, 3, 4, 6\}, \qquad A_\gamma = \{1, 2, 3, 6\}.$$

It should be clear that

$$\bigcup_{i \in I} A_i = A_\alpha \cup A_\beta \cup A_\gamma = \{1, 2, 3, 4, 5, 6\}$$

and

$$\bigcap_{i \in I} A_i = A_\alpha \cap A_\beta \cap A_\gamma = \{3\}$$

The following Theorem is almost immediate given the definitions of union and intersection: can you supply a formal proof?

Theorem 6.8. *Let $\{A_n : n \in I\}$ be an indexed collection of sets, and let $m \in I$. Then*

$$A_m \subseteq \bigcup_{n \in I} A_n \quad \text{and} \quad \bigcap_{n \in I} A_n \subseteq A_m.$$

Infinite Unions and Intersections: don't take limits!

The challenge with indexed sets often involves computing unions and intersections of *infinitely many* sets. Be very careful with this: it is very tempting to 'take limits' when this doesn't make sense. With this in mind, we dissect an important example.

For each $n \in \mathbb{N}$, consider the interval $A_n = \left[0, \frac{1}{n}\right)$. We analyze the collection $\{A_n : n \in \mathbb{N}\}$. First observe that $m \le n \implies \frac{1}{n} \le \frac{1}{m} \implies A_n \subseteq A_m$; the sets are therefore **nested**:

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq A_4 \supseteq \cdots \tag{$*$}$$

Since every set in the collection is a subset of A_1, it follows that this is the union,

$$\bigcup_{n=1}^{\infty} A_n = A_1 = [0, 1).$$

Before considering the full intersection, we first compute all finite intersections. Since the sets A_n are nested in the form $(*)$, it follows that any *finite* intersection is simply the smallest of the listed sets: i.e., for any constant $m \in \mathbb{N}$ we have

$$\bigcap_{n=1}^{m} A_n = A_m = \left[0, \frac{1}{m}\right).$$

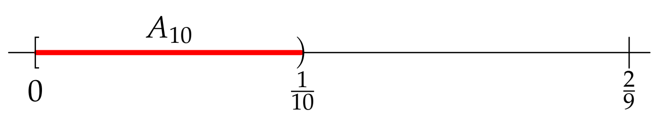
Observe that this is non-empty *for every m*. Now what about the infinite intersection? You might be tempted to take a limit and make an argument such as

$$\bigcap_{n=1}^{\infty} A_n \stackrel{?}{=} \lim_{m\to\infty} \bigcap_{n=1}^{m} A_n \stackrel{?}{=} \lim_{m\to\infty} \left[0, \frac{1}{m}\right) \stackrel{?}{=} \left[0, \lim_{m\to\infty} \frac{1}{m}\right) = [0, 0).$$

Quite apart from the issue that $[0, 0)$ is ugly and could only mean the empty set, we should worry about whether this is a legitimate use of limits. It isn't! We are only allows to take limits of sequences of numbers, not of *sets*. Perhaps you could forgive the abuse of limits if the approach yielded the correct conclusion. Unfortunately it doesn't: the infinite intersection is in fact non-empty, and we claim the following.

Theorem 6.9. $\bigcap\limits_{n=1}^{\infty} A_n = \{0\}$.

Before we give a formal proof, it is instructive to see a calculation. Let us show, for example, that $\frac{2}{9} \notin \bigcap\limits_{n=1}^{\infty} A_n$. To prove that $\frac{2}{9}$ is not in the intersection of *all* the A_n, it is enough to exhibit a single integer m such that $\frac{2}{9} \notin A_m$. The picture shows that we can choose $m = 10$: since $\frac{1}{10} < \frac{2}{9}$, we have $\frac{2}{9} \notin [0, \frac{1}{10}] = A_{10}$. Since $\frac{2}{9} \notin A_{10}$, we conclude that $\frac{2}{9} \notin \bigcap\limits_{n=1}^{\infty} A_n$.



Proof. We prove that $x \in \bigcap_{n=1}^{\infty} A_n \iff x = 0$.
Suppose that $x \in \bigcap_{n=1}^{\infty} A_n$. Then $x \in \left[0\,,\,\frac{1}{n}\right)$ for all n. Otherwise said,

$$\forall n \in \mathbb{N}, \text{ we have } 0 \leq x < \frac{1}{n}. \tag{†}$$

Certainly $x = 0$ satisfies these inequalities.
Now suppose, for a contradiction, that $x > 0$. Since $\lim_{n\to\infty} \frac{1}{n} = 0$, we can certainly choose[a] N large enough so that $\frac{1}{N} \leq x$. But this says that $x \notin A_N$, which contradicts (†).
The intersection contains no positive elements, and we conclude that

$$\bigcap_{n=1}^{\infty} A_n = \{0\}.$$

■

[a]Explicitly, you may choose choose $N = \lceil \frac{1}{x} \rceil$, or anything larger. Here $\lceil x \rceil$ is the *ceiling function:* the smallest integer greater than or equal to x.

By modifying the sets A_n to either include or exclude endpoints, we can obtain slightly different results. Consider each of the following in turn. How would the argument for computing each intersection differ from what we did above?

- If $B_n = \left(0\,,\,\frac{1}{n}\right)$, then $\bigcap_{n=1}^{\infty} B_n = \emptyset$.
- If $C_n = \left(0\,,\,\frac{1}{n}\right]$, then $\bigcap_{n=1}^{\infty} C_n = \emptyset$.
- If $D_n = \left[0\,,\,\frac{1}{n}\right]$, then $\bigcap_{n=1}^{\infty} D_n = \{0\}$.

The moral of these examples is that you cannot naïvely apply limits to sequences of sets. Your intuition is often a good guide, but that doesn't mean you should trust it blindly!

Here are a few more examples.

Examples. 6.3.1 Let $A_n = [n, n+1) \subseteq \mathbb{R}$, for each $n \in \mathbb{Z}$. For example,

$$A_3 = [3, 4), \quad \text{and} \quad A_{-17} = [-17, -16).$$

In this case the sets A_n are pairwise disjoint, and we have

$$\bigcup_{n\in\mathbb{Z}} A_n = \mathbb{R}, \quad \text{and} \quad \bigcap_{n\in\mathbb{Z}} A_n = \emptyset.$$

To prove the former, note that $\forall x \in \mathbb{R}$ we have $x \in [n, n+1)$ where $n = \lfloor x \rfloor$ is the greatest integer which is less than or equal to x: i.e. $x \in A_{\lfloor x \rfloor}$.

6.3.2 For each $n \in \mathbb{N}$, let $A_n = [-n, n]$. Each of the sets A_n is a closed interval. E.g.,

$$A_1 = [-1, 1], \qquad A_2 = [-2, 2], \qquad A_3 = [-3, 3].$$

It should be clear that $n \leq m \implies A_n \subseteq A_m$ so that we have a *nested* sequence of sets:

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$$

It follows immediately that the intersection is $\bigcap_{n \in \mathbb{N}} A_n = A_1 = [-1, 1]$.
With a little thinking you might hypothesize that the union is $\bigcup_{n \in \mathbb{N}} A_n = \mathbb{R}$. To prove this, assume that $x \in \mathbb{R}$ is non-zero, and observe that

$$-\lceil |x| \rceil \leq x \leq \lceil |x| \rceil \implies x \in A_{\lceil |x| \rceil}$$

Since $0 \in A_1$, it follows that $\mathbb{R} \subseteq \bigcup_{n \in \mathbb{N}} A_n$, whence these sets are equal.
If the notation is causing difficulty, consider for example,

$$-3.124 \in A_{\lceil 3.124 \rceil} = A_4.$$

6.3.3 For each $n \in \mathbb{N}$, let $A_n = \{x \in \mathbb{R} : \left|x^2 - 1\right| < \frac{1}{n}\}$. Before computing the union and intersection of these sets, it is helpful to write each set as a pair of intervals. Note that

$$\left|x^2 - 1\right| < \frac{1}{n} \iff -\frac{1}{n} < x^2 - 1 < \frac{1}{n} \iff \sqrt{1 - \frac{1}{n}} < |x| < \sqrt{1 + \frac{1}{n}}.$$

Therefore

$$A_n = \left(-\sqrt{1 + \tfrac{1}{n}}, -\sqrt{1 - \tfrac{1}{n}}\right) \cup \left(\sqrt{1 - \tfrac{1}{n}}, \sqrt{1 + \tfrac{1}{n}}\right).$$

As the picture suggests, the sets A_n are nested: $A_1 \supseteq A_2 \supseteq A_3 \supseteq A_4 \supseteq \cdots$.

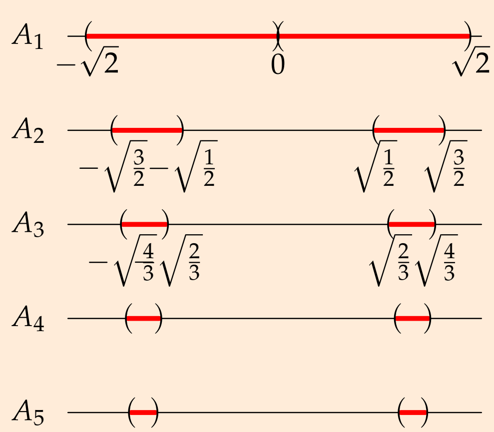
Since A_1 is the largest of the nested sets, we see that

$$\bigcup_{n \in \mathbb{N}} A_n = A_1 = (-\sqrt{2}, 0) \cup (0, \sqrt{2}).$$

For the intersection, note that

$$\begin{aligned} \forall n \in \mathbb{N},\ x \in A_n &\iff \forall n \in \mathbb{N},\ \left|x^2 - 1\right| < \frac{1}{n} \\ &\iff x^2 - 1 = 0. \end{aligned}$$

It follows that $\bigcap_{n \in \mathbb{N}} A_n = \{1, -1\}$.



Indexed Unions: Don't Confuse Sets and Elements

It is easy to confuse and important to distinguish between the sets

$$\{A_n : n \in I\} \qquad \text{and} \qquad \bigcup_{n \in I} A_n.$$

The first is a set whose *elements* are themselves sets. The second is the collection of all elements in *any* set A_n. Consider the following examples.

Examples. 6.3.1 For each $n \in \{1,2,3\}$, let A_n be the plane $\{(x,y,z) : x + ny + n^2 z = 1\} \subseteq \mathbb{R}^3$.

The indexed collection $\{A_1, A_2, A_3\}$ has *three* elements: each of the planes A_1, A_2, A_3 is an element in its own right.

The union $A_1 \cup A_2 \cup A_3$ is an *infinite* set consisting of all the *points* lying on any of the three planes.

For the intersection, a little work with simultaneous equations should convince you that

$$(x,y,z) \in \bigcap_{n \in \{1,2,3\}} A_n \iff \begin{cases} x + y + z = 1 \\ x + 2y + 4z = 1 \\ x + 3y + 9z = 1 \end{cases} \iff (x,y,z) = (1,0,0).$$

Thus $\bigcap A_n = \{(1,0,0)\}$. The planes are drawn below.

6.3.2 Let $I = \mathbb{R} \cup \{\infty\}$. For each $m \in I$, let A_m be the line[a] through the origin in $\mathbb{R}^2$ with gradient m.

Each element of $\{A_m : m \in I\}$ is a *line*: there is one for each direction through the origin.

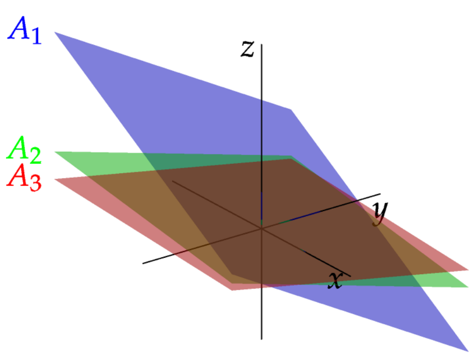
The union $\bigcup A_m$ consists of all of the *points* that lie on *any* line through the origin. Since any point in the plane lies on some line through the origin, we see that $\bigcup A_m = \mathbb{R}^2$.

It should be clear that all the lines intersect at the origin, and so $\bigcap A_m = \{(0,0)\}$.

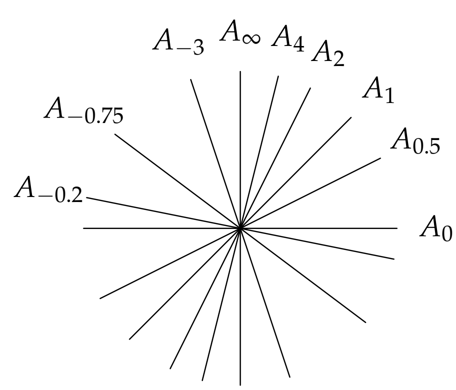
The collection of lines $\{A_m : m \in I\}$ is the famous *projective space* $\mathbb{P}(\mathbb{R}^2)$; this is a very different set from $\mathbb{R}^2$!

This example also shows that indexing sets don't have to be simple sets of integers. It is also possible to index the same set using $I = [0, \pi)$. If we define B_θ to be the line through the origin making an angle θ with the positive x-axis, we would then have $B_\theta = A_{\tan\theta}$.

[a] We include the vertical line A_∞.



Example 1: Three elements, or an infinite number?



Example 2: Elements in $\mathbb{P}(\mathbb{R}^2)$

Finite Decimals

Here is another example where our intuition of 'taking the limit' leads us astray. This time it is the union that behaves surprisingly.

For each $n \in \mathbb{N}$, let A_n be the set of decimals of length n. That is

$$A_n = \{0.a_1a_2\ldots a_n : \text{where each } a_i \in \{0,1,\ldots,9\}\}.$$

For example $0.134 \in A_3$. Since $0.134 = 0.1340$, we also have $0.134 \in A_4$. Once again we have a nested sequence of sets

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq A_4 \subseteq \cdots$$

The infinite intersection is therefore simply

$$\bigcap_{n\in\mathbb{N}} A_n = A_1 = \{0, 0.1, \ldots, 0.9\}.$$

Now consider a finite union: if $m \in \mathbb{N}$, then

$$\bigcup_{n=1}^{m} A_n = A_m = \{x \in [0,1) : x \text{ has a decimal representation of length} \leq m\}.$$

At this point, we might be inclined to take the limit as $m \to \infty$ of the *property* 'length m decimal.' If so, then it would seem that the infinite union should be the entire[20] interval $[0,1]$.
What is wrong with our reasoning? We have again abused the idea of limits: one cannot take the limit of a property! Instead we use the definition:

$$\begin{aligned} x \in \bigcup_{n\in\mathbb{N}} A_n &\iff \exists n \in \mathbb{N} \text{ such that } x \in A_n \\ &\iff \exists n \in \mathbb{N} \text{ such that } x \text{ is a decimal of length } n. \end{aligned}$$

It follows that

$$\bigcup_{n\in\mathbb{N}} A_n = \{x \in [0,1) : x \text{ has a \textit{finite} decimal representation}\}$$

In particular, there are no irrational numbers in $\bigcup_{n\in\mathbb{N}} A_n$:

If $x \in A_n$, then $y = 10^n x$ is an integer, whence $x = \dfrac{y}{10^n} \in \mathbb{Q}$.

Many rational numbers are also excluded. For example $\frac{1}{3} = 0.3333\cdots$ is not in any set A_n and is therefore not in the union.

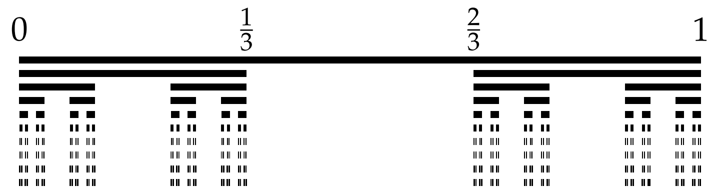
[20] We would include $1 = 0.9999\cdots$

The Cantor Set

We finish this section with a bit of fun. We can use infinite intersections to create self-similar sets, otherwise known as *fractals.* The *Cantor middle-third set* is a famous example.

Staring with the interval $C_0 = [0, 1]$, we construct a sequence of sets C_n for each $n \in \mathbb{N}_0$ by repeatedly removing the middle third of each of the intervals contained in C_n.

$C_0 = [0, 1]$,
$C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$,
$C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, etc.



The sequence is drawn up to C_9, with an animation below. To see the detail for the last few sets, try zooming in as far as you can.

Definition 6.10. The **Cantor set** $\mathcal{C}$ is the infinite intersection $\mathcal{C} = \bigcap\limits_{n=0}^{\infty} C_n$.

This set has several interesting properties.

Zero Measure (length) Intuitively, the *length* of a set of real numbers is the sum of the lengths of all the intervals contained in the set. Since we start with the interval $[0, 1]$ and remove a third of the set each time, it should be clear that

$$\text{length}(C_0) = 1, \qquad \text{length}(C_1) = \frac{2}{3}, \qquad \text{length}(C_2) = \left(\frac{2}{3}\right)^2, \quad \text{etc.}$$

Induction then gives us

$$\text{length}(C_n) = \left(\frac{2}{3}\right)^n.$$

As $n \to \infty$ this goes to zero, so the Cantor set contains no intervals. This at least seems reasonable from the picture.

Infinite Cardinality The Cantor set $\mathcal{C}$ contains the endpoints of every interval removed at any stage of its construction. In particular, $\frac{1}{3^n} \in \mathcal{C}$ for all $n \in \mathbb{N}_0$, and so $\mathcal{C}$ is an *infinite* set. Indeed it is more than merely infinite, it is *uncountably* so, as we shall see in Chapter 8.

Self-similarity If $\frac{1}{3}\mathcal{C}$ means 'take all the elements of $\mathcal{C}$ and divide them by three,' and $\frac{1}{3}\mathcal{C} + \frac{2}{3}$ means 'take all the elements of $\frac{1}{3}\mathcal{C}$ and add $\frac{2}{3}$,' then

$$\mathcal{C} = \frac{\mathcal{C}}{3} \cup \left(\frac{\mathcal{C}}{3} + \frac{2}{3}\right). \tag{$*$}$$

Otherwise said, $\mathcal{C}$ is made up of two shrunken copies of itself, a classic property of fractals. If you were to zoom into the Cantor set far enough that you couldn't see the whole set, you would not know what the scale was. In the following animation we are repeatedly zooming in on the second (of four) groups of points.

Optional: Analyzing the Cantor Set

To get further with the Cantor set, it is necessary to explicitly describe the elements of the set. This can be accomplished using the *ternary representation.* It can be shown that every number $x \in [0,1]$ may be written in the form[21]

$$x = \sum_{n=1}^{\infty} 3^{-n} a_n = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \cdots$$

where each $a_n \in \{0,1,2\}$. We write $x = [0.a_1 a_2 a_3 \cdots]_3$. For example:

$$[0.12]_3 = \frac{1}{3} + \frac{2}{3^2} = \frac{5}{9}, \qquad \frac{64}{243} = \frac{2}{3^2} + \frac{1}{3^3} + \frac{1}{3^5} = [0.02101]_3, \qquad 1 = [0.22222\cdots]_3.$$

For this last, use the formula for the sum of a geometric series to calculate $\sum\limits_{n=1}^{\infty} 2\left(\frac{1}{3}\right)^n = 2 \cdot \frac{1/3}{1-1/3} = 1.$

To convince yourself of the existence of a ternary representation, note that if $0 \le x < 1$ it follows that $x < 3$ and so, we can take

$$a_1 = \lfloor 3x \rfloor \in \{0,1,2\}$$

Now repeat, with $a_2 = \lfloor x - \frac{a_1}{3} \rfloor$, etc. It can also be shown that the only possibility whereby x can have two ternary expansions is if one of them terminates. The other will eventually become a sequence of repeating 2's. For example:[22]

$$[0.0222222\cdots]_3 = [0.1]_3 = \frac{1}{3} \quad \text{and} \quad [0.10122222\cdots]_3 = [0.102]_3 = \frac{1}{3} + \frac{2}{27} = \frac{11}{27}.$$

We can now describe precisely the elements of each of the sets C_n and consequently the Cantor set.

Theorem 6.11. *C_n is the set of all numbers $x \in [0,1]$ with a ternary expansion whose first n digits are only 0 or 2. It follows that $\mathcal{C}$ is the set of $x \in [0,1]$ with a ternary expansion containing only 0 and 2.*

The Theorem tells us that the Cantor set contains *a lot* of elements. For example:

$$[0.020202020\cdots]_3 = 2\sum_{n=1}^{\infty} 3^{-2n} = \frac{2/9}{1-1/9} = \frac{1}{4}$$

is an element of the Cantor set! What is strange is that $\frac{1}{4}$ is not the endpoint of any of the open intervals deleted during the construction of $\mathcal{C}$, and yet we've already established that $\mathcal{C}$ contains no intervals! Cantor introduced his set precisely because it was so challenging to the traditional concept of size: $\mathcal{C}$ seems to simultaneously have very few elements and enormously many.

[21] Analogous to a decimal representation $x = \sum\limits_{n=1}^{\infty} 10^{-n} a_n = \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \cdots$ where $a_n \in \{0,1,2,\ldots,9\}$.

[22] This is ticklish to prove, as is the corresponding result for decimals: compare with $1 = 0.99999999\cdots$

Proof. We prove by induction.
(*Base Case*) The proposition is clearly true for $C_0 = [0,1]$, as there is nothing to check.
(*Induction Step*) Assume that the proposition is true for some fixed $n \in \mathbb{N}_0$. Analogously to $(*)$ above, observe that C_{n+1} is built from two shrunken copies of C_n:

$$C_{n+1} = \frac{1}{3}C_n \cup \left(\frac{1}{3}C_n + \frac{2}{3}\right).$$

Now consider what division by 3 and addition of $\frac{2}{3}$ does to a ternary representation.

- Since $\frac{1}{3}\sum_{n=1}^{\infty} 3^{-n} a_n = \sum_{n=1}^{\infty} 3^{-n-1} a_1$, we see that multiplication by $\frac{1}{3}$ shifts a ternary representation one position to the right.[a]

$$\frac{1}{3}[0.a_1a_2a_3\ldots]_3 = [0.0a_1a_2a_3\ldots]_3$$

- Since $\frac{2}{3} = [0.2]_3$ we see that

$$\frac{2}{3} + \frac{1}{3}[0.a_1a_2a_3\ldots]_3 = [0.2a_1a_2a_3\ldots]_3$$

By the induction hypothesis, C_n contains only 0's and 2's in its first n entries. By moving ternary representations one step to the right and inserting 0 or 2 in the first position, we conclude that C_{n+1} contains only 0's and 2's in its first $n+1$ entries.
By induction the proposition is true for all $n \in \mathbb{N}_0$. ∎

[a] Compare to multiplication of a decimal by $\frac{1}{10}$.

Other fractal sets based on $\mathcal{C}$ include the Cantor dust $\mathcal{C} \times \mathcal{C}$, the Sierpiński carpet and gasket, and the von Koch snowflake.

Reading Quiz

6.3.1 Let I be a set and $\{A_n : n \in I\}$ a family of sets indexed by I. Then the definition of $\bigcup_{n\in I} A_n$ uses the __________ quantifier and the definition of $\bigcap_{n\in I} A_n$ uses the __________ quantifier.

(a) existential; existential
(b) existential; universal
(c) universal; existential
(d) universal; universal

6.3.2 Let I be a set and $\{A_n : n \in I\}$ a collection of sets indexed by I which is nested. What can you conclude? Select all that apply.

(a) $\bigcap_{n\in I} A_n \neq \emptyset$.
(b) $\bigcup_{n\in I} A_n = A_1$.

(c) The collection of sets is pairwise disjoint.

(d) Each A_n must be an interval.

6.3.3 True or False:

$$B \subseteq \bigcup_{n \in I} A_n \iff \forall n \in I,\ B \subseteq A_n$$

Practice Problems

6.3.1 (From previous exercises) For each non-negative real number $r \geq 0$ let

$$A_r = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = r^2\}$$

(a) Describe each of the sets A_r geometrically.

(b) Prove that $\bigcup_{r \in \mathbb{R}_0^+} A_r = \mathbb{R}^2$.

Video Solution

6.3.2 Let I be a set, $\{A_n : n \in I\}$ a family of sets indexed by I and B a set. Prove:

(a)

$$\left(\bigcup_{n \in I} A_n \right) \cap B = \bigcup_{n \in I} (A_n \cap B)$$

(b)

$$\left(\bigcap_{n \in I} A_n \right) \cup B = \bigcap_{n \in I} (A_n \cup B)$$

Video Solution

Exercises

6.3.1 For each integer n, consider the set $B_n = \{n\} \times \mathbb{R}$.

(a) Draw a picture of $\bigcup\limits_{n=2}^{4} B_n$ (in the Cartesian plane).

Hint: $\bigcup\limits_{n=2}^{4} B_n = B_2 \cup B_3 \cup B_4$.

(b) Draw a picture of the set $C = [1, 5] \times \{-2, 2\}$. *Careful!* $[1, 5]$ is an interval, while $\{-2, 2\}$ is a set containing two points.

(c) Compute $\left(\bigcup\limits_{n=2}^{4} B_n \right) \cap C$.

(d) Compute $\bigcup\limits_{n=2}^{4} (B_n \cap C)$.

(e) Compare $\left(\bigcup_{n=2}^{4} B_n\right) \cap C$ and $\bigcup_{n=2}^{4} (B_n \cap C)$. What do you notice?

6.3.2 (a) Determine $\bigcup_{r\in\{1,3,4\}} S_r$ and $\bigcap_{r\in\{1,3,4\}} S_r$, where S_r is the interval $[r-1, r+3]$.

(b) Determine $\bigcup_{i\in\mathbb{N}} \{i\}$ and $\bigcap_{i\in\mathbb{N}} \{i\}$.

(c) Determine $\bigcup_{X\in\mathcal{P}(\mathbb{Z})} X$ and $\bigcap_{X\in\mathcal{P}(\mathbb{Z})} X$.

6.3.3 Give an example of four different subsets A, B, C and D of $\{1,2,3,4\}$ such that all intersections of two subsets are different.

6.3.4 Find both the union and intersection of the following indexed collections of intervals. (Hint: Start by drawing a few sets in each collection.)

(a) $\{A_n\}_{n\in\mathbb{N}} = \{[0, 2+n] : n \in \mathbb{N}\}$

(b) $\{A_n\}_{n\in\mathbb{N}} = \{[1, 2+1), [1, 2+\frac{1}{2}), [1, 2+\frac{1}{3}), \ldots\}$

(c) $\{A_n\}_{n\in\mathbb{N}} = \{(\frac{-2n+1}{n}, 2n) : n \in \mathbb{N}\}$

(d) $\{A_n\}_{n\in\mathbb{N}} = \{(\frac{1}{4}, 1), (\frac{1}{8}, \frac{1}{2}), (\frac{1}{16}, \frac{1}{4}), (\frac{1}{32}, \frac{1}{8}), (\frac{1}{64}, \frac{1}{16}), \ldots\}$

6.3.5 For each non-negative real number $r \geq 0$ let

$$A_r = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = r^2\}$$

(a) Describe each of the sets A_r geometrically.

(b) Prove that $\bigcup_{r\in\mathbb{R}_0^+} A_r = \mathbb{R}^2$.

6.3.6 For each real number x, let $A_x = \{3, -2\} \cup \{y \in \mathbb{R} : y > x\}$. Find $\bigcup_{x\in\mathbb{R}} A_x$ and $\bigcap_{x\in\mathbb{R}} A_x$.

6.3.7 Use Definition 6.7 to prove the following results about nested sets.

(a) $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \implies \bigcup_{n\in\mathbb{N}} A_n = A_1$.

(b) $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \implies \bigcap_{n\in\mathbb{N}} A_n = A_1$.

6.3.8 Let $C_0(\mathbb{R})$ denote the set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ which satisfy $f(0) = 0$.
Let $A_f = \{x \in [0,1] : f(x) = 0\}$ (so, for example, if $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x(2x-1)$, then $A_f = \{0, \frac{1}{2}\}$). *Prove* that

$$\bigcup_{f\in C_0(\mathbb{R})} A_f = [0,1] \qquad \text{and} \qquad \bigcap_{f\in C_0(\mathbb{R})} A_f = \{0\}.$$

6.3.9 Let A_n be the set of decimals of length n, as described on page 172.

(a) Prove directly that the cardinality of A_n is 10^n.

(b) Prove by induction that $|A_n| = 10^n$.

(c) Prove that $\bigcup\limits_{n=1}^{\infty} A_n \subseteq \mathbb{Q}$.

(d) Prove by contradiction that $\frac{1}{3} \notin \bigcup\limits_{n=1}^{\infty} A_n$.

6.3.10 Suppose that the following are true:

- $\forall n \in \mathbb{N}, A_n \neq \emptyset$.
- $m \geq n \implies A_m \subseteq A_n$.

Prove or disprove the following conjectures:

(a) $\bigcup\limits_{n=1}^{293} A_n \neq \emptyset$

(b) $\bigcap\limits_{n=1}^{293} A_n \neq \emptyset$

(c) $\bigcup\limits_{n\in\mathbb{N}} A_n \neq \emptyset$

(d) $\bigcap\limits_{n\in\mathbb{N}} A_n \neq \emptyset$

6.3.11 Suppose we are working in a universal set $\mathcal{U}$ (so every set is considered a subset of $\mathcal{U}$). Give an explanation for why it makes sense to define $\bigcap_{n\in I} A_n = \mathcal{U}$ when $I = \emptyset$.

6.3.12 Let $\{A_n : n \in I\}$ and $\{B_n : n \in I\}$ be indexed families of sets. Give explicit examples for which the following hold:

(a)

$$\left(\bigcup_{n\in I} A_n\right) \cap \left(\bigcup_{n\in I} B_n\right) \neq \bigcup_{n\in I}(A_n \cap B_n)$$

(b)

$$\left(\bigcap_{n\in I} A_n\right) \cup \left(\bigcap_{n\in I} B_n\right) \neq \bigcap_{n\in I}(A_n \cup B_n)$$

6.3.13 (De Morgan's laws) Let $\{A_n : n \in I\}$ be an indexed family of sets and B a set. Prove

(a)

$$B \setminus \left(\bigcup_{n\in I} A_n\right) = \bigcap_{n\in I}(B \setminus A_n)$$

(b)

$$B \setminus \left(\bigcap_{n\in I} A_n\right) = \bigcup_{n\in I}(B \setminus A_n)$$

6.3.14 Let $\{A_n : n \in I\}$ be an indexed family of sets and B a set. Prove

(a)

$$\left(\bigcup_{n\in I} A_n\right) \setminus B = \bigcup_{n\in I}(A_n \setminus B)$$

(b)

$$\left(\bigcap_{n\in I} A_n\right) \setminus B = \bigcap_{n\in I}(A_n \setminus B)$$

6.3.15 We can take the Cartesian product of arbitrarily many sets. Let $\{A_n : n \in I\}$ be a family of sets. Define

$$\prod_{n\in I} A_n = \left\{ f : I \to \bigcup_{n\in I} A_n \mid f(n) \in A_n \right\}.$$

(a) If $I = \mathbb{N}$ and $A_n = \mathbb{R}$ for each $n \in \mathbb{N}$, can you give a more intuitive description of the elements of $\prod_{n\in\mathbb{N}} A_n$?

(b) Suppose we have two families $\{A_n : n \in I\}$ and $\{B_n : n \in I\}$. Prove

$$\left(\prod_{n\in I} A_n\right) \cap \left(\prod_{n\in I} B_n\right) = \prod_{n\in I}(A_n \cap B_n)$$

6.3.16 (Hard) Let $A_n = \{\frac{m}{n} \in \mathbb{Q} : 0 < m < n, m \in \mathbb{N}\}$, for each $n \in \mathbb{N}$.

(a) Write down A_1, A_2, A_3, A_4 explicitly.

(b) Prove that $A_m \subseteq A_{pm}$ for any $p \in \mathbb{N}$.

(c) Argue that $\bigcup\limits_{n\in\mathbb{N}} A_n = \mathbb{Q} \cap (0,1)$.

(d) Argue that further $\bigcup\limits_{n\in\mathbb{N}} A_{2n} = \mathbb{Q} \cap (0,1)$.

(e) Extend your proof to show that, for any fixed $p \in \mathbb{N}$, $\bigcup\limits_{n\in\mathbb{N}} A_{pn} = \mathbb{Q} \cap (0,1)$.

6.3.17 In this question we construct a fractal shape, similar to the von Koch curve. Let $F_0 = [0,1]$ be a straight line of length 1. Delete the segment between $\frac{1}{2}$ and $\frac{3}{4}$ to obtain the set

$$F_1 = [0, \tfrac{1}{2}] \cup [\tfrac{3}{4}, 1]$$

Now repeat: delete the third quarter of each of the two line segments in F_1 to obtain

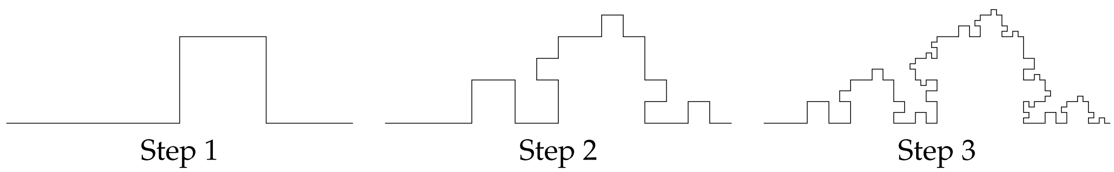
$$F_2 = [0, \tfrac{1}{4}] \cup [\tfrac{3}{8}, \tfrac{1}{2}] \cup [\tfrac{3}{4}, \tfrac{7}{8}] \cup [\tfrac{15}{16}, 1]$$

Suppose we repeat this process to create an infinite sequence of sets $F_0, F_1, F_2, F_3, F_4, \ldots$

(a) Prove that the total length of all of the line segments making up the set F_n is $\left(\frac{3}{4}\right)^n$.

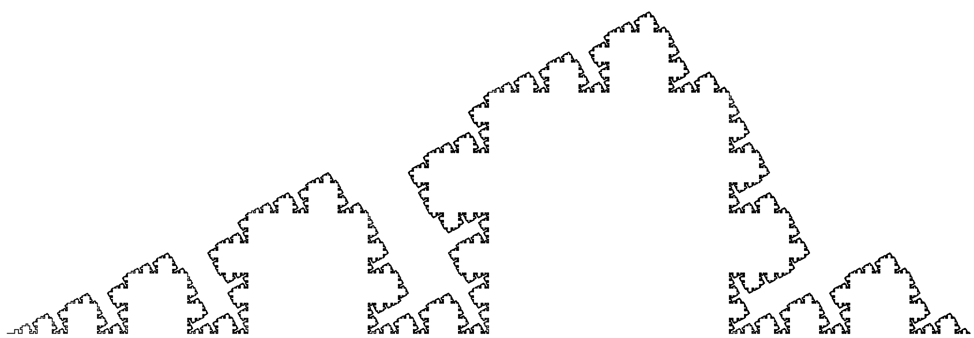
(b) Prove by contradiction that the intersection $\bigcap_{n=1}^{\infty} F_n$ does not contain any intervals of positive length.

(c) Now suppose that instead of simply deleting the third quarter of each line segment at each step, we replace it with the other three sides of a square. The first three steps in this process are shown below.



After each step, we are left with a curve. After step 1 the curve has length $\ell_1 = \frac{3}{2}$. After step 2 the length is $\ell_2 = \frac{9}{4}$. What is the *length* ℓ_n of the curve after n steps? Prove your assertion.

(d) Below is the result of repeating the steps in part 3 infinitely many times. What is the 'length' of the resulting fractal curve?



(e) Repeat parts (c) and (d) for the *area* under the curve at each step. Prove that the area between the fractal curve and the x-axis is $\frac{1}{8}$.

6.3.18 Let X be a set. A collection of sets $\tau \subseteq \mathcal{P}(X)$ is called a *topology* if the following conditions are satisfied:

(1) $\varnothing \in \tau$ and $X \in \tau$;

(2) τ is closed under arbitrary union. That is, if $\{U_n : n \in I\} \subseteq \tau$ for any index set I, then $\bigcup_{n \in I} U_n \in \tau$;

(3) τ is closed under finite intersection. That is, if $U_1, \ldots, U_n \in \tau$ for any $n \in \mathbb{N}$, then $U_1 \cap \cdots \cap U_n \in \tau$.

Elements of τ are called *open sets*.

(a) Let $X = \{a, b, c, d\}$. Let $\tau_1 = \{\varnothing, X\}$, $\tau_2 = \mathcal{P}(X)$, $\tau_3 = \{\varnothing, \{d\}, \{a, b\}, \{b, c\}, \{a, b, c\}, X\}$, and $\tau_4 = \{\varnothing, \{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}, X\}$. Show τ_1, τ_2 and τ_4 are topologies while τ_3 is not.

(b) Let X be an infinite set and define $\tau = \{A \in \mathcal{P}(X) : X \setminus A \text{ is finite}\} \cup \{\varnothing\}$. Show τ is a topology.

(c) The *standard topology* τ on $\mathbb{R}$ can be defined by declaring that a set $U \subseteq \mathbb{R}$ is open (i.e. an element of τ) if and only if for every $x \in U$, there is an open interval (a, b) such that $x \in (a, b)$ and $(a, b) \subseteq U$. Show this defines a topology on $\mathbb{R}$.

6.3.19 Let X be a set and τ a topology on X. A set $C \subseteq X$ is called *closed* if its complement is open, i.e., if $X \setminus C \in \tau$.

(a) Show the following properties of closed sets:

(i) $\emptyset$ and X are both closed.

(ii) If $\{A_n : n \in I\}$ is an arbitrary collection of closed sets, then $\bigcap_{n \in I} A_n$ is closed.

(iii) If $A_1, \ldots, A_n$ are closed sets, then $A_1 \cup \cdots \cup A_n$ is closed.

(b) In the standard topology on $\mathbb{R}$, show that a closed interval $[a, b]$ is a closed set but that a half-open interval $[a, b)$ is neither open nor closed.

7 Relations and Partitions

The mathematics of sets is rather basic, at least until one has a notion of how to relate elements of sets to each other. We are already familiar with examples of this:

7.0.1 The usual *order* of numbers (e.g. $3 < 7$) is a way of relating/comparing two elements of $\mathbb{R}$.

7.0.2 A *function* $f : A \to B$ relates elements in a set A with those in B.

It turns out that the concept of ordered pair (Cartesian product) is essential to relating elements.

7.1 Relations

Definition 7.1. Let A and B be sets. A **(binary) relation** $\mathcal{R}$ from A to B is a set of ordered pairs

$$\mathcal{R} \subseteq A \times B.$$

A **relation on** A is a relation from A to itself.
If $(x, y) \in \mathcal{R}$ we can also write $x \mathcal{R} y$, and say 'x is related to y.' Similarly $x \not\mathcal{R} y$ means $(x, y) \notin \mathcal{R}$.

Examples. 7.1.1 $\mathcal{R} = \{(1,3), (2,2), (2,3), (3,2), (4,1), (5,2)\}$ is a relation from $\mathbb{N}$ to $\mathbb{N}$. It is also a relation from $\{1,2,3,4,5\}$ to $\{1,2,3\}$. Various true statements about this relation include

$$(2,2) \in \mathcal{R}, \qquad (4,2) \notin \mathcal{R}, \qquad 2 \not\mathcal{R} 5, \qquad 3 \mathcal{R} 2$$

7.1.2 $\mathcal{R} = \Big([1,3) \times (3,4]\Big) \cup \left\{(2t+1, t^2) : t \in [\frac{1}{2}, 2]\right\}$ is a relation from $\mathbb{R}$ to $\mathbb{R}$. Be careful: it is easy to confuse interval notation with the notation for ordered pair!

7.1.3 The set $\mathcal{R} = \{(a,a) : a \in A\}$ is a relation on A, indeed

$$(x, y) \in \mathcal{R} \iff x = y$$

defines a relation on *any* set A. This example is where the term *equivalence relation* (Section 7.3) comes from. $x \mathcal{R} y \iff x = y$ simply says that $\mathcal{R}$ is 'equals.'

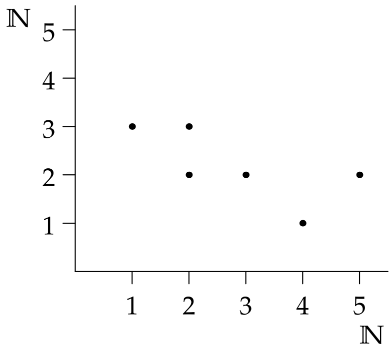
7.1.4 If $A = \{\text{all humans}\}$, we may define $\mathcal{R} \subseteq A \times A$ by

$$(a_1, a_2) \in \mathcal{R} \iff a_1, a_2 \text{ have a parent-child, or a sibling relationship.}$$

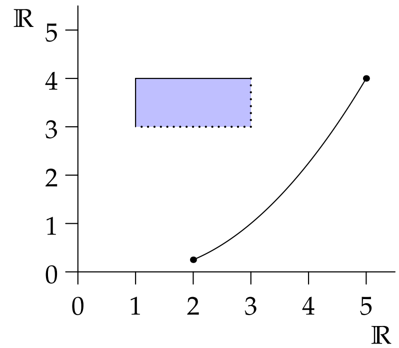
In this example, the mathematical use of the word relation is identical to that in English. For example, I am related to my sister, and my mother is related to me.

7.1.5 If A is a set, then $\subseteq$ is a relation on the power set $\mathcal{P}(A)$.
For example, if $A = \{1,2,3\}$ then $\{1\} \in \mathcal{P}(A)$ and $\{1,3\} \in \mathcal{P}(A)$. We'd say that $\{1\}$ is related to $\{1,3\}$ since $\{1\} \subseteq \{1,3\}$.
It should be clear that, under the relation $\subseteq$, that $\{1,3\}$ is not related to $\{1\}$.

When $\mathcal{R}$ is a relation between sets of numbers, we can often *graph* the relation. Examples 1 and 2 above would be graphed as follows:



Example 1.



Example 2.

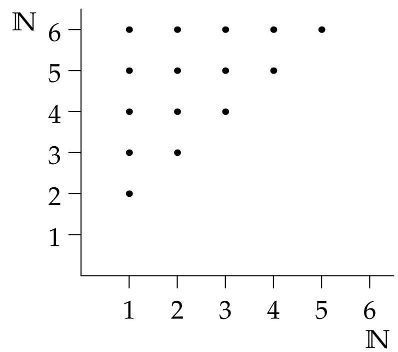
Not all relations between sets of numbers can be graphed: for example, graphing the relation $\mathcal{R} = \mathbb{Q} \times \mathbb{Q}$ is impossible!

To refer to the introduction, the standard ordering $<$ on $\mathbb{N}$ is a relation, and we can graph it: for all $x, y \in \mathbb{N}$, we define

$$x \, \mathcal{R} \, y \iff x < y$$

or equivalently,

$$\mathcal{R} = \{(x,y) \in \mathbb{N} \times \mathbb{N} : x < y\}$$



We can also think about functions in this language: if $f : \mathbb{R} \to \mathbb{R}$ is a function, then we could define

$$x \, \mathcal{R} \, y \iff y = f(x)$$

or equivalently

$$\mathcal{R} = \{(x,y) \in \mathbb{R}^2 : y = f(x)\}$$

We will return to this viewpoint on function in the Section 7.2.

Basic results regarding relations

With abstract relations, there are only a small number of things we can do.

Definition 7.2. If $\mathcal{R} \subseteq A \times B$ is a relation, then its **inverse** $\mathcal{R}^{-1} \subseteq B \times A$ is the set

$$\mathcal{R}^{-1} = \{(y,x) \in B \times A : (x,y) \in \mathcal{R}\}.$$

To find the elements of $\mathcal{R}^{-1}$, you simply switch the components of each ordered pair in $\mathcal{R}$.
Suppose $A = B$. We say that $\mathcal{R}$ is **symmetric** if $\mathcal{R} = \mathcal{R}^{-1}$.

The following results should seem natural, even if some of the proofs may not be obvious.

Theorem 7.3. *Given any relations $\mathcal{R}, \mathcal{S} \subseteq A \times B$:*

7.1.1 $(\mathcal{R}^{-1})^{-1} = \mathcal{R}$

7.1.2 $\mathcal{R} \subseteq \mathcal{S} \iff \mathcal{R}^{-1} \subseteq \mathcal{S}^{-1}$

7.1.3 $(\mathcal{R} \cup \mathcal{S})^{-1} = \mathcal{R}^{-1} \cup \mathcal{S}^{-1}$

7.1.4 $(\mathcal{R} \cap \mathcal{S})^{-1} = \mathcal{R}^{-1} \cap \mathcal{S}^{-1}$

7.1.5 If $A = B$, then $\mathcal{R} \cup \mathcal{R}^{-1}$ is symmetric

7.1.6 If $A = B$, then $\mathcal{R} \cap \mathcal{R}^{-1}$ is symmetric

Proof. Here are two of the arguments. Try the others yourself.

2. Assume that $\mathcal{R} \subseteq \mathcal{S}$, and suppose that $(x, y) \in \mathcal{R}^{-1}$. We must prove that $(x, y) \in \mathcal{S}^{-1}$. By the definition of inverse,

$$\begin{aligned} (x,y) \in \mathcal{R}^{-1} &\implies (y,x) \in \mathcal{R} \implies (y,x) \in \mathcal{S} \\ &\implies (x,y) \in \mathcal{S}^{-1}. \end{aligned}$$

   Therefore $\mathcal{R}^{-1} \subseteq \mathcal{S}^{-1}$. For the converse, suppose that $\mathcal{R}^{-1} \subseteq \mathcal{S}^{-1}$. Then, by an argument similar to the above, we see that $(\mathcal{R}^{-1})^{-1} \subseteq (\mathcal{S}^{-1})^{-1}$. Now use 1. to see that

$$\mathcal{R}^{-1} \subseteq \mathcal{S}^{-1} \implies \mathcal{R} \subseteq \mathcal{S}.$$

5. By 3,

$$(\mathcal{R} \cup \mathcal{R}^{-1})^{-1} = \mathcal{R}^{-1} \cup (\mathcal{R}^{-1})^{-1} = \mathcal{R}^{-1} \cup \mathcal{R} = \mathcal{R} \cup \mathcal{R}^{-1},$$

   and so $\mathcal{R} \cup \mathcal{R}^{-1}$ is symmetric. ■

Keep your proof skills sharp! Several parts of Theorem 7.3 look suspiciously similar to earlier results and it is easy to get confused. For example, 3 and 4 look almost like De Morgan's laws, except that $\cup$ and $\cap$ do not switch over. This is why it is important to be able to conjure up examples and *prove* such statements. There are many facts in mathematics: trying to memorize everything is too difficult! Instead, you will be forever conjecturing and having to justify your guesses. For example, suppose that you forget results 3 and 4: it seems reasonable to conjecture that

$$(\mathcal{R} \cup \mathcal{S})^{-1} = \begin{cases} \mathcal{R}^{-1} \cup \mathcal{S}^{-1} \\ \quad\text{or} \\ \mathcal{R}^{-1} \cap \mathcal{S}^{-1} \end{cases}$$

Now that you have two sensible guesses, you should be able to decide the correct one by thinking about examples and, if necessary, proving your assertion!

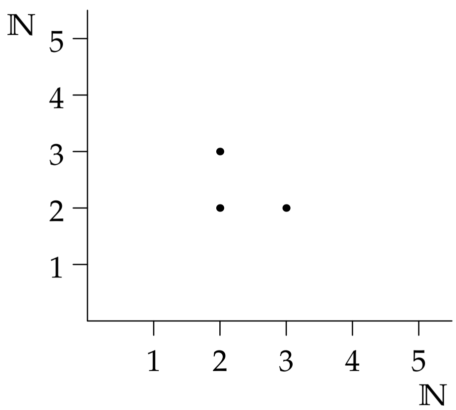
Example. Consider Example 1 from before: $\mathcal{R} = \{(1,3),(2,2),(2,3),(3,2),(4,1),(5,2)\} \subseteq \mathbb{N} \times \mathbb{N}$. This is not symmetric since, for example, $1\,\mathcal{R}\,3$ but $3 \not{\mathcal{R}}\, 1$. We compute

$$\mathcal{R}^{-1} = \{(3,1),(2,2),(3,2),(2,3),(1,4),(2,5)\},$$

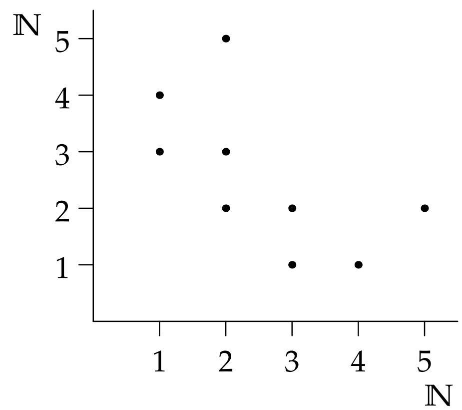
and observe that

$$\mathcal{R} \cap \mathcal{R}^{-1} = \{(2,2),(2,3),(3,2)\} \quad \text{and}$$
$$\mathcal{R} \cup \mathcal{R}^{-1} = \{(1,3),(3,1),(2,2),(2,3),(3,2),(4,1),(1,4),(5,2),(2,5)\}$$

are both symmetric.



The relation $\mathcal{R} \cap \mathcal{R}^{-1}$



The relation $\mathcal{R} \cup \mathcal{R}^{-1}$

These pictures should confirm something intuitive: if you are able to graph a symmetric relation, then the graph will have symmetry about the line $y = x$. Indeed, $\mathcal{R}^{-1}$ is obtained by reflecting $\mathcal{R}$ in the line $y = x$. Recall how to graph an inverse functions from calculus. . .

Reading Questions

7.1.1 A relation $\mathcal{R} \subseteq A \times B$ is ________________

(a) a nonempty subset of $A \times B$
(b) a proper subset of $A \times B$
(c) a function from A to B
(d) a subset of $A \times B$

7.1.2 If $A \subseteq \mathbb{R}$, then the graph of a symmetric relation $\mathcal{R} \subseteq A \times A$ has what kind of symmetry?

(a) symmetric about the x-axis
(b) symmetric about the y-axis
(c) symmetric about the line $y = x$
(d) symmetric across the origin

7.1.3 True or False: if $\mathcal{R}$ is symmetric, then it must contain an even number of elements.

Practice Problems

7.1.1 Let $L_{a,b,c} = \{(x,y) : ax + by = c\} \subseteq \mathbb{R}^2$.

(a) Describe $L_{a,b,c}$ geometrically.

(b) Let $A = \mathbb{R}^2$ and $B = \{L_{a,b,c} : a,b,c \in \mathbb{R}\}$. Define $\mathcal{R} \subseteq A \times B$ by $(x,y)\,\mathcal{R}\,L_{a,b,c}$ if and only if $ax + by = c$. For each of the following, determine if it is true or false.

(i) $(1,0)\,\mathcal{R}\,L_{1,1,1}$

(ii) $(3,-2)\,\not\mathcal{R}\,L_{1,1,1}$

(iii) if $(x,y)\,\mathcal{R}\,L_{a,b,c}$ and $(x,y)\,\mathcal{R}\,L_{d,e,f}$ for some (x,y) then $L_{a,b,c} = L_{d,e,f}$

(iv) suppose $(x,y)\,\not\mathcal{R}\,L_{a,b,c}$, then there exists $d,e,f \in \mathbb{R}$ such that $(x,y)\,\mathcal{R}\,L_{d,e,f}$ and $L_{a,b,c} \cap L_{d,e,f} = \emptyset$

7.1.2 Let X be a set. Let $\mathcal{R} \subseteq \mathcal{P}(X) \times \mathcal{P}(X)$ be the relation $A\,\mathcal{R}\,B \iff A \subseteq B$.

(a) Show that $A\,(\mathcal{R} \cap \mathcal{R}^{-1})\,B$ implies $A = B$.

(b) If $X = \{a,b\}$, compute $\mathcal{R}^{-1}$ explicitly as a set of ordered pairs.

Exercises

7.1.1 Let $\mathcal{R}$ be the relation on $\{0,1,2\}$ defined by

$$0\,\mathcal{R}\,0 \qquad 0\,\mathcal{R}\,1 \qquad 2\,\mathcal{R}\,1$$

(a) Write $\mathcal{R}$ as a set of ordered pairs.

(b) What is the inverse of $\mathcal{R}$?

7.1.2 (a) Let $\mathcal{R}$ be the relation on $\mathbb{R}$ defined by $a\,\mathcal{R}\,b \iff |a-b| = 1$. Is this relation symmetric?

(b) Let $\sim$ be the relation on $\mathbb{R}$ defined by

$$a \sim b \Leftrightarrow \exists x \in \mathbb{Q} \setminus \{0\} \text{ such that } a = x^2 b.$$

Is this relation symmetric?

7.1.3 Draw pictures of the following relations on the set of real numbers $\mathbb{R}$.

(a) $\mathcal{R} = \{(x,y) : y \leq x \text{ and } y \leq 2 \text{ and } y \leq 2 - x\}$.

(b) $\mathcal{S} = \{(x,y) : (x-4)^2 + (y-1)^2 \leq 9\}$.

Also draw the inverse of each relation.

7.1.4 A relation is defined on $\mathbb{N}$ by $a\,\mathcal{R}\,b \iff \frac{a}{b} \in \mathbb{N}$. Let $c,d \in \mathbb{N}$. Under what conditions is it permissable to write $c\,\mathcal{R}^{-1}\,d$?

7.1.5 Let $\mathcal{R} \subseteq \mathbb{N}^2$ be the relation $m\,\mathcal{R}\,n$ iff $m \mid n$. Compute $\mathcal{R} \cap \mathcal{R}^{-1}$.

7.1.6 Let $\mathcal{R} \subseteq \{1,2,3,4\} \times \{1,2,3,4\}$ be the relation

$$\mathcal{R} = \{(1,3),(1,4),(2,2),(2,4),(3,1),(3,2),(4,4)\}.$$

(a) Compute $\mathcal{R}^{-1}$.

(b) Compute the relations $\mathcal{R} \cup \mathcal{R}^{-1}$ and $\mathcal{R} \cap \mathcal{R}^{-1}$, and check that they are symmetric.

7.1.7 For the relation $\mathcal{R} = \{(x, y) : x \leq y\}$ defined on $\mathbb{N}$, what is $\mathcal{R}^{-1}$, and what is the intersection $\mathcal{R} \cap \mathcal{R}^{-1}$?

7.1.8 Let A be a set with $|A| = 4$. What is the maximum number of elements that a relation $\mathcal{R}$ on A can contain such that $\mathcal{R} \cap \mathcal{R}^{-1} = \emptyset$?

7.1.9 Give formal proofs of the remaining cases (1, 3, 4 & 6) of Theorem 7.3.

7.1.10 Let $\mathcal{R}$ and $\mathcal{S}$ be two symmetric relations on a set A.

(a) Show $\mathcal{R} \cap \mathcal{S}$ is symmetric.

(b) Does $\mathcal{R} \cup \mathcal{S}$ have to be symmetric? Give a proof or counterexample.

7.1.11 Let $\mathcal{R}$ be a relation on a set A and define $\mathcal{S} = \mathcal{R} \cup \mathcal{R}^{-1}$. We know that $\mathcal{S}$ is symmetric. Prove that $\mathcal{S}$ is the intersection of all *symmetric* relations on A which contain $\mathcal{R}$. Otherwise said: if

$$\mathrm{T} = \left\{ \mathcal{T} \subseteq A \times A : \mathcal{T} \text{ symmetric and } \mathcal{R} \subseteq \mathcal{T} \right\}$$

then

$$\mathcal{S} = \bigcap_{\mathcal{T} \in \mathrm{T}} \mathcal{T}$$

$\mathcal{S}$ is known as the *symmetric closure* of $\mathcal{R}$.

7.2 Functions revisited

Now that we have the language of relations, we can properly define functions. Recall that a function $f : A \to B$ is a rule that assigns one, and only one, element of B to each element of A. We may therefore view f as a collection of ordered pairs in $A \times B$:

$$\big\{(a, f(a)) : a \in A\big\}.$$

This set is nothing more than the *graph* of the function, and, being a set of ordered pairs, it is a relation.

Definition 7.4. Let $\mathcal{R} \subseteq A \times B$ be a relation from A to B. The **domain** and **range** of $\mathcal{R}$ are the sets

$$\operatorname{dom}(\mathcal{R}) = \{a \in A : (a,b) \in \mathcal{R} \text{ for some } b \in B\},$$
$$\operatorname{range}(\mathcal{R}) = \{b \in B : (a,b) \in \mathcal{R} \text{ for some } a \in A\}.$$

A **function** from A to B is a relation $f \subseteq A \times B$ satisfying the following conditions:

7.2.1 $\operatorname{dom}(f) = A$,

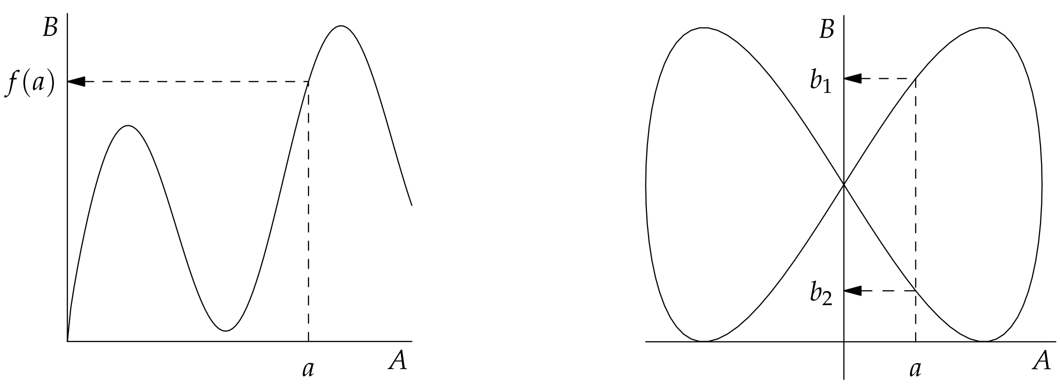
7.2.2 $(a, b_1), (a, b_2) \in f \implies b_1 = b_2$.

The two conditions can be thought of as saying:

7.2.1 Every element of A is related to *at least one* element of B.

7.2.2 Every element of A is related to *at most one* element of B.

Putting these together, we see that a relation $f \subseteq A \times B$ is a function if *every* $a \in A$ is the first entry of one (and only one) ordered pair $(a, b) \in f$. The second condition is the vertical line test, familiar from calculus.



$b_1 = b_2 = f(a)$: a function

$b_1 \neq b_2$: not a function

We can also think about injectivity and surjectivity (recall Definition 3.14) in this context. A function $f \subseteq A \times B$ is:

- *Injective* if no two pairs in f share the same second entry.
- *Surjective* if every $b \in B$ appears as the second entry of at least one pair in f.
- *Bijective* if every $b \in B$ appears as the second entry of one (and only one) ordered pair $(a, b) \in f$.

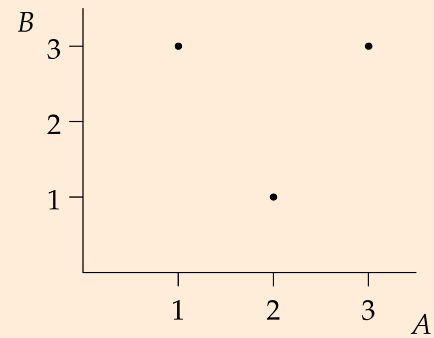
Example. Let $A = B = \{1,2,3\}$ and consider the relation

$$f = \{(1,3),(2,1),(3,3)\}.$$

Observe that $\text{dom}(f) = \{1,2,3\} = A$, and that each element of A appears exactly once as the first element in a pair $(a,b) \in f$. The relation therefore satisfies both conditions necessary to be a function. In more elementary language we would write $f(1) = 3$, $f(2) = 1$ and $f(3) = 3$.

Since 3 appears twice as a second entry of an ordered pair in f we see that f is *not injective.*

Since 2 never appears as the second entry of an ordered pair in f we see that f is *not surjective.*



A function $f : A \to B$

Example. Let A be any set and define a relation $\text{id}_A : A \to A$ by

$$\text{id}_A = \{(a,a) : a \in A\}.$$

Then id_A is a bijective function (check this!) called the *identity function on A*.

The Inverse of a Function

Since every function is a relation, it is a straightforward business to define the inverse of a function.

Definition 7.5. The **inverse** of a function $f \subseteq A \times B$ is the inverse relation $f^{-1} \subseteq B \times A$.

To compute an inverse relation we simply reverse the components of each ordered pair: the following should therefore be clear.

Theorem 7.6. $\text{dom}(f^{-1}) = \text{range}(f)$ *and* $\text{range}(f^{-1}) = \text{dom}(f)$.

In general, you should expect the inverse of a function to be merely a relation and not a function in its own right. We shall shortly (Theorem 7.7) discuss when the inverse relation is a function.

Example (cont.)**.** Consider the above example.

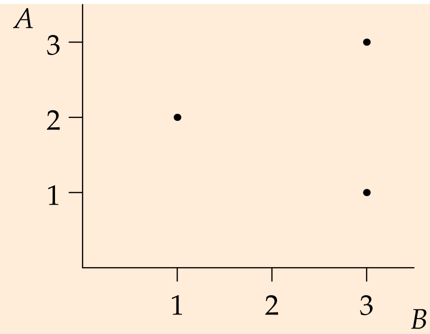
The inverse relation

$$f^{-1} = \{(3,1),(1,2),(3,3)\} \subseteq B \times A$$

is *not* a function due to failing *both* conditions of Definition 7.4.

- $\text{dom}(f^{-1}) = \{1,3\}$ is not the whole of B.
- $(3,1) \in f^{-1}$ and $(3,3) \in f^{-1}$, but $1 \neq 3$.

Both failures are clearly visible in the picture.



$f^{-1} \subseteq B \times A$: not a function

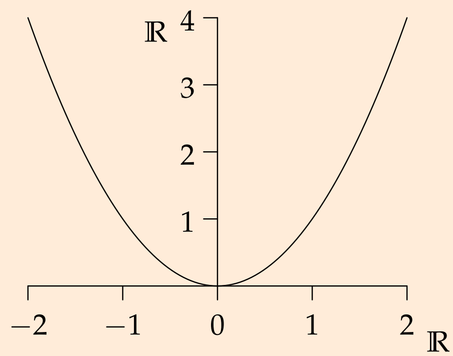
Before we consider exactly when the inverse of a function is a function in its own right, we consider a few more examples.

Examples. 7.2.1 Let $A = B = \mathbb{R}$ and $f = \{(x,x^2) : x \in \mathbb{R}\}$. This is simply the function with formula $f(x) = x^2$. The inverse relation $f^{-1} \subseteq \mathbb{R} \times \mathbb{R}$ is then

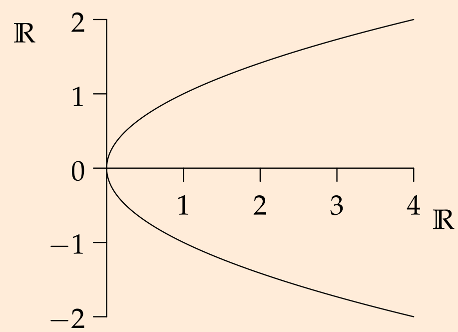
$$f^{-1} = \{(x^2,x) : x \in \mathbb{R}\} = \{(y, \pm\sqrt{y}) : y \geq 0\}.$$

In this case, f^{-1} is *not a function.* In the language of Definition 7.4:

- $\text{dom}(f^{-1}) = \mathbb{R}_0^+ \neq B$. E.g., $-1 \in B$ but $-1 \notin \text{dom}(f^{-1})$.
- $(4,2)$ and $(4,-2)$ are distinct elements of f^{-1} with the same first entry.



$f : A \to B$



$f^{-1} \subseteq B \times A$: not a function

It should be obvious that f is neither injective nor surjective: in the language of relations,

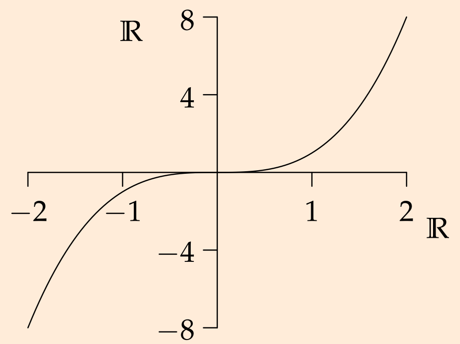
Not injective $(2,4)$ and $(-2,4)$ are distinct elements of f with the same second entry.

Not surjective For instance, -1 never appears as the second entry of any pair in f.

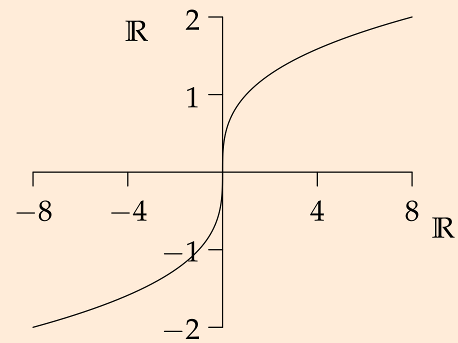
Observe how these are merely a rewriting of what it means for f^{-1} to fail to be a function.

7.2.2 Let $A = B = \mathbb{R}$ and $f = \{(x, x^3) : x \in \mathbb{R}\}$, so that f has formula $f(x) = x^3$. This time, the inverse is also a function and we could write $f^{-1}(y) = \sqrt[3]{y}$:

$$f^{-1} = \left\{(x^3, x) : x \in \mathbb{R}\right\} = \left\{(y, \sqrt[3]{y}) : y \in \mathbb{R}\right\}.$$



$f : A \to B$



$f^{-1} : B \to A$ is a function

All three of our examples help to illustrate the following important result.

Theorem 7.7. *A relation $f^{-1} \subseteq B \times A$ is a function $\iff$ f is bijective (both injective and surjective).*

Proof. Recalling Definition 7.4, we see that

$$f^{-1} \text{ is a function} \iff \begin{cases} \operatorname{dom}(f^{-1}) = B, \\ \qquad \textit{and} \\ (b, a_1), (b, a_2) \in f^{-1} \implies a_1 = a_2. \end{cases}$$

The first of these is equivalent to $\operatorname{range}(f) = B$, which says that f is surjective.
The second is equivalent to $(a_1, b), (a_2, b) \in f \implies a_1 = a_2$, which says that f is injective. ∎

Here is a final example, where the function f is harder to visualize.

Example. Let $A = \mathbb{R}$, $B = \mathbb{Q}$ and define f using the formula

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

In the language of relations, this is $f = \left\{(x, x) : x \in \mathbb{Q}\right\} \cup \left\{(x, 0) : x \notin \mathbb{Q}\right\}$.
This is a surjective function since every element of $B = \mathbb{Q}$ appears as the second entry in an ordered pair $(a, b) \in f$. It is not injective since zero appears more than once in the second entry. For example,

$$(\sqrt{2}, 0), \ (\sqrt{3}, 0) \in f.$$

Written in the more common manner, we are observing that $f(\sqrt{3}) = f(\sqrt{2})$.

The inverse f^{-1} is not a function, and it fails to be so precisely because f is non-injective. For example

$(0, \sqrt{2})$ and $(0, \sqrt{3})$ are distinct elements of f^{-1} with the same first component.

Inverse Images Recall that in Section 3.4, we can defined the *preimage* of a subset $V \subseteq B$ under a function $f : A \to B$ by

$$f^{-1}(V) = \{a \in A : f(a) \in V\}.$$

In particular, if $\{b\} \subseteq B$ has only one element, then its preimage is

$$f^{-1}(\{b\}) = \{a \in A : f(a) = b\}.$$

Both are *subsets* of A. For instance, in the last example the preimage of $\{0\}$ consists of zero and all irrational numbers!

$$f^{-1}(\{0\}) = \{0\} \cup (\mathbb{R} \setminus \mathbb{Q})$$

When $f^{-1} \subseteq B \times A$ is a function, each preimage of a singleton consists of one point of A: thus $f^{-1}(\{b\}) = \{a\}$. *Only* in such a case are we entitled to write $f^{-1}(b) = a$.

Aside. Equality of functions

There are two competing notions of what it means for two functions to be *equal.*

Same domain, same graph, same codomain $f = g$ means that f and g are the same subset of the *same* $A \times B$. This notion is preferred by set theorists because it sticks rigidly to the idea that a function is a *relation,* and it requires both the domain A and codomain B to be explicit.

Same domain, same graph $f = g$ means that $f \subseteq A \times B$, $g \subseteq A \times C$, and

$$(a, b) \in f \iff (a, b) \in g.$$

This notion considers what a function *does* to be fundamental; if two functions do the same thing to elements of the same domain then they are the same. This looser notion of equality is used more often, especially in elementary calculus.

The second conception of equality, while intuitive, has a problem. For example, let

$$f : \mathbb{R} \to \mathbb{R}, \quad \text{and} \quad g : \mathbb{R} \to [-1, 1] \quad \text{satisfy} \quad f(x) = g(x) = \sin x.$$

Although f and g have the same graph, the different codomains of f and g mean that these are *different functions* with respect to the first notion. Under the second notion, they are the *same function.* However, g is surjective while f is not, so wouldn't we prefer f and g to be non-equal?[a]

The same problem does not arise when considering domains. For example, in calculus you might have compared functions such as

$$f(x) = x^2 + 2, \quad \text{and} \quad g(x) = \frac{(x^2 + 2)(x - 1)}{x - 1}.$$

The implied domains of these functions are $\text{dom}(f) = \mathbb{R}$ and $\text{dom}(g) = \mathbb{R} \setminus \{1\}$. Even though these have the same graph whenever *both* are defined, regardless of which notion you choose we have $f \neq g$, since the functions have *different domains.*

[a] In elementary calculus, we usually say that a function is invertible if it is 1–1. In order for this to make sense, we have to ignore surjectivity and use the second notion of functional equality.

Reading Questions

7.2.1 What does it mean for a *relation* $\mathcal{R} \subseteq A \times B$ to be a *function*? Select all that apply.

(a) $\text{dom}(\mathcal{R}) = A$

(b) $\text{range}(\mathcal{R}) = B$

(c) for any $a \in A$, if $(a, b_1), (a, b_2) \in \mathcal{R}$, then $b_1 = b_2$

(d) for any $b \in \text{range}(\mathcal{R})$, if $(a_1, b), (a_2, b) \in \mathcal{R}$, then $a_1 = a_2$

7.2.2 Let $f : A \to B$ be a function. If $f^{-1} : B \to A$ is a function, this means in particular that $\text{dom}(f^{-1}) = B$. This is equivalent to what property of f?

(a) injectivity

(b) surjectivity

(c) $\text{dom}(f) = A$

(d) f is a symmetric relation.

7.2.3 True or False: a relation $\mathcal{R}$ has a domain and range if and only if it is a function.

Practice Problems

7.2.1 Let $f : A \to A$ be a function. Viewing f as a relation, if f is symmetric, what can be said about f?

7.2.2 (a) Express the function $f : \mathbb{R} \to \mathbb{R} : x \mapsto x^2$ as a relation.

(b) What is the inverse relation f^{-1}?

(c) Use Definition 7.4 to prove that the relation f^{-1} is *not* a function.

(d) Prove directly from Definition 3.14 that f is not injective and not surjective. Compare your arguments with your answer to part (c).

Exercises

7.2.1 Suppose that $f \subseteq \{1, 2, 3, 4\} \times \{1, 2, 3, 4, 5, 6, 7\}$ is the relation

$$f = \{(1, 1), (2, 3), (3, 5), (4, 7)\}.$$

(a) Show that f is a function $f : \{1, 2, 3, 4\} \to \{1, 2, 3, 4, 5, 6, 7\}$. Can you find a concise formula $f(x)$ to describe f?

(b) Is f injective? Justify your answer.

(c) Suppose that $g \subseteq \{1,2,3,4\} \times B$ is another relation so that the *graphs* of f and g are identical: i.e.

$$\{(a, f(a)) : a \in \{1,2,3,4\}\} = \{(a, g(a)) : a \in \{1,2,3,4\}\}.$$

as sets. If g is a bijective function, what is B?

7.2.2 Decide whether each of the following relations are functions. For those which are, decide whether the function is injective and/or surjective.

(a) $\mathcal{R} = \{(x,y) \in [-1,1] \times [-1,1] : x^2 + y^2 = 1\}$
(b) $\mathcal{S} = \{(x,y) \in [-1,1] \times [0,1] : x^2 + y^2 = 1\}$
(c) $\mathcal{T} = \{(x,y) \in [0,1] \times [-1,1] : x^2 + y^2 = 1\}$
(d) $\mathcal{U} = \{(x,y) \in [0,1] \times [0,1] : x^2 + y^2 = 1\}$

7.2.3 In Example 2 on page 191, explain why the function f is both injective and surjective using the language of relations: i.e., in the same manner as we analyzed Example 1.

7.2.4 Find a function $f : \mathbb{R} \to \mathbb{R}$ such that simultaneously satisfies: (1) $f(x) \neq x$ for all $x \in \mathbb{R}$ and (2) $f = f^{-1}$. Is it possible to find such a function from $\{1,2,3\} \to \{1,2,3\}$?

7.2.5 For each of the examples on page 191, compute the following preimages:

(a) $f^{-1}(\{0,1\})$
(b) $f^{-1}\Big([0,1)\Big)$
(c) $f^{-1}\Big((-\infty,0]\Big)$
(d) $f^{-1}\Big(\{-8\} \cup [-7,2] \cup (3,9)\Big)$

7.2.6 Let A and B be nonempty and $f : A \to B$ be a function.

(a) Prove that f is surjective if and only if $f^{-1}(\{b\})$ has at least one element, for all $b \in B$.
(b) Prove that f is injective if and only if $f^{-1}(\{b\})$ has at most one element, for all $b \in B$.
(c) Prove that f is bijective if and only if $f^{-1}(\{b\})$ has exactly one element, for all $b \in B$.

7.2.7 (a) Express the function $f : \mathbb{R} \to \mathbb{R} : x \mapsto x^4 + 3$ as a relation.
(b) What is the inverse relation f^{-1}?
(c) Use Definition 7.4 to prove that the relation f^{-1} is *not* a function.
(d) Prove directly from Definition 3.14 that f is not injective and not surjective. Compare your arguments with your answer to part (c).

7.2.8 Repeat the previous question for $f : \mathbb{R} \to \mathbb{R} : x \mapsto \sqrt{x^2 - 4x + 5}$.

7.2.9 Give a formal proof of Theorem 7.6.

7.2.10 Prove or disprove the following: if $f : A \to B$ is a function, and $U, V \subseteq B$, then

$$f^{-1}(U \cap V) = f^{-1}(U) \cap f^{-1}(V)$$

7.2.11 Let A and B be nonempty and $f : A \to B$ be a function.

(a) Suppose f is a bijection. Show that f^{-1} satisfies $f \circ f^{-1} = \mathrm{id}_B$ and $f^{-1} \circ f = \mathrm{id}_A$.

(b) Suppose that there is a function $g : B \to A$ such that $f \circ g = \mathrm{id}_B$ and $g \circ f = \mathrm{id}_A$. Show f is a bijection and that $g = f^{-1}$.

(c) Suppose there are functions $g : B \to A$ and $h : B \to A$ such that $f \circ g = \mathrm{id}_B$ and $h \circ f = \mathrm{id}_A$. Show that $g = h$ and that both are also equal to f^{-1}.

7.2.12 Let A and B be nonempty and $f : A \to B$ be a function.

(a) Prove that f is surjective if and only if there is $g : B \to A$ such that $f \circ g = \mathrm{id}_B$.

(b) Prove that f is injective if and only if there is $h : B \to A$ such that $h \circ f = \mathrm{id}_A$.

7.2.13 Let A and B be nonempty and $f : A \to B$ be a function.

(a) Prove that f is surjective if and only if for all sets C and functions $g, h : B \to C$, $g \circ f = h \circ f$ implies $g = h$.

(b) Prove that f is injective if and only if for all sets C and functions $g, h : C \to A$, $f \circ g = f \circ h$ implies $g = h$.

7.3 Equivalence Relations

In mathematics, the notion of *equality* is not as simple as one might think. The idea of two numbers being equal is straightforward, but suppose we want to consider two paths between given points as 'equal' if and only if they have the same length? Since two 'equal' paths might look very different, is this a good notion of equality? Mathematicians often want to gather together objects that have a common property and then treat them as if they were a single object. This is done using equivalence relations and equivalence classes.

First recall the alternative notation for a relation on a set A: if $\mathcal{R} \subseteq A \times A$ is a relation on A, then $x \,\mathcal{R}\, y$ has the same meaning as $(x, y) \in \mathcal{R}$. We might read $x \,\mathcal{R}\, y$ as 'x is $\mathcal{R}$-related to y.'

Definition 7.8. A relation $\mathcal{R}$ on a set A may be described as **reflexive, symmetric** or **transitive** if it satisfies the following properties:

Reflexivity $\forall x \in A,\ x \,\mathcal{R}\, x$ (every element of A is related to itself)

Symmetry $\forall x, y \in A,\ x \,\mathcal{R}\, y \implies y \, R \, x$ (if x is related to y, then y is related to x)

Transitivity $\forall x, y, z \in A,\ x \,\mathcal{R}\, y \text{ and } y \,\mathcal{R}\, z \implies x \,\mathcal{R}\, z$ (if x is related to y, and y is related to z, then x is related to z)

Symmetry is exactly the same notion as in Definition 7.2.

Examples. 7.3.1 Let $A = \mathbb{R}$ and let $\mathcal{R}$ be $\le$. Thus $2 \le 3$, but $7 \not\le 4$. We check whether $\mathcal{R}$ satisfies the above properties.

Reflexivity True. $\forall x \in \mathbb{R},\ x \le x$.

Symmetry False. For example, $2 \le 3$ but $3 \not\le 2$.

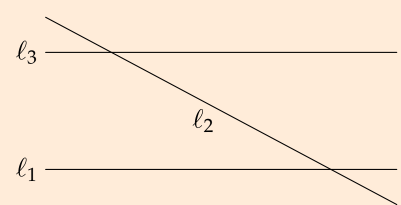
Transitivity True. $\forall x, y, z \in \mathbb{R}$, if $x \le y$ and $y \le z$, then $x \le z$.

7.3.2 Let A be the set of lines in the plane and define $\ell_1 \, R \, \ell_2 \iff \ell_1$ and ℓ_2 intersect.

Reflexivity True. Every line intersects itself, so $\ell \,\mathcal{R}\, \ell$ for all $\ell \in A$.

Symmetry True. For all lines $\ell_1, \ell_2 \in A$, if ℓ_1 intersects ℓ_2, then ℓ_2 intersects ℓ_1..

Transitivity False. As the picture illustrates, we may let ℓ_1 and ℓ_3 be parallel lines, and ℓ_2 cross both of these. Then $\ell_1 \,\mathcal{R}\, \ell_2$ and $\ell_2 \,\mathcal{R}\, \ell_3$, but $\ell_1 \not\!\mathcal{R}\, \ell_3$.



Definition 7.9. An **equivalence relation** is a relation $\sim$ which is reflexive, symmetric and transitive.

The symbol $\sim$ is almost universally used for an abstract equivalence relation. It can be read as 'related to,' 'tilde,' or 'twiddles.' The two examples above are *not* equivalence relations because they fail one of the three conditions. We now exhibit the simplest equivalence relation.

Example. Equals '=' is an equivalence relation on any set, hence the name!

Read the definitions of reflexive, symmetric and transitive until you are certain of this fact. There are countless other equivalence relations: here are a few.

Examples. 7.3.1 For all $x, y \in \mathbb{Z}$, we define the relation $\sim$ by

$$x \sim y \iff x - y \text{ is even.}$$

We claim that $\sim$ is an equivalence relation on $\mathbb{Z}$.

Reflexivity $\forall x \in \mathbb{Z},\ x - x = 0$ is even, hence $x \sim x$.

Symmetry $\forall x, y \in \mathbb{Z},\ x \sim y \implies x - y$ is even $\implies y - x$ is even $\implies y \sim x$.

Transitivity $\forall x, y, z \in \mathbb{Z}$, if $x \sim y$ and $y \sim z$, then $x - y$ and $y - z$ are even. But the sum of two even numbers is even, hence $x - z = (x - y) + (y - z)$ is even, and so $x \sim z$.

7.3.2 Let $A = \{\text{all students taking this course}\}$. For all $x, y \in A$, let

$$x \sim y \iff x \text{ achieves the same letter-grade as } y.$$

Then $\sim$ is an equivalence relation on A; here is the proof.

Reflexivity $\forall x \in A,\ x \sim x$ since everyone scores the same as themself!

Symmetry $\forall x, y \in A,\ x \sim y \implies x$ achieves the same letter-grade as y $\implies y$ achieves the same letter-grade as x $\implies y \sim x$

Transitivity $\forall x, y, z \in A$, if $x \sim y$ and $y \sim z$, then x achieves the same as y who achieves the same as z, whence x achieves the same as z. Thus $x \sim z$.

7.3.3 We define an equivalence relation on $\mathbb{Z}$ by

$$\forall x, y \in \mathbb{Z},\ \ x \sim y \iff x^2 \equiv y^2 \pmod 5.$$

Reflexivity $\forall x \in \mathbb{Z},\ x \sim x$ since x^2 is always congruent to itself!

Symmetry $\forall x, y \in \mathbb{Z},\ x \sim y \implies x^2 \equiv y^2 \pmod 5 \implies y^2 \equiv x^2 \pmod 5 \implies y \sim x$

Transitivity $\forall x, y, z \in \mathbb{Z}$, if $x \sim y$ and $y \sim z$, then $x^2 \equiv y^2$ and $y^2 \equiv z^2 \pmod 5$. But then $x^2 \equiv z^2 \pmod 5$ and so $x \sim z$.

The most important thing to observe in each of these examples is that **an equivalence relation separates elements of a set into subsets where elements share a common property** (even/oddness, letter-grade, etc.). The next definition formalizes this idea.

Definition 7.10. Let $\sim$ be an equivalence relation on a set X. The **equivalence class** of an element $x \in X$ is the set

$$[x] = \{y \in X : y \sim x\}.$$

Otherwise said, $y \sim x \iff y \in [x]$. The set of all equivalence classes is known as the **quotient** of X by $\sim$ or simply 'X mod $\sim$,' and is denoted

$$X/_{\sim} = \Big\{[x] : x \in X\Big\}$$

Let us think about the definition of equivalence class in the context of our previous examples.

Examples. 7.3.1 $[0] = \{y \in \mathbb{Z} : y \sim 0\} = \{y \in \mathbb{Z} : y \text{ is even}\}$ is the set of even numbers. Note that $[0] = [2] = [4] = [6]$, etc. The other equivalence class is $[1] = \{y \in \mathbb{Z} : y - 1 \text{ is even}\}$, which is the set of odd numbers. The quotient set is

$$\mathbb{Z}/_{\sim} = \big\{[0], [1]\big\} = \big\{\{\text{even numbers}\}, \{\text{odd numbers}\}\big\}.$$

7.3.2 There is one equivalence class for each letter grade awarded. Each equivalence class contains all the students who obtain a particular letter-grade. If we call the equivalence classes $\mathrm{A}^+, \mathrm{A}, \mathrm{A}^-, \mathrm{B}^+, \ldots, \mathrm{F}$, where, say, $\mathrm{B} = \{\text{students obtaining a B-grade}\}$, then

$$\{\text{Students}\}/_{\sim} = \{\mathrm{A}^+, \mathrm{A}, \mathrm{A}^-, \mathrm{B}^+, \ldots, \mathrm{F}\}.$$

7.3.3 The equivalence classes for this example are a little tricky. First observe that

$$x \equiv y \pmod 5 \implies x^2 \equiv y^2 \pmod 5,$$

so that there are at most five equivalence classes; those of 0, 1, 2, 3 and 4. Are they distinct? If we square each of these and consider the remainder modulo 5, we obtain

$x \pmod 5$	0	1	2	3	4
$x^2 \pmod 5$	0	1	4	4	1

Notice that $1 \sim 4$, so they share an equivalence class. Similarly $2 \sim 3$. Indeed the distinct equivalence classes are

$$[0] = \{x \in \mathbb{Z} : x \equiv 0 \pmod 5\}$$
$$[1] = \{x \in \mathbb{Z} : x \equiv 1, 4 \pmod 5\}$$
$$[2] = \{x \in \mathbb{Z} : x \equiv 2, 3 \pmod 5\}$$

In this case the quotient is the set

$$\mathbb{Z}/_{\sim} = \Big\{[0], [1], [2]\Big\}.$$

Here is one further example of an equivalence relation, this time on $\mathbb{R}^2$. Be careful with the notation: $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is already a Cartesian product, so a relation on $\mathbb{R}^2$ is a subset of $\mathbb{R}^2 \times \mathbb{R}^2$!

Example. Let $\sim$ be the relation on $\mathbb{R}^2$ defined by $(x, y) \sim (v, w) \iff x^2 + y^2 = v^2 + w^2$. We claim that this is an equivalence relation.

Reflexivity $\forall (x, y) \in \mathbb{R}^2,\ x^2 + y^2 = x^2 + y^2$.

Symmetry $\forall (x, y), (v, w) \in \mathbb{R}^2,\ (x, y) \sim (v, w) \implies x^2 + y^2 = v^2 + w^2$
$\implies v^2 + w^2 = x^2 + y^2$
$\implies (v, w) \sim (x, y)$

Transitivity $\forall (x, y), (v, w), (p, q) \in \mathbb{R}^2$, if $(x, y) \sim (v, w)$ and $(v, w) \sim (p, q)$, then $x^2 + y^2 = v^2 + w^2$ and $v^2 + w^2 = p^2 + q^2$. But then $x^2 + y^2 = p^2 + q^2$ and so $(x, y) \sim (p, q)$.

$\sim$ is therefore an equivalence relation. But what are the equivalence classes? By definition,

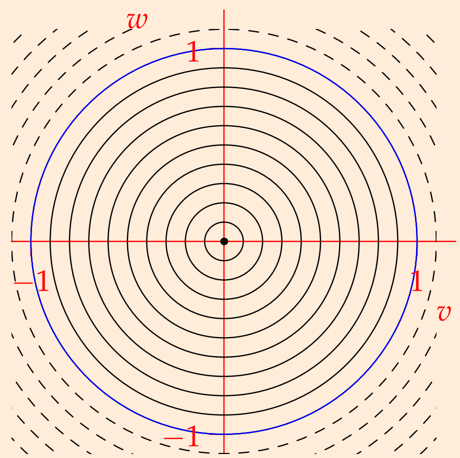
$$[(x, y)] = \left\{ (v, w) \in \mathbb{R}^2 : v^2 + w^2 = x^2 + y^2 \right\}.$$

This isn't particularly helpful. Indeed it is easier to think of each of these sets as

$$\left\{ (v, w) \in \mathbb{R}^2 : v^2 + w^2 \text{ is } \textit{constant} \right\}.$$

Each equivalence class is therefore a *circle* centered at the origin! Some of the equivalence classes are drawn in the picture: the class $[(1, 0)]$ is highlighted. Moreover, the quotient set is

$$\mathbb{R}^2/_{\sim} = \{\text{circles centered at the origin}\}.$$



Reading Quiz

7.3.1 True or False: a relation $\sim$ on a set X is *reflexive* if $\exists x \in X$ such that $x \sim x$.

7.3.2 Suppose that $x, y, z \in X$ and $\sim$ is an equivalence relation on X. Express each of the following assertions in terms of the properties satisfied by an equivalence relation.

(1) $x \in [y]$ and $y \in [z] \implies x \in [z]$.

(2) $x \in [x]$.

(3) $x \in [y] \iff y \in [x]$.

(a) (1) is reflexivity, (2) is symmetry, and (3) is transitivity

(b) (1) is transitivity, (2) is symmetry, and (3) is reflexivity

(c) (1) is transitivity, (2) is reflexivity, and (3) is antisymmetry

(d) (1) is transitivity, (2) is reflexivity, and (3) is symmetry

7.3.3 Suppose $\mathcal{R}$ is an equivalence relation on a set X. Then $\mathcal{R}^{-1}$ is __________ also an equivalence relation.

(a) never

(b) sometimes

(c) always

Practice Problems

7.3.1 Define $\mathcal{R}$ on $\mathbb{N}_{\geq 2}$ by $a\,\mathcal{R}\,b$ if and only if $\gcd(a,b) > 1$. Determine whether or not $\mathcal{R}$ is reflexive, symmetric, or transitive.

7.3.2 Let $\sim$ be the relation on $\mathbb{R}$ defined by $x \sim y$ if and only if $x - y \in \mathbb{Z}$.

(a) Prove that $\sim$ is an equivalence relation.

(b) List three distinct elements of the equivalence class $[5/2]$. In general, what is an equivalence class $[x]$ as a set?

(c) Describe the quotient $\mathbb{R}/\sim$.

Exercises

7.3.1 A relation $\mathcal{R}$ is *antisymmetric* if $((x,y) \in \mathcal{R}) \wedge ((y,x) \in \mathcal{R}) \implies x = y$. Give examples of relations $\mathcal{R}$ on $A = \{1,2,3\}$ having the stated property.

(a) $\mathcal{R}$ is both symmetric and antisymmetric.

(b) $\mathcal{R}$ is neither symmetric nor antisymmetric.

(c) $\mathcal{R}$ is transitive but $\mathcal{R} \cup \mathcal{R}^{-1}$ is not transitive.

7.3.2 A relation $\mathcal{R}$ on a set X is called a *partial order* if it is reflexive, antisymmetric, and transitive. Show the divisibility relation $|$ is a partial order on $\mathbb{N}$.

7.3.3 Let $\mathcal{S} = \{(x,y) \in \mathbb{R}^2 : \sin^2 x + \cos^2 y = 1\}$.

(a) Give an example of two real numbers x, y such that $x\,\mathcal{S}\,y$.

(b) Is $\mathcal{S}$ reflexive? Symmetric? Transitive? Justify your answers.

7.3.4 Each of the following relations $\sim$ is an equivalence relation on $\mathbb{R}^2$. Identify the equivalence classes and draw several of them.

(a) $(a,b) \sim (c,d) \iff ab = cd$.

(b) $(v,w) \sim (x,y) \iff v^2 w = x^2 y$.

7.3.5 (a) Let $\sim$ be the relation defined on $\mathbb{Z}$ by $a \sim b \iff a + b$ is even. Show that $\sim$ is an equivalence relation and determine the distinct equivalence classes.

(b) Suppose that 'even' is replaced by 'odd' in part (a). Which of the properties reflexive, symmetric, transitive does $\sim$ possess?

7.3.6 For each of the following relations $\mathcal{R}$ on $\mathbb{Z}$, decide whether $\mathcal{R}$ is reflexive, symmetric, or transitive, and whether $\mathcal{R}$ is an equivalence relation.

(a) $a \mathcal{R} b \iff a \equiv b \pmod 3$ **or** $a \equiv b \pmod 4$.

(b) $a \mathcal{R} b \iff a \equiv b \pmod 3$ **and** $a \equiv b \pmod 4$.

7.3.7 For the purposes of this question, we call a real number x *small* if $|x| \leq 1$. Let $\mathcal{R}$ be the relation on the set of real numbers defined by

$$x \mathcal{R} y \iff x - y \text{ is small.}$$

Prove or disprove: $\mathcal{R}$ is an equivalence relation on $\mathbb{R}$.

7.3.8 Let $A = \{1, 2, 3, 4, 5, 6\}$. The distinct equivalence classes resulting from an equivalence relation $\sim$ on A are $\{1, 4, 5\}$, $\{2, 6\}$, and $\{3\}$. What is $\sim$? Give your answer as a subset of $A \times A$.

7.3.9 $\subseteq$ is a relation on any set of sets. Is $\subseteq$ reflexive, symmetric, transitive? Prove your assertions.

7.3.10 Let S be the set of all polynomials of degree at most 3. An element $s \in S$ can then be expressed as

$$s(x) = ax^3 + bx^2 + cx + d, \qquad \text{where } a, b, c, d \in \mathbb{R}.$$

A relation $\mathcal{R}$ on S is defined by

$$p \mathcal{R} q \iff p \text{ and } q \text{ have a common root.}$$

For example $p(x) = (x-1)^2$ and $q(x) = x^2 - 1$ have the root 1 in common so that $p \mathcal{R} q$. Determine which of the properties reflexive, symmetric and transitive are possessed by $\mathcal{R}$.

7.3.11 Let $A = \{2^m : m \in \mathbb{Z}\}$. A relation $\sim$ is defined on the set $\mathbb{Q}^+$ of positive rational numbers by

$$a \sim b \iff \frac{a}{b} \in A$$

(a) Show that $\sim$ is an equivalence relation.

(b) Describe the elements in the equivalence class [3].

7.3.12 A relation is defined on the set $A = \{a + b\sqrt{2} : a, b \in \mathbb{Q},\, a + b\sqrt{2} \neq 0\}$ by $x \sim y \iff \frac{x}{y} \in \mathbb{Q}$. Show that $\sim$ is an equivalence relation and determine the distinct equivalence classes.

7.3.13 Let $\mathcal{R}$ and $\mathcal{S}$ be equivalence relations on a set X.

(a) Show that $\mathcal{R} \cap \mathcal{S}$ is an equivalence relation.

(b) Does $\mathcal{R} \cup \mathcal{S}$ have to be an equivalence relation? Prove or give counterexample.

(c) Determine what the equivalence classes of $\mathcal{R} \cap \mathcal{S}$ are in terms of the equivalence classes for $\mathcal{R}$ and $\mathcal{S}$ individually.

7.3.14 Let $f : A \to B$ be a function. Define $\sim$ on A by $a \sim b$ if and only if $f(a) = f(b)$.

(a) Show $\sim$ is an equivalence relation.

(b) Describe the equivalence classes of $\sim$.

(c) Show that f is injective if and only if $\sim$ is equality.

7.3.15 Let X be a set and $\sim$ an equivalence relation on X. Define $\pi : X \to X/\sim$ by $\pi(x) = [x]$. Show that π is a surjection. Prove that π is an injection if and only if $\sim$ is equality.

7.3.16 Let $\mathcal{R}$ be a relation on a set X. Show that $\mathcal{S} = \mathcal{R} \cup \{(x,x) : x \in X\}$ is the smallest reflexive relation on X containing $\mathcal{R}$. That is,

(a) $\mathcal{S}$ is reflexive.

(b) $\mathcal{R} \subseteq \mathcal{S}$.

(c) If $\mathcal{T}$ is a reflexive relation on X such that $\mathcal{R} \subseteq \mathcal{T}$, then $\mathcal{S} \subseteq \mathcal{T}$.

We call $\mathcal{S}$ the *reflexive closure* of $\mathcal{R}$.

7.3.17 Recall the description of the real projective line (page 171): if A_m is the line through the origin with gradient m, then

$$\mathbb{P}(\mathbb{R}^2) = \{A_m : m \in \mathbb{R} \cup \{\infty\}\}.$$

Define a relation $\sim$ on $\mathbb{R}^2_* = \mathbb{R}^2 \setminus \{(0,0)\}$ by $(a,b) \sim (c,d) \iff ad = bc$.

(a) Prove that $\sim$ is an equivalence relation.

(b) Find the equivalence classes of $\sim$. How do the equivalence classes differ from the lines A_m?

7.3.18 Let X be a set. Suppose we have a function $\text{cl} : \mathcal{P}(X) \to \mathcal{P}(X)$ which satisfies the following properties:

(i) (Reflexivity) $A \subseteq \text{cl}(A)$ for all $A \in \mathcal{P}(X)$;

(ii) (Monotonicity) if $A \subseteq B$, then $\text{cl}(A) \subseteq \text{cl}(B)$ for all $A, B \in \mathcal{P}(X)$;

(iii) (Idempotence) $\text{cl}(\text{cl}(A)) = \text{cl}(A)$ for all $A \in \mathcal{P}(X)$;

(iv) (Exchange) if $a \in \text{cl}(A \cup \{b\}) \setminus \text{cl}(A)$, then $b \in \text{cl}(A \cup \{a\})$ for all $a, b \in X$.

Define $\sim$ on $X \setminus \text{cl}(\emptyset)$ by $a \sim b$ if and only if $a \in \text{cl}(\{b\})$.

(a) Show $\sim$ is an equivalence relation.

(b) Show $a \sim b$ if and only if $\text{cl}(\{a\}) = \text{cl}(\{b\})$.

(c) Prove that $[a] = \text{cl}(\{a\}) \setminus \text{cl}(\emptyset)$.

7.3.19 Suppose that $\mathcal{R}, \mathcal{S}$ are relations on some set X. Define the *composition* $\mathcal{R} \circ \mathcal{S}$ to be the relation

$$(a,c) \in \mathcal{R} \circ \mathcal{S} \iff \exists b \in X \text{ such that } (a,b) \in \mathcal{R} \text{ and } (b,c) \in \mathcal{S}.$$

(a) If $\mathcal{R} = \{(1,1),(1,2),(2,3),(3,1),(3,3)\}$ and $\mathcal{S} = \{(1,2),(1,3),(2,1),(3,3)\}$, find $\mathcal{R} \circ \mathcal{S}$.

(b) Suppose that $\mathcal{R}$ and $\mathcal{S}$ are reflexive. Prove that $\mathcal{R} \circ \mathcal{S}$ is reflexive.

(c) Suppose that $\mathcal{R}$ and $\mathcal{S}$ are symmetric. Prove that $(x,y) \in \mathcal{R} \circ \mathcal{S} \iff (y,x) \in \mathcal{S} \circ \mathcal{R}$.

(d) Give an example of symmetric relations $\mathcal{R}, \mathcal{S}$ such that $\mathcal{R} \circ \mathcal{S}$ is *not* symmetric. Conclude that if $\mathcal{R}, \mathcal{S}$ are equivalence relations, then $\mathcal{R} \circ \mathcal{S}$ need not be an equivalence relation.

7.3.20 Let $\mathcal{R}$ be a relation on a set X. Inductively define $\mathcal{R}^n$ for $n \in \mathbb{N}$ as follows:

- $\mathcal{R}^1 = \mathcal{R}$
- $\mathcal{R}^{n+1} = \mathcal{R} \circ \mathcal{R}^n$.

Set

$$\mathcal{R}^+ = \bigcup_{n \in \mathbb{N}} \mathcal{R}^n.$$

Show that $\mathcal{R}^+$ is the smallest transitive relation on X containing $\mathcal{R}$. That is,

(a) $\mathcal{R}^+$ is transitive.

(b) $\mathcal{R} \subseteq \mathcal{R}^+$.

(c) If $\mathcal{T}$ is a transitive relation on X such that $\mathcal{R} \subseteq \mathcal{T}$, then $\mathcal{R}^+ \subseteq \mathcal{T}$. [Hint: use induction to show $\mathcal{R}^n \subseteq \mathcal{T}$ for all $n \in \mathbb{N}$.]

We call $\mathcal{R}^+$ the *transitive closure* of $\mathcal{R}$.

(d) Let $X = \{a, b, c, d\}$ and $\mathcal{R} = \{(a, b), (b, c), (c, d)\}$. Compute $\mathcal{R}^+$.

7.3.21 (*Only for those who have studied Linear Algebra*) Let $\sim$ be the relation on the set of 2×2 real matrices given by $A \sim B \iff \exists M$ such that $B = MAM^{-1}$.

(a) Prove that $\sim$ is an equivalence relation.

(b) What is the equivalence class of the identity matrix?

(c) Show that $\left(\begin{smallmatrix} -11 & 15 \\ -5 & 9 \end{smallmatrix}\right) \sim \left(\begin{smallmatrix} 4 & 10 \\ 0 & -6 \end{smallmatrix}\right)$ (*Hint: think about diagonalizing*)

(d) (Hard) Suppose that $L : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map and β, γ are bases of $\mathbb{R}^2$. Suppose that $A = [L]_\beta$ and $B = [L]_\gamma$ are the matrix representations of L with respect to the two bases. Prove that $A \sim B$.

(e) (Hard) Suppose that A, B have the same, but distinct, eigenvalues $\lambda_1 \neq \lambda_2$. Prove that $A \sim B$. *Again use diagonalization, the challenge here is to make your proof work even when the eigenvalues are complex numbers.*

7.4 Partitions

Recall the important observation about our equivalence relation examples: every element of the original set of objects ends up in *exactly one equivalence class.* For instance, every integer is either even or odd but not both. The equivalence classes *partition* the original set in the same way that cutting a cake partitions the crumbs: each crumb ends up in exactly one slice. We shall prove in a moment that equivalence relations *always* do this. Before doing so we reverse the discussion.

Definition 7.11. Let X be a set and $\{A_n : n \in I\}$ be a collection of non-empty subsets $A_n \subseteq X$. We say that X is **partitioned by** the collection of subsets if

7.4.1 $X = \bigcup_{n \in I} A_n$. (the A_n together make up X)

7.4.2 If $A_m \neq A_n$, then $A_m \cap A_n = \emptyset$. (*distinct* A_n are pairwise disjoint[a])

We describe the collection $\mathcal{A}$ as a **partition** of X.

[a]Recall that two sets A, B are *disjoint* if $A \cap B = \emptyset$: see Definition 3.7. In this definition we *don't* require the sets A_n all to be different, some could be identical to each other.

The conditions can be viewed as saying that every element of X lies in (1.) *at least one* subset A_n and (2.) *at most one* subset A_n: otherwise said, every element of X lies in *exactly one* subset.

Example. Partition the set $X = \{1, 2, 3, 4, 5\}$ into subsets

$$A_1 = \{1, 3\}, \qquad A_2 = \{2, 4\}, \qquad A_3 = \{5\}.$$

Now consider the relation $\mathcal{R}$ on X, defined by

$$\mathcal{R} = \{(1, 1), (1, 3), (3, 1), (3, 3), (2, 2), (2, 4), (4, 2), (4, 4), (5, 5)\}.$$

What does $\mathcal{R}$ have to do with the partition? It should be clear that $\mathcal{R}$ could be defined by insisting that

$$x \mathcal{R} y \iff x \text{ and } y \text{ are in the } \textit{same} \text{ subset } A_n.$$

Run through your mental checklist: is $\mathcal{R}$ reflexive? symmetric? transitive? Indeed $\mathcal{R}$ is an equivalence relation! Moreover, the equivalence classes of $\mathcal{R}$ are precisely the sets A_1, A_2 and A_3. For instance, 1 is related to itself and 3, but isn't related to anything else. Indeed

$$[1] = [3] = \{1, 3\} = A_1, \qquad [2] = [4] = \{2, 4\} = A_2, \qquad [5] = \{5\} = A_5.$$

The example suggests that partitioning a set defines a natural equivalence relation. Combining this with our observations in the previous section and you should be starting to believe that *partitions and equivalence relations are essentially the same thing.* Before we prove this important fact, here are some further examples of partitions.

Examples. 7.4.1 The integers can be partitioned according to their remainder modulo 3: define

$$A_r = \{z \in \mathbb{Z} : z \equiv r \pmod 3\}.$$

Then $\mathbb{Z} = A_0 \cup A_1 \cup A_2$. This is certainly a partition:

- Every integer z has remainder of 0, 1 or 2 after division by 3, and so every integer is in some set A_r.
- No integer has two distinct remainders modulo 3, so the sets A_0, A_1, A_2 are disjoint.

7.4.2 More generally, if $n \in \mathbb{N}$, then the set of integers $\mathbb{Z}$ is partitioned into n sets $A_0, \ldots, A_{n-1}$ where

$$A_r = \{z \in \mathbb{Z} : z \equiv r \pmod n\}$$

is the set of integers with remainder r upon dividing by n. We are appealing to the Division Algorithm (Theorem 4.2) which tells us that every integer z has a *unique remainder* $r \in \{0, 1, \ldots, n-1\}$.

7.4.3 The set of real numbers $\mathbb{R}$ is partitioned into the sets of rational and irrational numbers: $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})$.

Finally, here is an example of a relation which doesn't produce a partition.

Example. Let $X = \{1, 2, 3, 4\}$ and define a relation $\mathcal{R}$ on X by

$$\mathcal{R} = \{(1,3), (1,4), (2,2), (2,3), (3,1), (3,2), (4,3), (4,4)\}.$$

Also define the subsets

$$A_n = \{x \in X : (n, x) \in \mathcal{R}\}.$$

Thus A_n is the set of all elements of X which are related to n. We quickly see that

$$A_1 = \{3,4\}, \quad A_2 = \{2,3\}, \quad A_3 = \{1,2\}, \quad A_4 = \{3,4\}.$$

The collection of sets A_n is as follows:

$$\{A_n\}_{n \in X} = \{A_1, A_2, A_3, A_4\} = \Big\{\{3,4\}, \{2,3\}, \{1,2\}\Big\},$$

where we only have *three* sets in the collection since $A_4 = A_1$. This collection is not a partition because, for instance, $2 \in \{2,3\} \cap \{1,2\}$. In the language of Definition 7.11, we have

$$\{2,3\} \neq \{1,2\} \quad \text{but} \quad \{2,3\} \cap \{1,2\} \neq \emptyset.$$

More importantly, you should convince yourself that $\mathcal{R}$ is *not* an equivalence relation.

Equivalence Relations and Partitions

Before we present the fundamental result of the chapter, we prove a helpful lemma.

Lemma 7.12. *Suppose that $\sim$ is an equivalence relation. Then $x \sim y \iff [x] = [y]$.*

Proof. ($\Leftarrow$) By reflexivity, $x \in [x]$. If $[x] = [y]$, then we have $x \in [y]$. Finally, recalling Definition 7.10, we see that that this is the same as saying $x \sim y$.

($\Rightarrow$) Suppose that $x \sim y$. We begin by showing the inclusion $[x] \subseteq [y]$. Let $z \in [x]$, then

$$z \sim x \text{ and } x \sim y \implies z \sim y \implies z \in [y]. \qquad \text{(Transitivity)}$$

Therefore $[x] \subseteq [y]$. By symmetry, we also have $y \sim x$: repeating the argument yields $[y] \subseteq [x]$, and thus $[x] = [y]$. ■

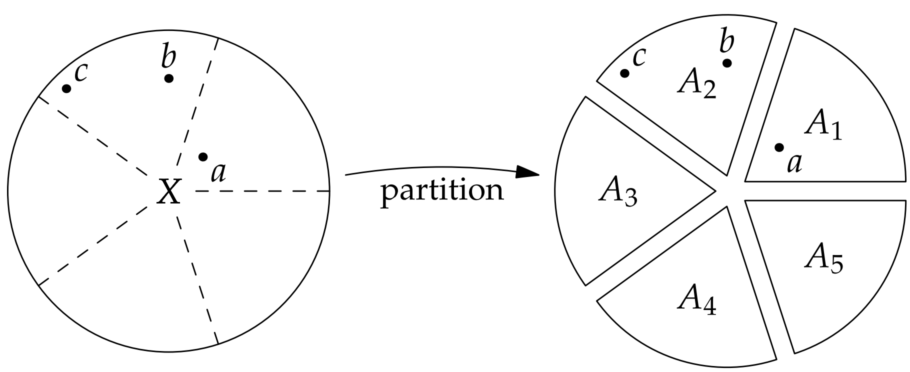
Theorem 7.13. *Let X be any set.*

7.4.1 If $\sim$ is an equivalence relation on X, then X is partitioned by the equivalence classes of $\sim$.

7.4.2 If $\{A_n : n \in I\}$ is a partition of X, then the relation $\sim$ on X defined by

$$x \sim y \iff \exists n \in I \text{ such that } x \in A_n \text{ and } y \in A_n$$

is an equivalence relation.



Each element of X ends up in exactly one subset. In the language of the Theorem, we have

$$A_1 = [a], \quad A_2 = [b] = [c], \quad b \sim c, \quad a \nsim b, \quad a \nsim c.$$

Some things to consider while reading the proof:

- Think about the picture! The result is nothing more than the notion of partitioning a cake by cutting it into slices. The slices are the equivalence classes of the obvious relation: two crumbs are related if and only if they lie in the same slice. The algebra that follows merely confirms that the picture is telling a legitimate story.

- In part 1. of the proof, look for where the reflexive, symmetric and transitive assumptions about $\sim$ are used. Why do we need $\sim$ to be an equivalence relation? Why does the proof fail if any of the three assumptions are dropped?
- Similarly, in part 2., look for where we use both parts of the definition of partition. Why are both assumptions required?

Proof. 7.4.1 Assume that $\sim$ is an equivalence relation on X. To prove that the equivalence classes of $\sim$ partition X, we must show two things:

(a) That every element of X is in some equivalence class.

(b) That the distinct equivalence classes are pairwise disjoint: if $[x] \neq [y]$, then $[x] \cap [y] = \emptyset$.

For (a), we only need reflexivity: $\forall x \in X$ we have $x \sim x$. Otherwise said, $x \in [x]$, whence every element of X is in the equivalence class defined by itself.

For (b), we prove by the contrapositive method and show that $[x] \cap [y] \neq \emptyset \implies [x] = [y]$. Assume that $[x] \cap [y] \neq \emptyset$. Then $\exists z \in [x] \cap [y]$. This gives

$$\begin{aligned} z \sim x \text{ and } z \sim y &\implies x \sim z \text{ and } z \sim y && \text{(Symmetry)} \\ &\implies x \sim y && \text{(Transitivity)} \\ &\implies [x] = [y] && \text{(Lemma 7.12)} \end{aligned}$$

We have proved (b) and therefore part 1. of the theorem.

7.4.2 Now suppose that $\{A_n : n \in I\}$ is a partition of X and define $\sim$ by

$$x \sim y \iff \exists n \in I \text{ such that } x \in A_n \text{ and } y \in A_n.$$

We must prove the reflexivity, symmetry and transitivity of $\sim$.

Reflexivity Every $x \in X$ is in some A_n. Thus $x \sim x$ for all $x \in X$.

Symmetry If $x \sim y$, then $\exists n \in I$ such that $x, y \in A_n$. But then $y, x \in A_n$ and so $y \sim x$.

Transitivity Let $x \sim y$ and $y \sim z$. Then $\exists p, q \in I$ such that $x, y \in A_p$ and $y, z \in A_q$. Since $\{A_n : n \in I\}$ is a partition and $y \in A_p \cap A_q$, we necessarily have $A_p = A_q$. Thus $x, z \in A_p$ and so $x \sim z$.

We have shown $\sim$ is an equivalence relation, and the proof is complete. ■

Reading the proof carefully, you should see that reflexivity in part 2. comes from the fact that $X = \bigcup_{n \in I} A_n$, while transitivity is due to the pairwise disjointness of the pieces of the partition. Symmetry is essentially free because the definition of $\sim$ is symmetric in x and y.

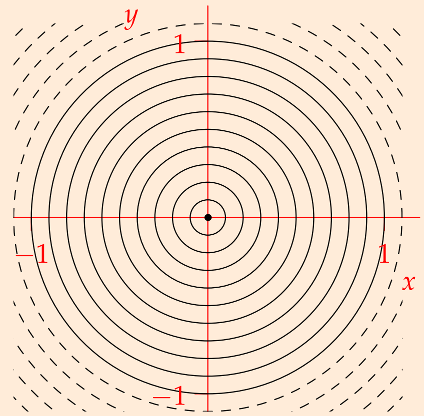
The ability to partition sets and view the resulting subsets as individual objects is crucial to advanced mathematics. The importance of the Theorem comes from the fact that equivalence relations provide a straightforward *algebraic* method of working with partitions.

Geometric Examples

The language of equivalence relations and partitions is used heavily in geometry and topology to describe complex shapes. We finish this section with several examples. Since examples of partitions are especially easy to visualize with curves in the plane, we first return to the example on page 199 and describe things in our new language.

Example. For each real number $r \geq 0$, define the set

$$A_r = \left\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = r^2\right\}.$$



This is simply the circle of radius r centered at the origin. We check that $\{A_r : r \in \mathbb{R}_0^+\}$ is a partition of $\mathbb{R}^2$.

- Every point of the plane lies on some circle. Precisely, $(x,y) \in A_{\sqrt{x^2+y^2}}$ since $\sqrt{x^2+y^2}$ is the distance of (x,y) from the origin. Thus $\mathbb{R}^2 = \bigcup_{r \in \mathbb{R}_0^+} A_r$.
- If $r_1 \neq r_2$, then the concentric circles A_{r_1} and A_{r_2} do not intersect. Thus $A_{r_1} \cap A_{r_2} = \emptyset$.

Now define a relation $\sim$ on $\mathbb{R}^2$ via

$$(x,y) \sim (v,w) \iff \exists r \geq 0 \text{ such that } (x,y), (v,w) \text{ both lie on the circle } A_r.$$

By Theorem 7.13 this is an equivalence relation. We can also check explicitly: dropping any mention of the radius r, we see that

$$(x,y) \sim (v,w) \iff x^2 + y^2 = v^2 + w^2.$$

This is exactly the equivalence relation described on page 199. The equivalence classes are precisely the sets A_r. Indeed for a given point (v,w),

$$[(v,w)] = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = v^2 + w^2\} = A_{\sqrt{v^2+w^2}}$$

is just the circle of radius $\sqrt{v^2+w^2}$.

The Möbius Strip Take a rectangle, for example $X = [0,6] \times [0,1]$, and partition into the following subsets.

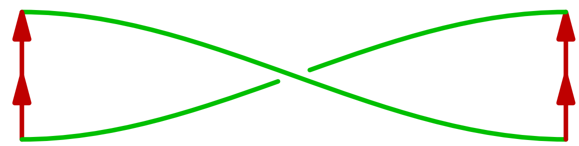
- If a point does not lie on the left or right edge of the rectangle, place it in a subset by itself: $\{(x,y)\}$ for $x \neq 0, 6$,
- If a point does lie on the left or right edge of the rectangle, place it in a subset with one point from the other edge: $\{(0,y), (6, 1-y)\}$ for any y.

The rectangle is drawn below, where the points on the left and right edges are colored red. The arrows indicate how the edges are paired up. For example the point $(0, 0.8)$ (high on the left near the tip of the arrow) is paired with $(6, 0.2)$ (low on the right edge of the rectangle).

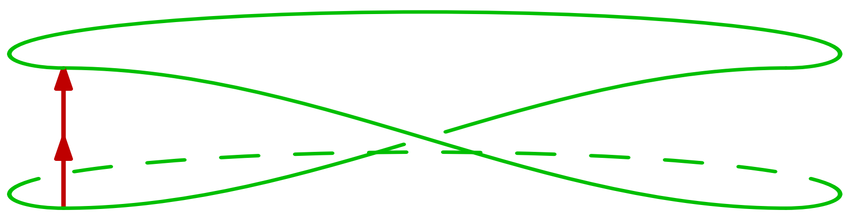
These subsets clearly partition the rectangle X. The partitions define an equivalence relation $\sim$ on X in accordance with Theorem 7.13. Note that there are infinitely many equivalence classes. The question is how we should interpret the quotient set ${}^{X}\!/_{\sim}$?

This is easier to visualize than you might think. Since each point on the left edge of the rectangle lies in an equivalence class with a point on the right edge, we imagine gluing the two edges together in such a way that the corresponding points touch. In the picture, we imagine holding X like a strip of paper, giving it a twist, and then gluing the edges together. This is the classic construction of a Möbius strip. The advantage of the quotient set calculation is that it is very easy to work with points in the original rectangle. As long as you permanently assume that equivalent points of the rectangle correspond to the same point of the Möbius strip you can easily work only in the rectangle.

Rectangle



Half twist



Glue arrows to obtain Möbius strip

The Cylinder We could construct a cylinder similarly to the Möbius strip, by identifying edges of the rectangle but *without* applying the half-twist. Instead we do something a little different.

Let $X = \mathbb{R}^2$ with equivalence relation $\sim$ defined by

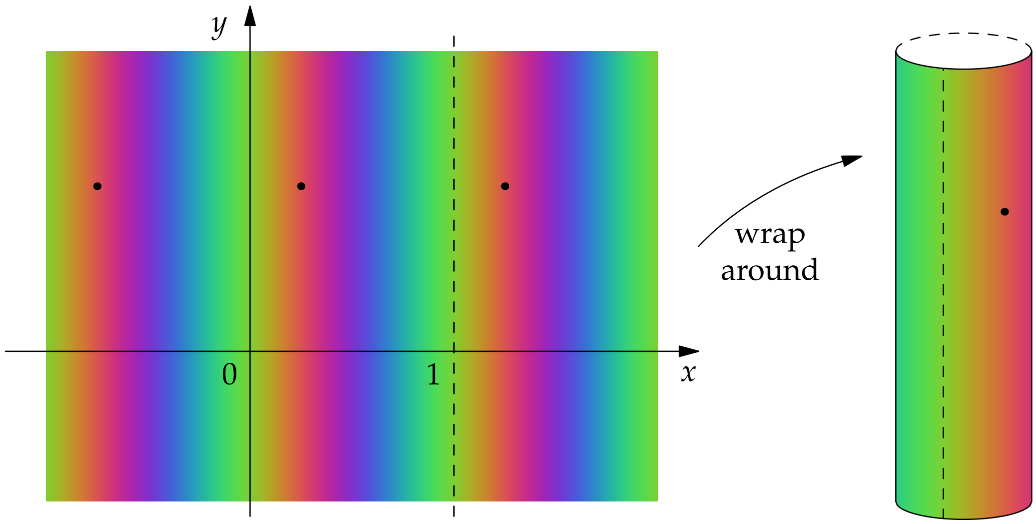
$$(a, b) \sim (c, d) \iff a - c \in \mathbb{Z} \quad \textit{and} \quad b = d.$$

The equivalence classes are horizontal strings of points with the same y co-ordinate. If we imagine wrapping $\mathbb{R}^2$ repeatedly around a cylinder of circumference 1, all of the points in a given equivalence class will now line up. The set of equivalence classes ${}^{\mathbb{R}^2}\!/_{\sim}$ can therefore be visualized as a cylinder.

Alternatively, you may imagine piercing a roll of toilet paper and unrolling it. The single puncture now becomes a row of (almost![23]) equally spaced holes.

In the picture, the left hand side is (part of) the plane $\mathbb{R}^2$, displayed so that points in each equivalence class have the same height and color. The three horizontal dots all lie in the same equivalence class. When we roll up the plane, all three points end up at the same point on the cylinder.

[23]Unfortunately for the analogy, toilet paper has purposeful thickness!



More complex shapes can be created by other partitions/relations. If you want a challenge in visualization, consider why the equivalence relation

$$(a,b) \sim (c,d) \iff a-c \in \mathbb{Z} \quad \textit{and} \quad b-d \in \mathbb{Z}$$

on $\mathbb{R}^2$ defines a torus (the surface of a ring-doughnut).

Reading Quiz

7.4.1 Which of the following statements are true? Select all that apply.

(a) If X is partitioned into the equivalence classes of some equivalence relation $\sim$, then each element of X lies in some equivalence class $[x]$.

(b) Suppose that X is partitioned into subsets and that $x, y, z \in X$. If x, y lie in the same subset, and y, z lie in the same subset of the partition, then it is possible for x and z to lie in different subsets.

(c) $\{\emptyset, \{a\}, \{b, c\}\}$ is a partition of $\{a, b, c\}$.

(d) Every subset in a partition of a set must have the same size.

7.4.2 Which of the following describe the relationship between partitions and equivalence relations? Select all that apply.

(a) Equivalence relations have nothing to do with partitions in general.

(b) For any set X and equivalence relation $\sim$ on X, the quotient set $X/\sim$ is a partition of X.

(c) There exists an infinite set X and a partition $\mathcal{A}$ of X such that for any equivalence relation $\sim$ on X, there is $A \in \mathcal{A}$ for which $A \neq [x]$ for any $x \in X$.

(d) Given any partition $\mathcal{A}$ of X, there is an equivalence relation whose equivalence classes are exactly the subsets of X in $\mathcal{A}$.

Practice Problems

7.4.1 Let X be a nonempty set. Then $\{X\}$ and $\{\{x\} : x \in X\}$ are both partitions of X. For both partitions, determine the equivalence relation whose equivalence classes form the subsets of the partition.

7.4.2 For each of the collections, determine whether the collections partition $\mathbb{R}^2$. Justify your answers, and sketch several of the sets A_n.

(a) $A_n = \{(x,y) \in \mathbb{R}^2 : y = 2x + n\}$, for $n \in \mathbb{Z}$.

(b) $A_n = \{(x,y) \in \mathbb{R}^2 : y = x^2 + n\}$, for $n \in \mathbb{R}$.

(c) $A_n = \{(x,y) \in \mathbb{R}^2 : y = \cos(x - n)\}$, for $n \in \mathbb{R}$.

Exercises

7.4.1 For each of the collections $\{A_n : n \in \mathbb{R}\}$, determine whether the collections partition $\mathbb{R}^2$. Justify your answers, and sketch several of the sets A_n.

(a) $A_n = \{(x,y) \in \mathbb{R}^2 : y = 2x + n\}$.

(b) $A_n = \{(x,y) \in \mathbb{R}^2 : y = (x - n)^2\}$.

(c) $A_n = \{(x,y) \in \mathbb{R}^2 : xy = n\}$.

(d) $A_n = \{(x,y) \in \mathbb{R}^2 : y^4 - y^2 = x - n\}$.

7.4.2 Let X be the set of all humans. If $x \in X$, we define the set

$$A_x = \{\text{people who had the same breakfast } \textit{or} \text{ lunch as } x\}.$$

(a) Does the collection $\{A_x : x \in X\}$ partition X? Explain your answer.

(b) Is your answer different if the *or* in the definition of A_x is changed to *and*?

If Jane and Tom had both had the same breakfast and lunch, then $A_{Jane} = A_{Tom}$ so there are likely many fewer distinct *sets A_x than there are humans!*

7.4.3 Let $X = \{1,2,3\}$. Define the relation $\mathcal{R} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,3)\}$ on X.

(a) Which of the properties reflexive, symmetric, transitive are satisfied by $\mathcal{R}$?

(b) Compute the sets A_1, A_2, A_3 where $A_n = \{x \in X : x \mathcal{R} n\}$. Show that $\{A_1, A_2, A_3\}$ do not form a partition of X.

(c) Repeat parts (a) and (b) for the relations $\mathcal{S}$ and $\mathcal{T}$ on X, where

$$\mathcal{S} = \{(1,1),(1,3),(3,1),(3,3)\}$$
$$\mathcal{T} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,3)\}$$

Some of the sets A_1, A_2, A_3 might be the same in each of your examples. If, for example, $A_1 = A_3$, then the collection $\{A_1, A_2, A_3\}$ only contains two sets: $\{A_1, A_2\}$. Is this a partition? Compare with the example on page 205.

7.4.4 For each of the following, give an example of an infinite set X and an equivalence relation $\sim$ on X such that

(a) $\sim$ has finitely many equivalence classes.

(b) $\sim$ has infinitely many classes, each of which have finitely many elements.

(c) $\sim$ has infinitely many classes, each of which have infinitely many elements.

(d) $\sim$ has a class of size n for each $n \in \mathbb{N}$.

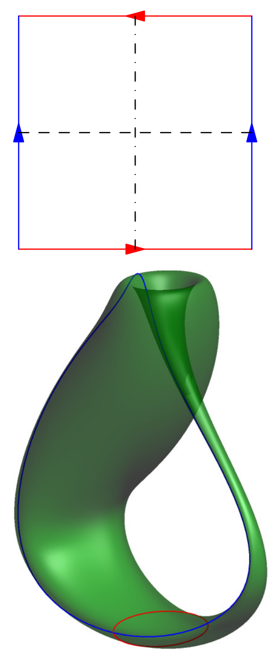
7.4.5 Let A and B be nonempty sets and $f : A \to B$ be a function.

(a) Show that $\{f^{-1}(\{b\}) : b \in \text{range}(f)\}$ forms a partition of A.

(b) Determine the equivalence relation $\sim$ associated to this partition (in the sense of Theorem 7.13).

7.4.6 Using the equivalence relation description of the Möbius strip, prove that you may cut a Möbius strip round the middle and yet still end up with a single loop.
Where would you cut the defining rectangle and how can you tell that you still have one piece?

7.4.7 (Hard!) A *Klein bottle* can be visualized as follows. Define an equivalence relation $\sim$ on the unit square $X = [0,1] \times [0,1]$ so that:



- $(0, y) \sim (1, y)$ for $0 \le y \le 1$.
- $(x, 0) \sim (1 - x, 1)$ for $0 \le x \le 1$.

The result is the picture: the blue edges are identified in the same direction and the red edges in the opposite. Attempting to visualize this in 3D requires a willingness to stretch and distort the square, but results in the green bottle. The original red and blue arrows have become curves on the bottle. If you are using Acrobat Reader, click on the bottle and move it around.

(a) Suppose you cut the Klein bottle along the horizontal dashed line of the defining square. What is the resulting object? What happens to the green bottle?

(b) Now cut the square along the vertical dashed line. What do you get this time?

Can you visualize where the two dashed lines are on the green bottle?

7.5 Well-definition, Rings and Congruence

We return to our discussion of congruence (recall Section 4.1) in the context of equivalence relations and partitions. The important observation is that *congruence modulo n is an equivalence relation on $\mathbb{Z}$*, each equivalence class being the set of all integers sharing a remainder modulo n.

Theorem 7.14. *For a fixed $n \in \mathbb{N}$, define $x \sim_n y \iff x \equiv y \pmod{n}$. Then $\sim_n$ is an equivalence relation on $\mathbb{Z}$.*

The theorem is a restatement of Example 2 on page 205, in conjunction with Theorem 7.13. You should prove this yourself, as practice in using the definition of equivalence relation.
The equivalence classes are precisely those integers which are congruent modulo n: the integers which share the same remainder.

$$\begin{aligned} [a] &= \big\{x \in \mathbb{Z} : x \equiv a \pmod{n}\big\} \\ &= \big\{x \in \mathbb{Z} : x \text{ has the same remainder as } a \text{ when divided by } n\big\} \\ &= \big\{x \in \mathbb{Z} : x - a \text{ is divisible by } n\big\} \end{aligned}$$

In this language, we can restate what it means for two equivalence classes to be equal.

Theorem 7.15. $[a] = [b] \iff a \equiv b \pmod{n} \iff \exists k \in \mathbb{Z}$ *such that* $b = a + kn$.

If the meaning of *any* of the above is unclear, re-read the previous two sections: they are critically important!
The equivalence classes of $\sim_n$ partition the integers $\mathbb{Z}$. According to Theorem 7.15, there are exactly n equivalence classes, whence we may describe the quotient set as

$$\mathbb{Z}/_{\sim_n} = \big\{[0], [1], \ldots, [n-1]\big\}.$$

We use this set to define an extremely important object.

Definition 7.16. Define operations $+_n$ and $\cdot_n$ on the set $\mathbb{Z}/_{\sim_n}$ as follows:

$$[x] +_n [y] := [x + y], \qquad [x] \cdot_n [y] := [x \cdot y].$$

The **ring** $\mathbb{Z}_n$ is the set $\mathbb{Z}/_{\sim_n}$ together with the operations $+_n$ and $\cdot_n$.

The operation $+_n$ is telling us how to add *equivalence classes*, that is, how to produce a new equivalence class from two old ones. It is important to understand that $+_n$ is *not the same* operation as $+$: we are *defining* $+_n$ using $+$. The former combines equivalence classes, while the latter sums integers. The operation $\cdot_n$ similarly tells us how to multiply equivalence classes. The challenge here is that you have to think of each equivalence class as a single object.

Example. When we write

$$[3] +_8 [6] = [3+6] = [9] = [1],$$

we are thinking about the equivalence classes [3] and [6] as individual objects rather than as collections of elements: remember that $[3] = \{\ldots, -5, 3, 11, 19, \ldots\}$ is an infinite set! There is, moreover, a matter of choice: since, for example, $[3] = [11]$ and $[6] = [22]$ we should be able to observe that

$$[3] +_8 [6] = [11] +_8 [22].$$

Is this true? If not, then the operation $+_8$ would not be particularly useful. Thankfully this is not a problem: according to the definition of $+_8$, we have

$$[11] +_8 [22] = [11+22] = [33] = [1],$$

exactly as we would wish.

Let us think a little more abstractly. Suppose we are given equivalence classes X and Y, how do we compute $X +_n Y$? Here is the process.

7.5.1 *Choose* elements $x \in X$ and $y \in Y$ so that $X = [x]$ and $Y = [y]$.

7.5.2 Add x and y to get a new element $x + y \in \mathbb{Z}$.

7.5.3 Then $X +_n Y$ is the equivalence class $[x + y]$.

The issue is that there are *infinitely many possibilities* for the elements $x \in X$ and $y \in Y$ chosen at step 1. If $+_n$ is to make sense, we must obtain the *same* equivalence class $[x + y]$ **regardless of our choices of $x \in X$ and $y \in Y$.**

Definition 7.17. A concept is **well-defined** if it is **independent of all choices used in the definition.**

Theorem 7.18. *The operations $+_n$ and $\cdot_n$ are well-defined.*

The choices made in the definitions of $+_n$ and $\cdot_n$ were of representative elements x and y of the equivalence classes $[x]$ and $[y]$. All representatives of these classes have the form

$$x + kn \in [x] \quad \text{and} \quad y + ln \in [y]$$

for some integers k, l. It therefore suffices to prove that

$$\forall k, l \in \mathbb{Z}, \quad [x + kn] +_n [y + ln] = [x] +_n [y] \quad \text{and} \quad [x + kn] \cdot_n [y + ln] = [x] \cdot_n [y].$$

We are now in a position to prove the Theorem.

Proof. We prove that $+_n$ is well-defined.

$$\begin{aligned}
[x+kn] +_n [y+ln] &= [(x+kn)+(y+ln)] && \text{(by definition of } +_n) \\
&= [x+y+(k+l)n] \\
&= [x+y] && \text{(by Theorem 7.15)} \\
&= [x] +_n [y] && \text{(by definition of } +_n)
\end{aligned}$$

The argument for $\cdot_n$ is similar. ∎

You should re-read Theorem 4.8 until you are comfortable that we are doing the same thing!

Aside. Aside: Ugly notation

Given the usefulness of $\mathbb{Z}_n$ and the cumbersome nature of the above notation, it is customary to drop the square brackets and subscripts and simply write

$$\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}, \qquad x + y := x + y \pmod n, \qquad x \cdot y := xy \pmod n.$$

When using this description of $\mathbb{Z}_n$, you should realize that we are working with equivalence classes, not numbers. In this context, $-3 \in \mathbb{Z}_8$ makes perfect sense, for it really means $[-3] \in \mathbb{Z}_8$. This is perfectly fine, since $[-3] = [5]$ as equivalence classes, whence it is legitimate to write $-3 = 5$ in $\mathbb{Z}_8$. Until you are 100% sure that you know when 3 represents an equivalence class and when it represents a number, you should keep the brackets in place: in particular it might be a good idea to keep using them until you have passed *this course*!

Reading Quiz

7.5.1 Which of the following are true and which false?

(a) $[28] = [5]$ in $\mathbb{Z}_6$.
(b) $[24] + \big([3] + [17]\big) = [-10]$ in $\mathbb{Z}_9$.
(c) $[2]^3 + [3]^3 = [4]^3$ in $\mathbb{Z}_{29}$.

7.5.2 Is the following True or False?

$$[x] + [y] = [z] \iff x + y = z.$$

Practice Problems

7.5.1 Suppose $\gcd(a, n) = 1$. Show that there exists b such that $[a] \cdot [b] = [1]$ in $\mathbb{Z}_n$.

7.5.2 (a) Show that $\gcd(a, n) = 1$ if and only if there exist m and k such that $ma + kn = 1$.
(b) Use part (a) to prove that if there is b such that $[a] \cdot [b] = [1]$ in $\mathbb{Z}_n$, then $\gcd(a, n) = 1$.

Exercises

7.5.1 (a) Explicitly check that $[7] + [21] = [98] + [-5]$ in $\mathbb{Z}_{13}$.

(b) Suppose that $[5] \cdot [7] = [8] \cdot [9]$ makes sense. Find the value of n if we are working in the ring $\mathbb{Z}_n$.

7.5.2 (a) Prove the second half of Theorem 7.18, that $\cdot_n$ is well-defined.

(b) Prove by induction that the operation of raising to the power $m \in \mathbb{N}$ is well-defined in $\mathbb{Z}_n$. I.e., prove that

$$\forall m \in \mathbb{N},\ \forall [x] \in {}^{\mathbb{Z}}\!/_{\sim_n} \text{ we have } [x^m] = [x]^m.$$

Be careful! n is fixed, your induction variable is m. What base case(s) do you need?

7.5.3 Suppose that p is prime and that in $\mathbb{Z}_p$, we have $[a] \neq [0]$. Show $[a]^2 \neq [0]$. [Hint: See Exercise 5 in Section 5.4.]

7.5.4 Give an explicit proof of Theorem 7.14.

7.5.5 Consider the relation $\sim$ defined on $\mathbb{Z} \times \mathbb{N} = \{(x, y) : x \in \mathbb{Z}, \text{ and } y \in \mathbb{N}\}$ by

$$(a, b) \sim (c, d) \iff ad = bc.$$

(a) Prove that $\sim$ is an equivalence relation.

(b) List several elements of the equivalence class of $(2, 3)$. Repeat for the equivalence class of $(-3, 7)$. What do the equivalence classes have to do with the set of rational numbers $\mathbb{Q}$?

(c) Define operations $\oplus$ and $\otimes$ on ${}^{\mathbb{Z} \times \mathbb{N}}\!/_{\sim}$ by

$$[(a, b)] \oplus [(c, d)] = [(ad + bc, bd)], \qquad [(a, b)] \otimes [(c, d)] = [(ac, bd)].$$

Prove that $\oplus$ and $\otimes$ are well-defined.

Try to do this question without *using division! We will return to this example in the next section.*

7.6 Functions and Partitions

To complete our discussion of partitions and equivalence relations, we consider how to define a function whose domain is a set of equivalence classes. We take congruence as our motivating example.

Suppose we want to define a function $f : \mathbb{Z}_4 \to \mathbb{Z}_6$. Say $f(x) = 3x \pmod 6$. This certainly looks like a function, but is it? Remember that 'x' and '$3x$' are really equivalence classes, so we should say[24]

$$f\big([x]_4\big) = [3x]_6, \qquad \text{where} \quad [x]_4 \in \mathbb{Z}_4 \quad \text{and} \quad [3x]_6 \in \mathbb{Z}_6.$$

Is *this* a function? To make sure, we need to check that *any* representative $a \in [x]_4$ gives the same result. That is, we need to prove that

$$a \equiv b \pmod 4 \implies 3a \equiv 3b \pmod 6.$$

This is not so hard:

$$\begin{aligned} a \equiv b \pmod 4 &\implies \exists n \in \mathbb{Z} \text{ such that } a = b + 4n \\ &\implies 3a = 3b + 12n \implies 3a \equiv 3b \pmod 6. \end{aligned}$$

It might appear to be a minor difference, but attempting to define $g : \mathbb{Z}_4 \to \mathbb{Z}_6$ by $g(x) = 2x \pmod 6$ does *not* result in a function. If it were, then we should have

$$a \equiv b \pmod 4 \implies 2a \equiv 2b \pmod 6.$$

But this is simply not true: for example $4 \equiv 0 \pmod 4$, but $8 \not\equiv 0 \pmod 6$. It might look like g is a function, but it is not well-defined because $[4] = [0]$ in $\mathbb{Z}_4$ and $g\big([4]\big) \neq g\big([0]\big)$ in $\mathbb{Z}_6$.

Just as in Definition 7.17, the process of verifying that a rule really is a function is called checking *well-definition.* In general, if we are defining a function

$$f : {}^{X}\!/_{\sim} \to A$$

whose domain is a quotient set, then it is usually necessary to construct f by saying what happens to a *representative* x of an equivalence class $[x]$:

$$f\big([x]\big) = \text{'do something to } x\text{'}. \tag{$*$}$$

We need to make sure that the 'something' is *independent of the choice of element x*.

Definition 7.19. Suppose that $f : {}^{X}\!/_{\sim} \to A$ is a rule of the form $(*)$. We say that f is a **well-defined** function if

$$[x] = [y] \implies f\big([x]\big) = f\big([y]\big).$$

If you think carefully, this is nothing more than condition 2. of Definition 7.4.

[24] The notation $[x]_4$ is helpful for reminding us which equivalence relation is being applied. When dealing with functions between different quotient sets, it is easy to become confused.

Examples. 7.6.1 Show that $f : \mathbb{Z}_n \to \mathbb{Z}_n$ defined by $f([x]) = [x^2 + 4]$ is well-defined.

We must check that $x \equiv y \pmod n \implies x^2 + 4 \equiv y^2 + 4 \pmod n$. But this is trivial!

7.6.2 For which integers k is the rule $f_k : \mathbb{Z}_4 \to \mathbb{Z}_6$ defined by $f_k([x]_4) = [kx]_6$ a well-defined function?

We start with a special case. If $k = 1$, then we can attempt to construct a table of values for $f_1([x]_4)$:

$[x]_4$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$	$[4]_4$	$[5]_4$	$[6]_4$	$[7]_4$	$[8]_4$	$\cdots$
$f_1([x]_4)$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$	$[0]_6$	$[1]_6$	$[2]_6$	$\cdots$

The problem is immediately visible! In $\mathbb{Z}_4$ we have $[0]_4 = [4]_4$, however $f_1([0]_4) = [0]_6$ and $f_1([4]_4) = [4]_6$ *which are not equal in* $\mathbb{Z}_6$! It follows that f_1 is not a function.

Rather than try out all possible values of k, we proceed systematically. If f_k is to be well-defined, we require $x \equiv y \pmod 4 \implies kx \equiv ky \pmod 6$. Now

$$\begin{aligned} x \equiv y \pmod 4 &\implies \exists n \in \mathbb{Z} \text{ such that } x - y = 4n \\ &\implies kx - ky = 4kn. \end{aligned}$$

For f_k to be well-defined, we need to see that $k(x - y) = 4kn$ is a multiple of 6 *independently* of x and y. Thus f_k is well-defined if and only if $6 \mid 4kn$ *for all* $n \in \mathbb{Z}$. This is the case if and only if $6 \mid 4k$. Otherwise said,

$$f_k \text{ is well-defined} \iff 6 \mid 4k \iff 3 \mid 2k \iff 3 \mid k.$$

Given that we want $[kx]_6 \in \mathbb{Z}_6$, we need only consider $k \in \{0, 1, 2, 3, 4, 5\}$: equivalent values of k modulo 6 won't change the definition of f_k. It follows that there are only *two* well-defined functions $f_k : \mathbb{Z}_4 \to \mathbb{Z}_6 : x \mapsto kx$, namely $f_0([x]_4) = [0]_6$ and $f_3([x]_4) = [3x]_6$. Here they are in tabular form (dropping the brackets):

x	0	1	2	3	4	5	6	7	8	$\cdots$
$f_0(x)$	0	0	0	0	0	0	0	0	0	$\cdots$
$f_3(x)$	0	3	0	3	0	3	0	3	0	$\cdots$

It should be clear that well-defined functions f_k produce tables whose $f_k(x)$ line is *periodic with period four.* To ram this point home, here is the table when $k = 5$:

x	0	1	2	3	4	5	6	7	8	$\cdots$
$f_5(x)$	0	5	4	3	2	1	0	5	4	$\cdots$

This is palpably not a function! You should compare these examples with those on page 96 and with Exercise 3.4.**??**. Are these earlier example still functions when the domains are assumed to be a ring $\mathbb{Z}_n$ rather than simply a set of integers?

Functions on the Cylinder and Torus

Recall our construction on page 209, where we viewed the cylinder as the set ${}^{\mathbb{R}^2}/_{\sim}$ with respect to the equivalence relation

$$(a,b) \sim (c,d) \iff a-c \in \mathbb{Z} \quad \text{and} \quad b=d.$$

We wish to define a function f whose domain is a cylinder. Using the equivalence relation, this is the same as defining a function $f : {}^{\mathbb{R}^2}/_{\sim} \to A$ where A is our chosen codomain. *Well-definition* requires that f satisfy

$$(a,b) \sim (c,d) \implies f\Big(\big[(a,b)\big]\Big) = f\Big(\big[(c,d)\big]\Big).$$

Since $(a,b) \sim (a+1,b)$, we require $f\Big(\big[(a,b)\big]\Big) = f\Big(\big[(a+1,b)\big]\Big)$, for all $a,b \in \mathbb{R}$. Otherwise said, $f\Big(\big[(x,y)\big]\Big)$ must be periodic in x with period one. It is easy to see that

$$f\Big(\big[(x,y)\big]\Big) = y^2 \sin(2\pi x)$$

is a suitable choice of function $f : {}^{\mathbb{R}^2}/_{\sim} \to \mathbb{R}$.

More generally, to define a function whose domain is the torus

$$T^2 = {}^{\mathbb{R}^2}/_{\sim} \quad \text{where} \quad (a,b) \sim (c,d) \iff a-c \in \mathbb{Z} \quad \text{and} \quad b-d \in \mathbb{Z},$$

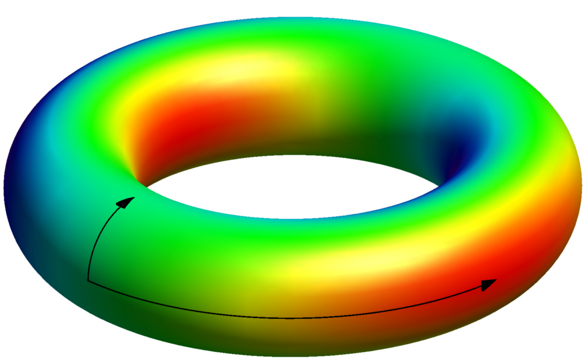
requires a function which is periodic in *both* x and y. The function

$$f\Big(\big[(x,y)\big]\Big) = \sin(2\pi x)\cos(2\pi y)$$

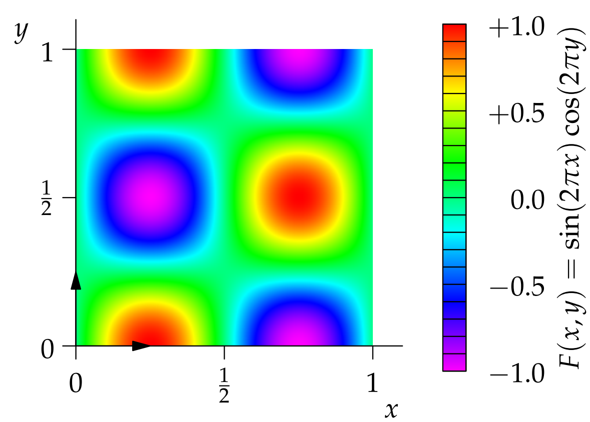
is plotted below, with the color on the torus indicating the value of f. It is easier to instead consider the function

$$F : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto \sin(2\pi x)\cos(2\pi y).$$

This is also plotted, with the same color for each value. The point is that F is really f in disguise, but has the advantage of being much easier to work with.



The function f: domain T^2
The arrows in the two pictures correspond



The function F restricted to $[0,1) \times [0,1)$

The Canonical Map

To do this justice, and to give you a taste for the details which are necessary in pure mathematics, here is the important definition.

Definition 7.20. Suppose that $\sim$ is an equivalence relation on a set X. The function $\gamma : X \to {}^{X}\!/_{\sim}$ defined by $\gamma(x) = [x]$ is the **canonical map**[a].

[a]Canonical, in mathematics, just means natural or obvious.

For us, the purpose of the canonical map is to allow us to construct functions $f : {}^{X}\!/_{\sim} \to A$.

Theorem 7.21. *Suppose that $\sim$ is an equivalence relation on X.*

7.6.1 *If $f : {}^{X}\!/_{\sim} \to A$ is a function, then $F : X \to A$ defined by $F = f \circ \gamma$ satisfies*

$$x \sim y \implies F(x) = F(y).$$

7.6.2 *If $F : X \to A$ satisfies $x \sim y \implies F(x) = F(y)$, then there is a unique function $f : {}^{X}\!/_{\sim} \to A$ satisfying $F = f \circ \gamma$.*

Proof. 7.6.1 This is trivial:
$$\begin{aligned} x \sim y &\implies [x] = [y] \implies \gamma(x) = \gamma(y) \\ &\implies f(\gamma(x)) = f(\gamma(y)) \implies F(x) = F(y). \end{aligned}$$

7.6.2 $f : {}^{X}\!/_{\sim} \to A$ can only be the function *defined* by $f([x]) = F(x)$. We show that this is well-defined:

$$[x] = [y] \implies x \sim y \implies F(x) = F(y) \implies f([x]) = f([y]).$$ ■

The proof, like much of mathematics, is a masterpiece in concision that seems to be doing nothing at all. The point is that functions of the form $f : {}^{X}\!/_{\sim} \to A$ are *difficult* to work with. The Theorem says that we never need to explicitly use such functions, and can instead work with *simpler* functions of the form $F : X \to A$. The only condition is that we must have $x \sim y \implies F(x) = F(y)$. Essentially, F is f in disguise!

$$\begin{tikzcd} X \arrow[rr, "F"] \arrow[dr, "\gamma"'] & & A \\ & {}^{X}\!/_{\sim} \arrow[ur, "f"'] & \end{tikzcd}$$

This result will be resurrected when you study Groups, Rings & Fields as part of the famous *First Isomorphism Theorem.*

Reading Quiz

7.6.1 Let X be a set and $\sim$ an equivalence relation on X. What does it mean for a function $f : X/\sim \to B$ to be well-defined?

(a) It means f is an injection.

(b) It means $[x] = [y]$ if and only if $f(x) = f(y)$.

(c) It means f is a surjection.

(d) It means that $x \sim y$ implies $f(x) = f(y)$.

7.6.2 True or False: the rule $[x] \mapsto x : \mathbb{Z}_n \to \mathbb{Z}$ is not well-defined.

Practice Problems

7.6.1 Let k be a constant integer. If $f : \mathbb{Z}_5 \to \mathbb{Z}_{18} : [x]_5 \mapsto [kx]_{18}$ is a well-defined function, what period must the sequence of values $f([0]_5), f([1]_5), f([2]_5), \ldots$ have?

7.6.2 In Theorem 7.21 show that F is a surjection if and only if f is a bijection.

Exercises

7.6.1 (a) Prove or disprove: $f : \mathbb{Z}_3 \to \mathbb{Z}_5 : x \mapsto x^3 \pmod 5$ is well-defined.

(b) Prove or disprove: $f : \mathbb{Z}_{10} \to \mathbb{Z}_{20} : x \mapsto x^2 \pmod{20}$ is well-defined.

7.6.2 Determine whether the following are well-defined.

(a) Define $f : \mathbb{Z}_8 \to \mathbb{Z}_{16}$ by $f([a]_8) = [2a]_{16}$.

(b) Define $g : \mathbb{Z}_8 \to \mathbb{Z}_{12}$ by $g([a]_8) = [2a]_{12}$.

7.6.3 Can we view $F(x, y) = (y^2 - 1)\sin^2(\pi x)$ as a function whose domain is the cylinder, as described on page 219? Explain your answer.

7.6.4 (a) Compute $(x + 4n)^2$.

(b) Suppose that $\forall n \in \mathbb{Z}$, we have $(x + 4n)^2 \equiv x^2 \pmod m$. Find all the integers m for which this is a true statement.

(c) For what $m \in \mathbb{N}_{\geq 2}$ is the function $f : \mathbb{Z}_4 \to \mathbb{Z}_m : x \mapsto x^2 \pmod m$ well-defined.

7.6.5 A rule $f : {}^{X}\!/_{\sim} \to A$ is well-defined if $[x] = [y] \implies f([x]) = f([y])$.

(a) State what it means for $f : {}^{X}\!/_{\sim} \to A$ to be *injective*. What do you observe?

(b) Prove that $f : \mathbb{Z}_7 \to \mathbb{Z}_{35} : x \mapsto 15x$ is a well-defined, injective function.

(c) Repeat part (b) for the function $f : \mathbb{Z}_{100} \to \mathbb{Z}_{300} : x \mapsto 9x$. Compare your arguments for well-definition and injectivity.
This forces you to write your argument abstractly, rather than using a table! You may find it useful that $9 \cdot (-11) \equiv 1 \pmod{100}$.

7.6.6 Define a partition of the sphere $S^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$ into subsets of the form

$$\{(x, y, z), (-x, -y, -z)\}.$$

Each subset consists of two points directly opposite each other on the sphere (antipodal points). Let $\sim$ be the equivalence relation whose equivalence classes are the above subsets.

(a) $f : {}^{S^2}/_{\sim} \to \mathbb{R} : [(x,y,z)] \mapsto xyz$ is not well-defined. Explain why.

(b) Prove that $f : {}^{S^2}/_{\sim} \to \mathbb{R}^3 : [(x,y,z)] \to (yz, xz, xy)$ is a well-defined function. *The image of this function is Steiner's famous Roman Surface, another example, like the Klein Bottle, of a generalization of the Möbius Strip.*

7.6.7 Recall Exercise 7.5.5, where we defined an equivalence relation $\sim$ on $\mathbb{Z} \times \mathbb{N}$.

(a) Prove that the function $f : {}^{\mathbb{Z} \times \mathbb{N}}/_{\sim} \to \mathbb{Q}$ defined by $f\Big([(x,y)]\Big) = \frac{x}{y}$ is a *well-defined bijection.*

(b) Prove that f transforms the operations $\oplus$ and $\otimes$ into the usual addition and multiplication of rational numbers. That is:

$$f\Big([(a,b)] \oplus [(c,d)]\Big) = f\Big([(a,b)]\Big) + f\Big([(c,d)]\Big)$$
$$f\Big([(a,b)] \otimes [(c,d)]\Big) = f\Big([(a,b)]\Big) \cdot f\Big([(c,d)]\Big)$$

The technical term for this is that $f : \Big({}^{\mathbb{Z} \times \mathbb{N}}/_{\sim}, \oplus, \otimes\Big) \to (\mathbb{Q}, +, \cdot)$ *is an isomorphism of rings.*

7.6.8 Let $\mathcal{C} = \{\text{circles centered at the origin}\} \cup \{(0,0)\}$.

(a) Find an equivalence relation $\sim$ on $\mathbb{R}^2$ such that $\mathbb{R}^2/\sim = \mathcal{C}$.

(b) Let $f : \mathcal{C} \to \mathbb{R}$ by $f(C) = $ radius of C. Theorem 7.21 says there is a function $F : \mathbb{R}^2 \to \mathbb{R}$ such that $F = f \circ \gamma$ where γ is the canonical map. Determine an explicit formula for F.

7.6.9 Define $\sim$ on $\mathbb{R}$ by $x \sim y$ if and only if $x - y \in \mathbb{Z}$.

(a) Show $\sim$ is an equivalence relation on $\mathbb{R}$.

(b) Find a function $F : \mathbb{R} \to [0,1)$ such that $x \sim y$ implies $F(x) = F(y)$.

(c) Use part (b) to find a well-defined function $f : \mathbb{R}/\sim \to [0,1)$. Is f a bijection?

7.6.10 Every function can be factored as a composition of a surjection followed by a bijection followed by an injection. Let $f : A \to B$ be a function. We show that $f = i \circ g \circ \gamma$ where i is an injection, g a bijection, and γ a surjection.

(a) Define $\sim$ on A by $a \sim b$ if and only if $f(a) = f(b)$. Show $\sim$ is an equivalence relation on A.

(b) Let $\gamma : A \to A/\sim$ be the canonical projection. Show there is a well-defined function $g : A/\sim \to \text{range}(f)$ such that $f^* = g \circ \gamma$ where $f^* : A \to \text{range}(f)$ is the function $f^*(a) = f(a)$.

(c) Show that g is a bijection.

(d) Let $i : \text{range}(f) \to B$ be the inclusion map $i(b) = b$. Show $f = i \circ f^*$, and conclude that $f = i \circ g \circ \gamma$.

8 Cardinalities of Infinite Sets

8.1 Cantor's Notion of Cardinality

During the late 1800's a German mathematician named Georg Cantor almost single-handedly overturned the foundations of mathematics. Prior to Cantor, mathematicians had understood a set to be nothing more than a collection of objects. Via the consideration of certain infinite sets (in particular his middle third set), Cantor showed this naïve idea to be woefully inadequate. Cantor met great resistance from many famous mathematicians and philosophers who felt his ideas to be unnatural. He even managed to inflame several religious scholars who believed his investigation of infinity to be an affront to the divine! Despite strong initial antipathy, Cantor's notion of cardinality is now universally accepted by mathematicians. More importantly, by exposing the contradictions inherent in contemporary set theory, he convinced mathematicians that a rigorous axiomatic approach was necessary. The result was a revolution in foundational mathematics, now known as *axiomatic set theory.* Indeed, Cantor's legacy is arguably the modern axiomatic nature of pure mathematics, where rigor dominates and mathematicians are obliged to follow logic wherever it leads, regardless of the bizarre paradoxes which might appear.

In this chapter we consider the basics of Cantor's contribution, essentially his extension of the concept of *cardinality* to infinite sets.

Recall that if A is a *finite* set, then $|A|$, the cardinality of A, is simply the number of elements in A. This definition obviously does not extend to infinite sets. However, cardinality has a stronger purpose than merely attaching a number to each set: it can be viewed as a *relation* and used to *compare* sets. It is this interpretation that turns out to apply to infinite sets. For example, suppose that

$$A = \{\text{fish}, \text{dog}\}, \quad \text{and} \quad B = \{\alpha, \beta, \gamma\}.$$

Even though the elements of the sets A and B are completely different, we may use cardinality to compare the sizes of A and B: since $|A| = 2$ and $|B| = 3$, we may write $|A| < |B|$ to indicate that B has more elements as A: colloquially, "B is larger than A."

It is at this point that Cantor enters the discussion. By Theorem 3.15 and Corollary 3.16, the condition $|A| < |B|$ is equivalent to the existence of an injective (one-to-one) function $f : A \to B$ and the non-existence of a bijection $g : A \to B$. For example, the function $f : A \to B$ defined by

$$\text{fish} \longmapsto \alpha, \qquad \text{dog} \longmapsto \beta,$$

is clearly injective. In a sense, Theorem 3.15 tells us how to compare the cardinalities of finite sets *without* counting their elements. Cantor's seemingly innocuous idea was to turn this *theorem* for finite sets into a *definition* of cardinality for all sets.

Definition 8.1. The **cardinalities** of two sets A, B are denoted $|A|$ and $|B|$. We compare cardinalities as follows:

- $|A| \leq |B| \iff \exists f : A \to B$ injective.
- $|A| = |B| \iff \exists f : A \to B$ bijective.

We write $|A| < |B| \iff |A| \leq |B|$ and $|A| \neq |B|$. That is $\exists f : A \to B$ injective but $\nexists g : A \to B$ bijective.

Cardinality is defined as an abstract *property* whereby two sets can be *compared.* Otherwise said, it is a *relation.* To define a cardinality $|A|$ as an object, we need the following theorem.

Theorem 8.2. *On any collection of sets, the relation $A \sim B \iff |A| = |B|$ is an equivalence relation.*

The cardinality of a set A can then be defined to be the equivalence class of A with respect to this relation: $|A| := [A]$. It is now clear that cardinality partitions any collection of sets: every set has a cardinality, and no set has more than one cardinality. We can moreover identify the cardinalities of finite sets with the cardinal numbers $0, 1, 2, 3, 4, \ldots$ in a natural way. To get further it is useful to introduce a symbol for the cardinality of the simplest infinite set.

Countably Infinite Sets

Definition 8.3. The cardinality of the set of natural numbers $\mathbb{N}$ is denoted $\aleph_0$, read *aleph-nought* or *aleph-null.* We say that a set A is **countably infinite,** or **denumerable**[a] if $|A| = \aleph_0$.

[a] Sometimes this is shortened to **countable,** although some authors use countable to mean 'finite or denumerable,' i.e. any A for which $|A| \leq \aleph_0$. Use **countably infinite** or **denumerable** to avoid confusion. $\aleph$ is the first letter of the Hebrew alphabet.

We will discuss in a moment why we need a new symbol, why ∞ doesn't suffice. First we consider an example of Definition 8.1 at work.

Example. Let $2\mathbb{N} = \{2, 4, 6, 8, 10, \ldots\}$ be the set of positive even integers. The function

$$f : \mathbb{N} \to 2\mathbb{N} : n \mapsto 2n$$

is a bijection. It follows that $|2\mathbb{N}| = |\mathbb{N}| = \aleph_0$ and we say that $2\mathbb{N}$ is countably infinite.

This example immediately demonstrates one of strange properties of infinite sets: $2\mathbb{N}$ is a *proper subset* of $\mathbb{N}$, and yet the two sets are in bijective correspondence with one another! You should feel like you want to say two contradictory things simultaneously:

- $\mathbb{N}$ has the same 'number of elements' as $2\mathbb{N}$.
- $\mathbb{N}$ has twice the 'number of elements' as $2\mathbb{N}$.

If this doesn't make you feel uncomfortable, then read it again! The remedy to your discomfort is to appreciate that *cardinality* and *number of elements* are different concepts. Replacing 'number of elements' with 'cardinality' in the two statements makes both true! Indeed it is completely legitimate to write $2\aleph_0 = \aleph_0$. The idea of a set having a proper subset with the same cardinality can be used as a *definition* of infinite set (see Exercise 8.1.18).

Here is another example of the same phenomenon; $\mathbb{N}$ has one more element than $\mathbb{N}_{\geq 2}$ and yet they have the same cardinality: $\aleph_0 + 1 = \aleph_0$.

Example. The function $g : \mathbb{N} \to \mathbb{N}_{\geq 2} : n \mapsto n+1$ is a bijection, whence $\mathbb{N}_{\geq 2} = \{2,3,4,5,\ldots\}$ is countably infinite.

Proving that a set is countably infinite While it is possible to use any number of clever theorems to prove the denumerability of a set A, the simplest thing to imagine listing the elements in some order so that A 'looks like' the natural numbers, or some other known countably infinite set. For instance, the above examples can be summarized by listing the elements of these sets below those of the natural numbers:

$\mathbb{N}$	1	2	3	4	5	6	7	8	9	10	$\cdots$
$2\mathbb{N}$	2	4	6	8	10	12	14	16	18	20	$\cdots$
$\mathbb{N}_{\geq 2}$	2	3	4	5	6	7	8	9	10	11	$\cdots$

The required bijective functions are then easy to read off! We use this technique to construct bijections which show the denumerability of two important examples.

Theorem 8.4. *The integers $\mathbb{Z}$ are countably infinite.*

Proof. We must construct a bijective function $f : \mathbb{N} \to \mathbb{Z}$. By experimenting with listing the integers, we write down the first few terms of a suitable function in tabular form:

n	1	2	3	4	5	6	7	8	9	10	$\cdots$
$f(n)$	0	1	-1	2	-2	3	-3	4	-4	5	$\cdots$

Two things should be clear from the table:

Surjectivity Every integer appears at least once in the second row.

Injectivity No integer appears more than once in the second row.

It follows that the function f is bijective. ■

You might object that the above argument is too quick, and perhaps you don't trust the reasoning. Does the table really define a function? Is it really obvious that the function is bijective? We can be more formal and explicit, but the cost is that the big picture becomes less clear. Our function may be written

$$f(n) = \begin{cases} \frac{1}{2}n & \text{if } n \text{ is even,} \\ -\frac{1}{2}(n-1) & \text{if } n \text{ is odd.} \end{cases}$$

Now we check that this is bijective:

(*Injectivity*) Let $m, n \in \mathbb{N}$, and suppose that $f(m) = f(n)$. Without loss of generality, there are three cases to consider.

(m, n *both even*) $f(m) = f(n) \implies \frac{m}{2} = \frac{n}{2} \implies m = n$.

(m, n *both odd*) $f(m) = f(n) \implies -\frac{1}{2}(m-1) = -\frac{1}{2}(n-1) \implies m = n$.

(m *even*, n *odd*) $f(m) = f(n) \implies \frac{m}{2} = -\frac{1}{2}(n-1) \implies m + n = 1$. But $m, n \in \mathbb{N}$, so $m + n \geq 2$, which is a contradiction.

Therefore f is injective.

(*Surjectivity*) With a little calculation, you should be able to see that, for any $z \in \mathbb{Z}$, there exists a positive integer n such that $f(n) = z$, namely:

$$z = \begin{cases} f(2z) & \text{if } z > 0, \\ f(1-2z) & \text{if } z \leq 0. \end{cases}$$

Hence f is surjective.

For basic examples you are encouraged to use the listing/pictorial construction rather than explicitly writing everything out. Training your intuition is more important than the formality here! Indeed we would likely have been unable to come up with an explicit formula for f without the table, and it is easier to get a feel for what f is using the table rather than the formula.

As you build up examples, you no longer have to compare countably infinite sets directly to the natural numbers. A set B is countably infinite if and only if there exists a bijection $f : A \to B$ where A is *any countably infinite set*. This holds because the composition of bijective function is also bijective (Theorem 3.18). For instance, we immediately see that the set of even integers $2\mathbb{Z}$ is countably infinite because

$$f : \mathbb{Z} \to 2\mathbb{Z} : z \mapsto 2z$$

is a bijection, and because we now know that $\mathbb{Z}$ is countably infinite. We use this approach to help prove the following result, the first of Cantor's truly counter-intuitive revelations.

Theorem 8.5. *The rational numbers $\mathbb{Q}$ are countably infinite.*

We prove the Theorem in stages. First we construct a bijection between the natural numbers $\mathbb{N}$ and the positive rational numbers $\mathbb{Q}^+$. We then modify this to obtain a bijection between the integers $\mathbb{Z}$ and the full set of rational numbers $\mathbb{Q}$. By the previous Theorem, it follows that $\mathbb{Q}$ must be countably infinite.

Proof. For each pair of natural numbers a, b, place the fraction $\frac{a}{b} \in \mathbb{Q}^+$ in the ath column and bth row of an infinite square as shown below. Now list the positive rational numbers by tracing the diagonals as shown, deleting any number that has already appeared in the list ($\frac{2}{2} = \frac{1}{1}$, $\frac{6}{4} = \frac{3}{2}$, etc.).

$$\begin{array}{cccccccc}
\frac{1}{1} & \frac{2}{1} & \frac{3}{1} & \frac{4}{1} & \frac{5}{1} & \frac{6}{1} & \frac{7}{1} & \cdots \\
\frac{1}{2} & \frac{2}{2} & \frac{3}{2} & \frac{4}{2} & \frac{5}{2} & \frac{6}{2} & \frac{7}{2} & \cdots \\
\frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \frac{4}{3} & \frac{5}{3} & \frac{6}{3} & \frac{7}{3} & \cdots \\
\frac{1}{4} & \frac{2}{4} & \frac{3}{4} & \frac{4}{4} & \frac{5}{4} & \frac{6}{4} & \frac{7}{4} & \cdots \\
\frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & \frac{5}{5} & \frac{6}{5} & \frac{7}{5} & \cdots \\
\frac{1}{6} & \frac{2}{6} & \frac{3}{6} & \frac{4}{6} & \frac{5}{6} & \frac{6}{6} & \frac{7}{6} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

The infinite square

$$\begin{array}{cccccccc}
\frac{1}{1} & \frac{2}{1} & \frac{3}{1} & \frac{4}{1} & \frac{5}{1} & \frac{6}{1} & \frac{7}{1} & \cdots \\
\frac{1}{2} & \cancel{\frac{2}{2}} & \frac{3}{2} & \cancel{\frac{4}{2}} & \frac{5}{2} & \cancel{\frac{6}{2}} & \frac{7}{2} & \cdots \\
\frac{1}{3} & \frac{2}{3} & \cancel{\frac{3}{3}} & \frac{4}{3} & \frac{5}{3} & \cancel{\frac{6}{3}} & \frac{7}{3} & \cdots \\
\frac{1}{4} & \cancel{\frac{2}{4}} & \frac{3}{4} & \cancel{\frac{4}{4}} & \frac{5}{4} & \cancel{\frac{6}{4}} & \frac{7}{4} & \cdots \\
\frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & \cancel{\frac{5}{5}} & \frac{6}{5} & \frac{7}{5} & \cdots \\
\frac{1}{6} & \cancel{\frac{2}{6}} & \cancel{\frac{3}{6}} & \cancel{\frac{4}{6}} & \frac{5}{6} & \cancel{\frac{6}{6}} & \frac{7}{6} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

Trace diagonals and delete repeats

We obtain the *ordered set*

$$\{a_1, a_2, a_3, a_4, \ldots\} = \left\{ \frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{1}{3}, \frac{3}{1}, \frac{4}{1}, \frac{3}{2}, \frac{2}{3}, \frac{1}{4}, \frac{1}{5}, \ldots \right\}.$$

Now define the function $f : \mathbb{N} \to \mathbb{Q}^+$ by $f(n) = a_n$. This is certainly a function. We claim that it is a bijection.

(*Injectivity*) Let $m, n \in \mathbb{N}$, and suppose that $f(n) = f(m)$. Then $a_m = a_n$. In the above construction we deleted any rational number which had already appeared in the list. Thus a_m can only equal a_n if $m = n$.

(*Surjectivity*) A positive rational number $\frac{a}{b}$ appears in the ath column and bth row of the square (and in many other places, $\frac{a}{b} = \frac{2a}{2b} = \cdots$). We only delete a fraction $\frac{a}{b}$ if it has already appeared in the list, therefore every positive rational lies in the range of f.

To finish things off, we extend the function to all rational numbers by

$$g : \mathbb{Z} \to \mathbb{Q} : n \mapsto \begin{cases} f(n) & \text{if } n > 0, \\ 0 & \text{if } n = 0, \\ -f(-n) & \text{if } n < 0. \end{cases}$$

We are merely using f to identify the negative integers with the negative rationals. It is immediate that $g : \mathbb{Z} \to \mathbb{Q}$ is a bijection. Appealing to Theorem 8.4, we deduce that $|\mathbb{Q}| = |\mathbb{Z}| = \aleph_0$, and so $\mathbb{Q}$ is countably infinite. ∎

This result should surprise you! Any sensible person should feel that there are far, far more rational numbers than integers, and yet the two sets have the same cardinality. Bizarre.

There are other countably infinite sets that appear to be even larger than $\mathbb{Q}$. For example, we can show that the Cartesian product $\mathbb{N} \times \mathbb{N}$ is countably infinite: use almost the same proof as for $\mathbb{Q}^+$ except that there are no repeats to delete. For a much larger-seeming yet still countably infinite set, consider the *algebraic numbers:*

$$\big\{x \in \mathbb{R} : p(x) = 0 \text{ for some polynomial } p \text{ with integer coefficients}\big\}.$$

Algebraic numbers are the zeros of polynomials with integer coefficients. Clearly any rational number $\frac{a}{b}$ is algebraic, since it satisfies $p(x) = 0$ for $p(x) = bx - a$. There are many more algebraic numbers than rational numbers: e.g. $\sqrt[5]{2} - 3$ is algebraic since it is a root of the polynomial $p(x) = (x+3)^5 - 2 = 0$. Not all real numbers are algebraic however: those which aren't, such as π and e, are termed *transcendental.*

The least infinite cardinal?

We originally introduced the symbol $\aleph_0$ to represent the cardinality of the 'simplest' infinite set. While the natural numbers are certainly infinite and straightforward, is there any more compelling reason why we should consider them to be the *most simple* infinite set? One reason lies in the following result.

Theorem 8.6. *A is a finite set if and only if $|A| < \aleph_0$.*

Otherwise said, every infinite set has cardinality *at least as large* as the natural numbers: $\aleph_0$ may be considered the least infinite cardinal.

Proof. ($\Longrightarrow$) The $n = 0$ case is left to the Exercises. Suppose that $|A| = n \geq 1$ so that we may list the elements of A as $\{a_1, \ldots, a_n\}$. We must prove two things:

8.1.1 $|A| \leq \aleph_0$. That is, $\exists f : A \to \mathbb{N}$ which is *injective.*

8.1.2 $|A| \neq \aleph_0$. That is, $\nexists g : A \to \mathbb{N}$ which is *bijective.* By symmetry this is equivalent to showing that there is no bijective function $h : \mathbb{N} \to A$.[a]

For part 1., simply define f by $f(a_k) = k$ for each $k \in \{1, 2, 3, \ldots, n\}$. This is injective since the distinct elements a_k of A map to distinct integers.
For part 2., suppose that $h : \mathbb{N} \to A$ is bijective. Consider the set

$$h\big(\{1, \ldots, n+1\}\big) = \big\{h(1), \ldots, h(n+1)\big\} \subseteq A.$$

Since A has n elements, by Dirichlet's box principle, at least two of the values $h(1), \ldots, h(n+1)$ must be equal. Therefore h is not injective and consequently not bijective. A contradiction.
($\Longleftarrow$) See Exercise 8.1.18. ∎

[a] If $g : A \to \mathbb{N}$ is a bijection, then $g^{-1} : \mathbb{N} \to A$ is also a bijection.

Of course, this doesn't answer the question of whether there exist infinite sets with larger cardinality than $\aleph_0$, though we shall answer this in the next section.

Aside. $\aleph_0$ **versus** ∞**: what's the difference?**

It can be difficult to grasp why $\aleph_0$ and ∞ are not the same thing. The problem is compounded by references to an 'infinite number' of objects whenever the cardinality of a set is not finite. This loose phrase is commonly used, but risks conflating the concepts of 'infinite set' and 'infinity.'

So what is the difference between $\aleph_0$ and ∞? If there aren't an 'infinite number' of natural numbers, how many are there? Theorem 8.6 says that $\aleph_0$ is 'larger than any natural number.' Is this not what we mean by infinity? The reason we need a new symbol $\aleph_0$, and why it and ∞ are different, is twofold:

8.1.1 As we shall see shortly, there are infinite sets with greater cardinality than $\aleph_0$: in a naïve sense, there are multiple infinities. The single symbol ∞ is insufficient to distinguish sets with different infinite cardinalities.

8.1.2 More philosophically, $\aleph_0$ is an *object* in its own right; an object to which the cardinality of some set may be equal. Indeed, by Theorem 8.2, $\aleph_0$ is an equivalence class.

By contrast, ∞ is typically not an object. The symbol ∞ is mostly used in *interval notation* and when talking about *limits:* in neither case does the symbol represent an object. For example:

- The interval $(2, \infty)$ is the set of all real numbers greater than 2. We don't say 'greater than 2 and *less than infinity.*'
- $\lim\limits_{x \to 3} \frac{1}{(x-3)^2} = \infty$ means that the function $f(x) = \frac{1}{(x-3)^2}$ gets unboundedly larger as x approaches 3. It is incorrect to say that $f(x)$ 'approaches infinity.' It is even worse to write $f(3) = \frac{1}{(3-3)^2} = \infty$.

The challenge of Cantor's notion of cardinality is to appreciate that the question, 'How many natural numbers are there?' is meaningless!

Reading Questions

8.1.1 A set A is *countably infinite* or *denumerable* if ________________. Select all that apply.

(a) There exists a surjection from $\mathbb{N}$ onto A.

(b) There exists an injection from $\mathbb{N}$ into A.

(c) There exists a bijection between A and $\mathbb{Q}$.

(d) There exists an injection from A into $\mathbb{N}$ and no injection from A into any finite set.

8.1.2 True or False: if A is a proper subset of B, then A has strictly smaller cardinality than B.

Practice Problems

8.1.1 Suppose that $A \neq \emptyset$. Prove that $|A| \leq |B|$ if and only if there is a surjection $g : B \to A$.

8.1.2 Let $a, b \in \mathbb{R}$ with $a < b$. Show that $|(a, b)| = |(0, 1)|$. Conclude that any two open intervals in $\mathbb{R}$ have the same cardinality.

Exercises

8.1.1 Refresh your proof skills by proving explicitly that the following functions are bijections:

(a) $f : \mathbb{N} \to 2\mathbb{N} : n \mapsto 2n$.

(b) $g : \mathbb{N} \to \mathbb{N}_{\geq 2} : n \mapsto n + 1$.

8.1.2 Construct a function $f : \mathbb{N} \to \mathbb{Z}_{\geq -3} = \{-3, -2, -1, 0, 1, 2, 3, 4, \ldots\}$ which proves that the latter set is countably infinite: you must show that your function is a bijection.

8.1.3 Prove that the set $3\mathbb{Z} + 2 = \{3n + 2 : n \in \mathbb{Z}\}$ is countably infinite.

8.1.4 Show that the set of all triples of the form $(n^2, 5, n + 2)$ with $n \in 3\mathbb{Z}$ is countably infinite by explicitly providing a bijection with a countably infinite set A. (*You must check that the set A is countably infinite, and that your map is indeed a bijection.*)

8.1.5 Imagine a hotel with an infinite number of rooms: Room 1, Room 2, Room 3, Room 4, etc.. Show that, even if the hotel is full, the guests may be re-accommodated so that there is always a room free for one additional guest.
Hint: consider the function $f : \mathbb{N} \to \mathbb{N} : n \mapsto n + 1$.

8.1.6 Let A be a set, and let B be a subset of A. Suppose B is countably infinite and $a \in A \setminus B$. Show $B \cup \{a\}$ is countably infinite.

8.1.7 Find an injection $f : \mathbb{Z} \to (0, 1)$.

8.1.8 Find an explicit bijection $f : [0, 1] \to (0, 1)$. Make sure to show your map is a bijection.

8.1.9 Prove that $A \subseteq B \implies |A| \leq |B|$. (*You need an injective function $f : A \to B$*)

8.1.10 Prove Theorem 8.2. (*You need little more than Theorem 3.18 on the composition of bijective functions.*)

8.1.11 Prove that the set $\mathbb{N} \times \mathbb{N}$ is countably infinite. You should base your proof on Theorem 8.5.

8.1.12 We know that $\mathbb{Q}$ is countably infinite, and we saw (Theorem 8.5) that there must exist a bijective function $f : \mathbb{N} \to \mathbb{Q}$. Show that $g : \mathbb{N} \times \mathbb{N} \to \mathbb{Q} \times \mathbb{Q}$ defined by $g(m, n) = (f(m), f(n))$ is a bijection. Appeal to the previous question to show that $\mathbb{Q} \times \mathbb{Q}$ is countably infinite.

8.1.13 Here we consider the $n = 0$ case of Theorem 8.6. Recall the definition of function in Section 7.2.

(a) If $|A| = 0$, then $A = \emptyset$. Suppose that $f : \emptyset \to \mathbb{N}$ is a function. Use Definition 7.4 to prove that $f = \emptyset$.

(b) State what it means, in the language of Definition 7.4, for a function $f : A \to \mathbb{N}$ to be injective. Show that $f = \emptyset$ is an injective function.

(c) Suppose that B is a set with $|B| \geq 1$. Prove by contradiction that there are no functions $h : B \to \emptyset$. Conclude that $0 < \aleph_0$.

8.1.14 Let A be a countably infinite set. Show that for any $n \in \mathbb{N}$, there is a partition $\mathcal{A} = \{A_1, \ldots, A_n\}$ of A such that each subset A_i in the partition is also countably infinite.

8.1.15 Suppose that the set A_n is countably infinite for each $n \in \mathbb{N}$. We may then list the elements of each set: $A_n = \{a_{n1}, a_{n2}, a_{n3}, a_{n4}, \ldots\}$. Now list the elements of the sets $A_1, A_2, A_3, \ldots$ as follows:

$$\begin{aligned} A_1 &= \{a_{11}, a_{12}, a_{13}, a_{14}, \ldots\} \\ A_2 &= \{a_{21}, a_{22}, a_{23}, a_{24}, \ldots\} \\ A_3 &= \{a_{31}, a_{32}, a_{33}, a_{34}, \ldots\} \\ &\ \ \vdots \end{aligned}$$

Use this construction to prove that $\bigcup\limits_{n \in \mathbb{N}} A_n$ is a countably infinite set.

This result is often stated, 'A countable union of countable sets is countable.'

8.1.16 Let $A = \{x \in \mathbb{R} : p(x) = 0 \text{ for some polynomial } p \text{ with integer coefficients}\}$ be the set of *algebraic* numbers. We will show that A is countable.

(a) Let $M \in \mathbb{N}$. Prove that there are only finitely many choices of $d \in \mathbb{N}$ and $a_0, \ldots, a_d \in \mathbb{Z}$ such that $M = d + |a_0| + \cdots + |a_d|$.

(b) Let $P_M = \{a_d x^d + \cdots + a_1 x + a_0 : M = d + |a_0| + \cdots + |a_d|\}$. Explain why P_M is finite.

(c) Using the fact that a polynomial of degree d can have at most d roots in $\mathbb{R}$, show that

$$R_M = \{x \in \mathbb{R} : p(x) = 0 \text{ for some } p \in P_M\}$$

is finite.

(d) Prove that $A = \bigcup_{M \in \mathbb{N}} R_M$ and conclude by Exercise 15 that A must be countably infinite.

8.1.17 (Hard!) In this question we complete the proof of Theorem 8.6 by showing that if $|A| < \aleph_0$, then A is a finite set.

We prove by contradiction. Suppose that A is an infinite set such that $|A| < \aleph_0$. Then there exists an injective function $f : A \to \mathbb{N}$. List the elements of the image of f in increasing order:

$$\text{range}(f) = \{n_1, n_2, n_3, \ldots\}.$$

(a) Prove that $\operatorname{Im} f$ is an infinite set.

(b) Show that for all $k \in \mathbb{N}$, there exists a unique $a_k \in A$ satisfying $f(a_k) = n_k$.

(c) Define $g : \mathbb{N} \to A$ by $g(k) = a_k$. Prove that g is a bijection.

(d) Why do we obtain a contradiction?

8.1.18 (Hard) Prove that a set A is infinite if and only if it has a proper subset $B \subset A$ with the same cardinality $|B| = |A|$.

8.2 Uncountable Sets

Since $\mathbb{Q}$ seems so large, you might think that there cannot be any sets with strictly larger cardinality. But we haven't yet thought about the real numbers...

Definition 8.7. A set A is **uncountable** if $|A| > \aleph_0$, that is if there exists an injection $f : \mathbb{N} \to A$ but no bijection $g : \mathbb{N} \to A$.

Theorem 8.8. *The interval* $[0, 1]$ *of real numbers is uncountable.*

We denote the cardinality of the interval $[0, 1]$ by the symbol $\mathfrak{c}$ for *continuum*. The theorem may therefore be written $\mathfrak{c} > \aleph_0$.

Proof. First we require an injective function $f : \mathbb{N} \to [0, 1]$. The function defined by $f(n) = \frac{1}{n}$ clearly fits the bill, for

$$f(n) = f(m) \implies \frac{1}{n} = \frac{1}{m} \implies n = m.$$

Now we prove that there exists no bijection $g : \mathbb{N} \to [0, 1]$, arguing by contradiction. Suppose that g is such a bijection and consider the sequence of values $g(1), g(2), g(3), \ldots$ These are real numbers between 0 and 1, hence they may all be expressed as decimals:[a]

$$\begin{aligned} g(1) &= 0.b_{11}b_{12}b_{13}b_{14}b_{15}b_{16}\cdots \\ g(2) &= 0.b_{21}b_{22}b_{23}b_{24}b_{25}b_{26}\cdots \\ g(3) &= 0.b_{31}b_{32}b_{33}b_{34}b_{35}b_{36}\cdots \\ g(4) &= 0.b_{41}b_{42}b_{43}b_{44}b_{45}b_{46}\cdots \\ g(5) &= 0.b_{51}b_{52}b_{53}b_{54}b_{55}b_{56}\cdots \\ &\ \ \vdots \end{aligned} \qquad \text{where each } b_{ij} \in \{0, \ldots, 9\}.$$

Since g is bijective, it is certainly surjective. It follows that all of the values $c \in [0, 1]$ appear in the above list of decimals. Now define a new decimal

$$c = 0.c_1c_2c_3c_4c_5\cdots \qquad \text{where} \quad c_n = \begin{cases} 1 & \text{if } b_{nn} \neq 1, \\ 2 & \text{if } b_{nn} = 1. \end{cases}$$

c is a non-terminating decimal whose digits are 1's and 2's, whence it has no other representation. Since c disagrees with $g(n)$ at the nth decimal place, we have $c \neq g(n)$, $\forall n \in \mathbb{N}$. Hence c is *not* in the above list. However $c \in [0, 1]$ and g is surjective, whence $c \neq g(n)$ for some $n \in \mathbb{N}$: a contradiction. We conclude that $\mathfrak{c} \neq \aleph_0$.
Putting this together with the first part of the proof where $\mathfrak{c} \geq \aleph_0$, we conclude that $\mathfrak{c} > \aleph_0$. ∎

[a] A number has two decimal representations if and only if one of them terminates and the other ultimately becomes an infinite sequence of 9's. For the purposes of this proof it does not matter which representation is chosen when there is a choice. We are forced, however, to take $1 = 0.999999\cdots$, due to our insistence that all elements be written with zero units.

The second part of the proof is known as *Cantor's diagonal argument,* since we are comparing the constructed decimal c with the diagonal of an infinite square of integers. We have proved that the interval $[0,1]$ has a strictly larger cardinality than the set of integers. Since $[0,1] \subseteq \mathbb{R}$, it follows immediately that the real numbers are also uncountable. Indeed we shall see in a moment that the real numbers also have cardinality $\mathfrak{c}$, as does any interval (of positive width). More amazingly, the Cantor middle-third set (page 173) also has cardinality $\mathfrak{c}$, despite seeming vanishingly small.

More advanced ideas

Our countable and uncountable examples are merely scratching the surface of a truly weird subject. We conclude these notes with a couple more ideas.

The following theorem is very useful for being able to compare cardinalities. It allows us to prove that two sets have the same cardinality *without* explicitly constructing bijective functions. Injective functions are usually much easier to find.

Theorem 8.9 (Cantor–Schröder–Bernstein). *If* $|A| \leq |B|$ *and* $|B| \leq |A|$, *then* $|A| = |B|$.

The theorem seems like it should be obvious, but pause for a moment: it is *not* a result about *numbers!* A and B are *sets,* and so the theorem must be understood in the context of Definition 8.1. In this language the theorem becomes:

> Suppose there exist *injective* functions $f : A \to B$ and $g : B \to A$.
> Then there exists a *bijective* function $h : A \to B$.

The proof is beautiful, though a little long to reproduce here. If you are interested it can be found in any text on set theory. The applications of the theorem are more important to our purposes.

Theorem 8.10. *The interval* $(0,1)$ *has cardinality* $\mathfrak{c}$.

It is possible to explicitly define a bijection $h : (0,1) \to [0,1]$, although it is very messy. Instead we construct two injections.

Proof. $f : (0,1) \to [0,1] : x \mapsto x$ is clearly an injection, whence $|(0,1)| \leq |[0,1]| = \mathfrak{c}$. Now define

$$g : [0,1] \to (0,1) : x \mapsto \frac{1}{2}x + \frac{1}{4}.$$

g is certainly injective, and so $\mathfrak{c} \leq |(0,1)|$.
By the Cantor–Schröder–Bernstein Theorem, the sets $(0,1)$ and $[0,1]$ have the same cardinality $\mathfrak{c}$. ∎

In case you're feeling nervous, note that the function g in the proof isn't surjective: the range of g is the interval $[\frac{1}{4}, \frac{3}{4}] \neq (0,1)$. By a similar trick, covered in the Exercises, one can see that $\mathbb{R}$ also has cardinality $\mathfrak{c}$.

For a final punchline, we prove Cantor's Theorem, which says that the power set of any set A always has a strictly larger cardinality than A. In Theorem 6.6 we saw that $|\mathcal{P}(A)| = 2^{|A|}$ for finite sets A. We therefore already believe that Cantor's Theorem is true for finite sets. The proof that follows also works for infinite sets.

Theorem 8.11 (Cantor)**.** *If A is any set, then $|A| \lneq |\mathcal{P}(A)|$.*

Proof. If $A = \emptyset$, the result is trivial. Otherwise, we must show two things:

- $\exists f : A \to \mathcal{P}(A)$ which is injective.
- $\nexists g : A \to \mathcal{P}(A)$ which is bijective.

For the first, note that $f : a \mapsto \{a\}$ is a suitable injective function.
Now suppose for a contradiction that $\exists g : A \to \mathcal{P}(A)$ which is bijective. That is, $g(a)$ is a subset of A for each $a \in A$. Consider the set

$$X = \{a \in A : a \notin g(a)\}.$$

It is important to note that X *is a subset of A.*

We pause the proof for a moment, as the set X is somewhat tricky to think about. Before proceeding, let us consider an example. Suppose that $g : \{1,2\} \to \mathcal{P}(\{1,2\})$ is defined by

$$g(1) = \{1,2\}, \qquad g(2) = \{1\}.$$

Then $1 \in g(1)$ and $2 \notin g(2)$, whence the above set is $X = \{2\}$. Since we are trying to prove that no bijection $g : A \to \mathcal{P}(A)$ exists, it is important to note that the function g in our example is *not bijective!*

Proof Continued. By assumption, g is bijective, hence it is certainly surjective. Because the range of g is the power set $\mathcal{P}(A)$, the set X *lies in the image of g.* Otherwise said, there exists $b \in A$ such that $g(b) = X$. We ask whether b is an element of X. Think carefully about the definition of X, and observe that

$$\begin{aligned} b \in X &\iff b \notin g(b) && \text{(by the definition of } X\text{)} \\ &\iff b \notin X && \text{(since } X = g(b)\text{)} \end{aligned}$$

Look at what we have concluded: $b \in X \iff b \notin X$. This is clearly a contradiction!
It follows that there exists no bijection $g : A \to \mathcal{P}(A)$, and so $|A| \lneq |\mathcal{P}(A)|$. ■

The main implication of this is that *there is no largest cardinality!* We can always construct a larger set simply by taking the power set of what we already have. For example, $\mathcal{P}(\mathbb{R})$ has larger cardinality than $\mathbb{R}$. If you want a set with even larger cardinality, why not take $\mathcal{P}(\mathcal{P}(\mathbb{R}))$? Or $\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{R})))$. We can continue this process indefinitely.

Cantor's Theorem played a large part in pushing set theory towards axiomatization. Here is a conundrum motivated by the theorem: If a 'set' is just a collection of objects, then we may consider the 'set of all sets.' Call this A. Now consider the power set of A. Since $\mathcal{P}(A)$ is a set of sets, it must be a subset of A, whence $|\mathcal{P}(A)| \leq |A|$. However, by Cantor's Theorem, we have $|A| \lneq |\mathcal{P}(A))|$. The conclusion is the manifest absurdity

$$|A| \lneq |A|$$

The remedy is a thorough definition of 'set' which prevents the collection of all sets from being considered a set. This is where *axiomatic set theory* begins.

A word on the limits of proof

Throughout this course we have learned about some of the basic methods and and concepts used by the mathematician. In particular, we learned about various types of proof and how to use these proofs to demonstrate the truth of statements about mathematical objects. As we finish the course, it makes sense to reflect on the limits of our methods.

In the early 20th century, the discovery of various paradoxes and contradictions led to a foundational crises in mathematics. After all, it is difficult to build a house if you have cracks in your foundation! The result was an effort to put all of mathematics on a rigorous axiomatic basis by formulating a list of reasonable axioms from which all of mathematics could be derived, using basic logical reasoning. This axiomatic foundation ideally would satisfy the following conditions:

8.2.1 *consistency*, i.e. no contradiction would be derivable from the axioms;

8.2.2 *completeness*, i.e. all true mathematical statements would be derivable from the axioms.

The hope for such a foundation was crushed in 1931, when a young logician by the name of Kurt Gödel published his famous *Incompleteness Theorems* which showed that no such axiomatic system could exist. Essentially, Gödel showed that in any consistent axiomatic system that was strong enough to produce some basic arithmetic, there must be statements which are neither derivable nor refutable from the axioms. Perhaps even worse, no such system can prove its own consistency.

While the strongest aims of some of the early 20th century attempts at an axiomatic foundation cannot be accomplished, the research of that time was able to provide a foundation that most modern mathematicians deem adequate for current work. Perhaps the most popular approach is to base all of mathematics on set theory – you will see as your studies progress that many of the objects you study can be formalized as sets together with functions and relations between sets. We have seen in Chapter 7 that functions and relations are just themselves sets. Even numbers like $0, 1, 2$ or $\frac{12}{19}$ or $3.14\ldots$ can be thought of as sets, if one desires. In turn, set theory is often axiomatized using the ZFC axioms (short for Zermelo-Fraenkel set theory with the Axiom of Choice).

While the ZFC axioms are subject to the limitations imposed by Gödel's theorems, they have proven themselves by being able to formalize most of the mathematics actually used by current mathematicians, and have so far not produced any inconsistencies. Thus most mathematicians feel little need to dwell on the foundational issues of the previous century.

Reading Questions

8.2.1 A set A is *uncountable* if and only if ________________________________.

(a) there is a bijection between A and $[0, 1]$.
(b) there is a surjection from $\mathbb{R}$ onto A.
(c) there is an injection from $\mathbb{N}$ into A.
(d) there exists no injection from A into $\mathbb{N}$.

8.2.2 Which of the following sets are uncountable. Select all that apply.

(a) $(1, 2] \cup \{3\}$
(b) $\mathbb{N} \times [1, 2]$
(c) $\mathbb{R} \setminus \{\frac{1}{n} : n \in \mathbb{N}\}$
(d) $\mathbb{Q} \cap [1, 2]$

8.2.3 True or False: there is no set A such that there is a surjection from $\mathcal{P}(A)$ onto A.

Practice Problems

8.2.1 Let $\{0, 1\}^{\mathbb{N}}$ denote the set of all sequences $(x_1, x_2, \ldots)$ such that each x_i is 0 or 1. In other words, A is the set of all functions $f : \mathbb{N} \to \{0, 1\}$. The cardinality of $\{0, 1\}^{\mathbb{N}}$ is often $2^{\aleph_0}$. Show $2^{\aleph_0} = \mathfrak{c}$. [Hint: use the Cantor-Schröder-Bernstein theorem.]

Exercises

8.2.1 You may assume that $[0, 1]$ has cardinality $\mathfrak{c}$.

(a) Construct an explicit bijection $f : [0, 1] \to [3, 8]$ which proves that the interval $[3, 8]$ also has cardinality $\mathfrak{c}$. *Try a linear function mapping the endpoints of* $[0, 1]$ *to the endpoints of* $[3, 8]$.
(b) Let $a, b \in \mathbb{R}$ with $a < b$. Generalizing part (a), construct a bijection which proves that the closed interval $[a, b]$ has cardinality $\mathfrak{c}$.

8.2.2 (a) Suppose that $g : \{1, 2, 3, 4\} \to \mathcal{P}(\{1, 2, 3, 4\})$ is defined by

$$g(1) = \{1, 2, 3\}, \qquad g(2) = \{1, 4\}, \qquad g(3) = \emptyset, \qquad g(4) = \{2, 4\}.$$

Compute the set $X = \{a \in \{1, 2, 3, 4\} : a \notin g(a)\}$.

(b) Repeat part (a) for $g : \mathbb{N} \to \mathcal{P}(\mathbb{N}) : n \mapsto \{x \in 2\mathbb{N} : x \leq n\}$.

8.2.3 Let A be a countably infinite set. For any $n \in \mathbb{N}$, prove $A^n = \underbrace{A \times \cdots \times A}_{n \text{ times}}$ is countably infinite.

8.2.4 The proof of Cantor's Theorem makes use of a construction similar to *Russell's Paradox.* Let X be the set of all sets which are not members of themselves: explicitly

$$X = \{A : A \notin A\}.$$

(a) Assume that X is a set, and use it to deduce a contradiction: ask yourself if X is a member of itself.

(b) Russell's paradox is one avatar of an ancient logical conundrum which appears in many guises. For example, suppose that a town has one hairdresser, and suppose that the hairdresser is the person who cuts the hair of all the people, and only those people, who do not cut their own hair. Who then cuts the hairdresser's hair? Can you explain the connection with Russell's paradox/Cantor's Theorem?

The point of Russell's paradox is that we need a definition of 'set' which prevents objects like X from being considered sets.

8.2.5 Let $A = \{0,1\}^{\mathbb{N}}$ denote the set of all sequences $(x_1, x_2, \ldots)$ such that each x_i is 0 or 1. In other words, A is the set of all functions $f : \mathbb{N} \to \{0,1\}$. Use a diagonal argument similar to the proof of Theorem 8.8 to show A is uncountable.

8.2.6 Recall the Cantor set as described in the notes, where we proved that $\mathcal{C}$ is the set of all numbers in $[0,1]$ possessing a ternary expansion consisting only of zeros and twos. Modeling your answer on the proof that the interval $[0,1]$ is uncountable, prove that $\mathcal{C}$ is uncountable.

8.2.7 Let $\mathbb{I} = \mathbb{R} \setminus \mathbb{Q}$ be the set of irrational numbers.

(a) Prove that $|\mathbb{I}| \leq \mathfrak{c}$.

(b) Prove that $x \in \mathbb{Q} \implies x + \sqrt{2} \in \mathbb{I}$. Hence conclude that $\aleph_0 \leq |\mathbb{I}|$.

(c) Appeal to Exercise 8.1.15 to argue that the irrational numbers are uncountable.

It is true, though we haven't show it, that $|\mathbb{I}| = \mathfrak{c}$. *Doing so is more difficult!*

8.2.8 A real number $x \in \mathbb{R}$ is called transcendental if it is not algebraic, i.e. not the root of any polynomial with integer coefficients. Show there are uncountably many transcendental numbers. [Hint: see Exercise 16 in Section 8.]

8.2.9 (a) Prove that $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined by $f(m,n) = 2^m 3^n$ is injective.

(b) Use part (a) and the Cantor–Schröder–Bernstein Theorem to conclude that $|\mathbb{N} \times \mathbb{N}| = \aleph_0$.

(c) Extend your argument to conclude that, for any $k \in \mathbb{N}$,

$$\underbrace{|\mathbb{N} \times \cdots \times \mathbb{N}|}_{k\text{times}} = \aleph_0$$

(d) Use part (b) to provide an alternative proof that $|\mathbb{Q}^+| = \aleph_0$.

8.2.10 (a) Show that $|(0,1)| \leq |\mathbb{R} \setminus \mathbb{N}| \leq |\mathbb{R}|$.

(b) Construct a bijection $f : (0,1) \to (-\frac{\pi}{2}, \frac{\pi}{2})$. (*Try a linear function*)

(c) Show that $g : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R} : x \mapsto \tan x$ is a bijection.

(d) Use the Cantor–Schröder–Bernstein Theorem to conclude that $|\mathbb{R} \setminus \mathbb{N}| = |\mathbb{R}| = \mathfrak{c}$.

8.2.11 Show that the complex numbers $\mathbb{C}$ have the cardinality of the continuum $|\mathbb{C}| = \mathfrak{c}$.

8.2.12 Give an example of an uncountable I and $\{A_n : n \in I\}$ such that each A_n is countably infinite, and the following three conditions hold:

(i) if $m \neq n$, then $A_m \neq A_n$,

(ii) for all m, n, either $A_m \subseteq A_n$ or $A_m \supseteq A_n$,

(iii) $\bigcup_{n \in I} A_n$ is countably infinite.

8.2.13 (Hard!) Let $x \in [0, 1]$. The *binary expansion* of x is the sequence b_n of zeros and ones such that

$$x = \sum_{n=1}^{\infty} \frac{b_n}{2^n}.$$

Given the choice,[25] we choose the terminating binary expansion of x. With such a caveat, you are given that the binary expansion of $x \in [0, 1]$ is unique. Define a function $f : [0, 1] \to \mathcal{P}(\mathbb{N})$ by

$$f(x) = \{n \in \mathbb{N} : b_n = 1 \text{ in the binary expansion of } x\}.$$

(a) Prove that f is an injection, and that, consequently, $\mathfrak{c} \leq |\mathcal{P}(\mathbb{N})|$.

(b) Prove that the function $g : \mathcal{P}(\mathbb{N}) \to \mathcal{C}$ (the Cantor set) defined by

$$g(X) = \sum_{n \in X}^{\infty} \frac{2}{3^n}$$

is a *bijection.*

(c) Use Cantor–Schröder–Bernstein to conclude that $|\mathcal{P}(\mathbb{N})| = |\mathcal{C}| = \mathfrak{c}$.

8.2.14 Let A and B be sets.

(a) Define $|A| \cdot |B|$ to be $|A \times B|$. Note that when A and B are finite, this definition agrees with our usual notion of multiplication (i.e. if $|A| = m$ and $|B| = n$, then $|A| \cdot |B| = m \cdot n$).

(b) Show that if A and B are nonempty and at least one of them is infinite, then $|A| \cdot |B| = \max\{|A|, |B|\}$.

8.2.15 Let A and B be sets.

(a) Show $\max\{|A|, |B|\} \leq |A \cup B|$.

(b) Define $|A| + |B|$ to be $|(A \times \{0\}) \cup (B \times \{1\})|$. Show that when A and B are finite, this definition agrees with our usual notion of addition (i.e. if $|A| = m$ and $|B| = n$, then $|A| + |B| = m + n$).

(c) Show $|A \cup B| \leq |A| + |B|$.

(d) If at least one of A or B is infinite, show $|A| + |B| \leq \max\{|A|, |B|\}$. Conclude that $|A| + |B| = \max\{|A|, |B|\}$.

[25] The binary expansion of x is unique unless x has a terminanting expansion, in which case the the other expansion involves an infinite sequence of ones: e.g. $[0.011111\cdots]_2 = [0.1]_2$ in binary.