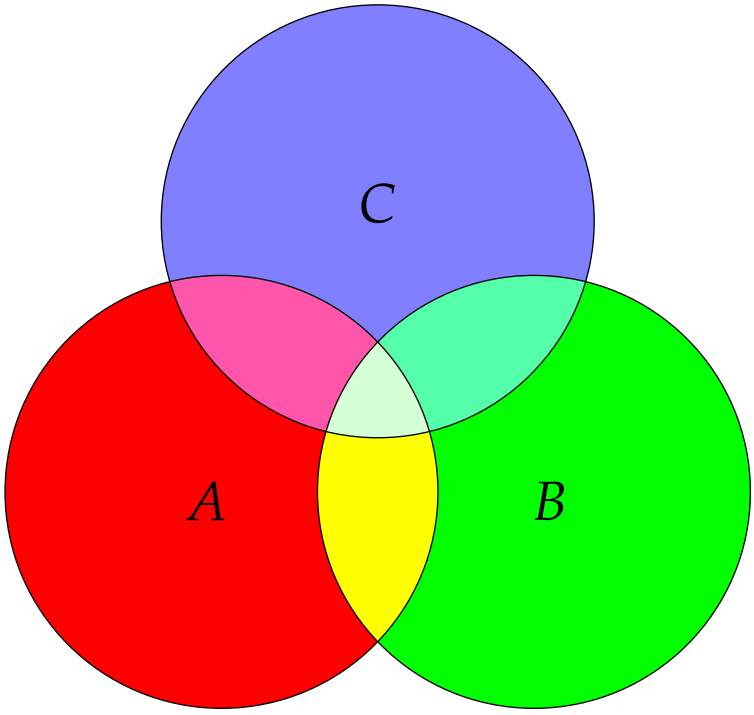
Math 13 — An Introduction to Abstract Mathematics

Neil Donaldson and Alessandra Pantano

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

March 10, 2024



Contents

Preface: What is Math 13 and who is it for?

Math 13 was created by the late Howard Tucker as a traditional discrete mathematics course; the number 13 was chosen as a joke! It eventually evolved to become the key transition class in the undergraduate program, introducing students to abstraction and proof, and serving as the primary pre-requisite for upper-division courses such as abstract algebra, analysis, linear algebra & number theory.

The typical student is simultaneously working through lower-division calculus and linear algebra. Knowledge of such material is unnecessary and students are encouraged to take the class early so as to ease the transition from algorithmic to abstract mathematics and so that a proof-mentality may be brought to other lower-division classes.

This text evolved from course notes dating back to 2008. Math 13 is something of a hydra due to the niche it occupies in UCI's program: part proof-writing, part discrete mathematics, and part introduction to specific upper-division topics. Logic is covered only at a basic level so that mathematical proofs may be engaged with quickly. Set theory is spread through the text; the intent is for dry 'grammar' topics to be absorbed via engagement with more accessible and fun ideas. By the end of the course, interested students should be prepared for a formal study of logic and set theory at the upper-division level.

Learning Outcomes

1. Developing the skills necessary to read and practice abstract mathematics.
2. Understanding the concept of proof and becoming acquainted with multiple proof techniques.
3. Learning what sort of questions mathematicians ask and what excites them.
4. Introducing upper-division mathematics by providing a taste of what is covered in several courses. For instance:

   Number Theory & Abstract Algebra How can we perform arithmetic with remainders? Can you figure out what on which *day* of the week you were born?

   Geometry and Topology How can we visualize and work with objects such as the Möbius strip? How can we use sequences of sets to produce objects (fractals) that appear similar at all scales?

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

Useful Texts The following texts are recommended if you want more exercises and material. The first two are available free online, while the remainder were previous textbooks for Math 13.

- *Book of Proof*, Richard Hammack
- *Mathematical Reasoning*, Ted Sundstrom
- *Mathematical Proofs: A Transition to Advanced Mathematics*, Chartrand/Polimeni/Zhang
- *The Elements of Advanced Mathematics*, Steven G. Krantz
- *Foundations of Higher Mathematics*, Peter Fletcher and C. Wayne Patty

1 Introduction: What is a Proof?

The essential concept in higher-level mathematics is that of *proof.* A basic dictionary entry might cover two meanings:

1. A test or trial of an assertion.
2. An argument that establishes the validity (truth) of an assertion.

In science and the wider culture, the first meaning predominates: a defendant was *proved* guilty in court; a skin cream is clinically *proven* to make you look younger; an experiment *proves* that the gravitational constant is 9.81ms^{-2}. A common mistake is to assume that a *proved assertion* is actually *true.* Two juries might disagree as to whether a defendant is guilty, and for many crimes the truth is uncertain hence the more nuanced legal expression *proved beyond reasonable doubt.*

In mathematics we use the second meaning: a proof establishes the incontrovertible *truth* of some assertion. To see what we mean, consider a simple claim (mathematicians use the word *theorem*).

Theorem 1.1. *The sum of any pair of even integers is even.*

Hopefully you believe this statement. But how do we *prove* it? We can *test* it by verifying examples ($4+6=10$ is even, $(-8)+30=22$ is even, etc.), but we cannot expect to verify *all* pairs this way. For a mathematical proof, we somehow need to test all possible examples simultaneously. To do this, it is essential that we have a clear idea of what is meant by an *even integer.*

Definition 1.2. An integer is *even* if it may be written in the form $2k$ where k is an integer.

Proof. Let x and y be even. Then $x=2k$ and $y=2l$ for some integers k and l. But then

$$x+y=2k+2l=2(k+l) \qquad (*)$$

is even. ■

The box ■ indicates that we've finished our argument. Traditionally the letters Q.E.D. were used, an acronym for the Latin *quod erat demonstrandum* (*which is what was to be demonstrated*).

Consider how the proof depends crucially on the definition.

- The theorem did not mention any *variables,* though these were essential to the proof. The variables k and l come for free *once you write the definition of evenness!* This is very common; a proof is often little more than rearranged definitions.
- According to the definition, $2k$ and $2l$ together represent *all possible pairs* of even integers. It is essential that k and l be *different symbols,* otherwise all you would be proving is that twice an even number is even!
- The calculation $(*)$ is the easy bit; without the surrounding sentences and the direct reference to the definition of evenness, the calculation means nothing.

There is some sleight of hand here; a mathematical proof establishes truth only by reference to one or more definitions. In this case, the definition and theorem also depend on the meanings of *integer* and *sum,* but we haven't rigorously defined either since to do so would take us too far afield. In any context, some concepts will be considered too basic to merit definition.

Theorems & Conjectures

Theorems are true mathematical statements that we can prove. Some are important enough to be named (the Pythagorean theorem, the fundamental theorem of calculus, the rank–nullity theorem, etc.), but most are simple statements such as Theorem 1.1.

In practice we are often confronted with *conjectures*: statements we suspect to be true, but which we don't (yet) know how to prove. Much of the fun and messy creativity of mathematics lies in formulating and attempting to prove (or disprove) conjectures.

A conjecture is the mathematician's equivalent to the scientist's hypothesis: a statement one would like to be true. The difference in approach takes us right back to the dual meaning of *proof.* The scientist *tests* their hypothesis using the scientific method, conducting experiments which attempt (and hopefully fail!) to show that the hypothesis is incorrect. The mathematician tries to *prove* that a conjecture is undeniably true by relying on logic. The job of a mathematical researcher is to formulate conjectures, prove them, and publish the resulting theorems. Creativity lies as much in the formulation as in the proof. Attempting to formulate your own conjectures is an essential part of learning mathematics; many will likely be false, but you'll learn a lot in the process!

Here are two conjectures to give us a taste of this process.

Conjecture 1.3. *If n is any odd integer, then $n^2 - 1$ is a multiple of 8.*

Conjecture 1.4. *If n is any positive integer, then $n^2 + n + 41$ is prime.*[1]

How can we decide if these conjectures are true or false? To get a feel for things, we start by computing with several small integers n. In practice, this process is likely what lead to the formulation of the conjectures in the first place!

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

Since 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? Establishing whether each conjecture is true or false requires 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 one</u> instance in which the statement is false.

Let us start with Conjecture 1.3. If n is an odd integer, then, by definition, we may write $n = 2k + 1$ for some integer k. Now compute the object of interest:

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

We need to investigate whether this is *always* a multiple of 8. Since k is an integer, $n^2 - 1$ is plainly a multiple of 4, so everything comes down to deciding whether $k(k+1)$ is *always even.* Do we believe

[1] A positive integer is *prime* if it cannot be written as the product of two integers, both greater than one.

this? We return to testing 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* or *disprove it*. Observe that $k(k+1)$ is the *product of two consecutive integers*. This is great, because for any two consecutive integers, one is even and the other odd, so their product must be even. Conjecture 1.3 is indeed a *theorem!*

Everything so far has been investigative. Scratch work is an essential part of the process, but it isn't something we should expect a reader to have to fight their way through. We therefore offer a formal proof. This is the final result of our deliberations; investigate, spot a pattern, conjecture, prove, and finally present our work in as clean and convincing a manner as we can.

Theorem 1.5. *If n is any odd integer, then n^2-1 is a multiple of 8.*

Proof. Let n be any odd integer. By definition, we may write $n = 2k+1$ for some integer k. Then

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

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. ∎

All that work, just for five lines of clean argument! But wasn't it *fun*?

When constructing elementary proofs it is common to feel unsure over how much detail to supply. We plainly relied on the definition of *oddness*, but we also used the fact that a product is even whenever either factor is even; does this need a proof? Since the purpose of a proof is to convince the reader, the appropriateness of an argument will depend on context and your audience: if you are trying to convince a middle-school student, maybe you should justify this step more fully, though the cost would be a longer argument that might be harder to grasp in its totality. A perfect proof that is best for all situations is unlikely to exist! A good rule is to imagine you are writing for another mathematician at the same level as yourself—if a fellow student believes your argument, that's a good sign of its validity.

Now consider Conjecture 1.4. The question is whether n^2+n+41 is prime for *every* positive integer n. When $n \leq 7$ the answer is yes, but examples do not make a proof! To investigate further, return to the definition of prime (Footnote 1): is there a positive integer n for which is n^2+n+41 can be factored as a product of two integers, both at least 2? A straightforward answer is staring us in the face! When $n = 41$ such a factorization certainly exists:

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

We call $n = 41$ a *counterexample*; it shows that there is at least one integer n for which n^2+n+41 is *not* prime. Conjecture 1.4 is therefore false (it has been *disproved*).

Planning and Writing Proofs

Your main responsibility in this course is the construction of proofs. Their sheer variety means that, unlike in elementary calculus, you cannot simply practice computing tens of similar problems until the process becomes automatic. So how do you learn to write proofs?

The first step is to *read* other arguments. Don't just accept them, make sure you *believe* them: check the calculations, verify claims, rewrite the argument in your own words adding any clarification you think necessary.

As you read others' arguments, the question will often arise: *how did they ever come up with this?* As our work on Theorem 1.5 shows, the source of a proof is often less magical than it appears; usually the author experimented until they found something that worked. Most of that experimentation gets hidden in the final proof which should be as clean and easy to read as possible. Imagine it as a concert performance after lots of private practice; no-one wants to hear wrong notes at the Carnegie Hall!

In order to bridge the gap, we recommend splitting the proof-writing process into several steps.

Interpret Make sense of the statement. What is it saying? Can you rephrase in a way that is clearer to you? What are you assuming? The most important part of this step is identifying the *logical structure* of the statement. We'll discuss this at length in the next chapter.

Brainstorm Convince *yourself* that the statement is true. First, look up the relevant definitions. Next, think of some instances where the conditions of the statement are met. Try out some examples, and ask yourself what makes the claim work in those instances. Examples can be crucial for building intuition about *why* the claim is true and can sometimes suggest a proof strategy. Review other theorems that relate to these definitions. Do you know any theorems that relate your assumptions to the conclusion? Have you seen a proof of a similar statement before?

Sketch Build the skeleton of your proof. Think again about what you are assuming and what you are are you trying to prove. As we'll see in the next chapter, it is often straightforward to write down reasonable *first* and *last* steps (the bread slices of a *proof-sandwich*). Try to connect these with informal arguments. If you get stuck, try a different approach.

This step is often the longest in the proof-writing process. It is also where you will be doing most of your calculations. You can be as messy as you want because *no-one ever has to see it*! Once you've learned a variety of different proof methods, this is a good stage at which to experiment with different approaches.

Prove Once you have a suitable sketch, it's time to prove the statement to the world. Translate your sketch into a linear story, written in complete sentences. Carefully word your explanations and avoid shorthand, though well-understood mathematical symbols like $\implies$ are encouraged. The result should be a clear, formal proof like you'd find in a mathematics textbook. Although you are providing a mathematical argument, your proof should read like prose.

Review Finally, *review* your proof. Assume the reader is meeting the problem for the first time and has not seen your sketch. Read your proof with skepticism; consider its readability and flow. Get read of unnecessary claims and revise the wording if necessary. Read your proof out loud. If you're adding extra words that aren't written down, include them in the proof. Finally, share your work with others. Do they understand it *without any additional input from you*?

Conjectures: *True* or *False*?

Higher-level mathematics is all about the important links between proofs, definitions, theorems and conjectures. We prove theorems (and solve homework problems) because they make us use, and aid our understanding of, definitions. We state definitions to help us formulate conjectures and prove theorems. One does not *know* mathematics, one *does* it. Mathematics is a *practice*; an art as much as it is a science.

With this in mind, do your best to prove or disprove the following conjectures. Don't worry if you're currently unsure as to the meanings of some of the terms or notation: ask! It will all be covered formally soon enough. At the end of the course, revisit these problems 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. If n is an integer greater than 2, then $n^2 - 1$ is not prime.
10. An integer is divisible by 5 when its last digit is 5.
11. If r is a rational number, then there is a non-zero integer n for which rn is an integer.
12. There is a smallest positive real number.
13. For all real numbers x, there exists a real number y for which $x < y$.
14. There exists a real number x such that, for all real numbers y, $x < y$.
15. 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.

2 Logic and the Language of Proofs

2.1 Propositions

In order to read and construct proofs, we need to start with the language in which they are written: *logic.* This is to mathematics what grammar is to English.

Definition 2.1. A *proposition* or *statement* is a sentence that is either true or false.

Examples 2.2. 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.

For a proposition to make sense, we must agree on the meaning of each concept it contains. When people argue over propositions, in practice they are often disagreeing about *definitions.* There are many concepts of God; we cannot begin to consider whether or not They exist until we agree *which* concept is being discussed! This also illustrates that the truth status of a proposition *need not be known* at the moment you state it; this is particularly common in mathematics.[2]

Truth Tables and Combining Propositions

To develop basic rules and terminology, it is helpful to consider *abstract* propositions: $P, Q, R, \ldots$. Given a small number of propositions, all possible combinations of truth states may be easily represented in tabular format: in a *truth table.* These are useful for defining new propositions.

Definition 2.3. Let P and Q be propositions. The truth tables below define three new propositions modeled on the words *and*, *or* and *not.*

- The *conjunction* $P \wedge Q$ is read "P and Q."
- The *disjunction* $P \vee Q$ is read "P or Q."
- The *negation* $\neg P$ is read "not P."

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

P	$\neg P$
T	F
F	T

The letters T/F stand for *true/false.* E.g., the second line of the first table says that if P is true and Q is false, then the proposition "P and Q" is false, while "P or 'Q" is true.

Example 2.4. Suppose P and Q are the following propositions:

P: "I like purple." Q: "I like chartreuse."

We form the new propositions described in the definition:

$P \wedge Q$: "I like purple and chartreuse." $P \vee Q$: "I like purple or chartreuse."

$\neg P$: "I do not like purple."

It is typical to modify phrasing to aid readability: "Not, I like purple" just sounds weird! Note also that or is *inclusive* in logic: if "I like purple or chartreuse" is true, then you might like *both*!

[2] More surprisingly, there are even propositions whose truth state is impossible to determine!

Let's continue by adding a third proposition:

R: "It's 9am."

What proposition is represented by the following English sentence?

"I like purple and I like chartreuse or it's 9am."

Is it $P \wedge (Q \vee R)$ or is it $(P \wedge Q) \vee R$? Without brackets, the sentence is unclear; the moral is that English is terrible at logic! Indeed, as the truth table shows, the two logical expressions really do mean different things!

P	Q	R	$Q \vee R$	$P \wedge (Q \vee R)$	$P \wedge Q$	$(P \wedge Q) \vee R$
T	T	T	T	T	T	T
T	T	F	T	T	T	T
T	F	T	T	T	F	T
T	F	F	F	F	F	F
F	T	T	T	F	F	T
F	T	F	T	F	F	F
F	F	T	T	F	F	T
F	F	F	F	F	F	F

Conditional and Biconditional Connectives

Of critical importance is the ability to have one proposition lead to another.

Definition 2.5. Given propositions P, Q, the *conditional* ($\Longrightarrow$) and *biconditional* ($\Longleftrightarrow$) *connectives* define new propositions as described in the truth table.

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

P	Q	$P \Longrightarrow Q$	$P \Longleftrightarrow Q$
T	T	T	T
T	F	F	F
F	T	T	F
F	F	T	T

Connective propositions can be read and written in many different ways:

$P \Longrightarrow Q$		$P \Longleftrightarrow Q$
P implies Q	P therefore Q	P if and only if Q
If P, then Q	Q follows from P	P iff Q
P only if Q	Q if P	P and Q are (logically) equivalent
P is sufficient for Q	Q is necessary for P	P is necessary and sufficient for Q

Example 2.6. The following sentences express, in English, the same conditional $P \Longrightarrow Q$.

- 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 for being Italian.
- Being Italian is necessary for being born in Rome.

Are you comfortable with what the propositions P and Q are here?

Connectives are central to mathematics for many reasons. In particular:

1. The vast majority of theorems we'll encounter may be written as a connective $P \implies Q$. For instance, revisit Theorem 1.1 and the discussion that follows:

   If x and y are even integers, then $x + y$ is even.

   Identifying the hypothesis and conclusion is essential if you want to understand a theorem!

2. Simple proofs typically involve chaining a sequence of connectives:

   $$P \implies P_2 \implies \cdots \implies P_n \implies Q$$

We'll revisit these ideas in Section 2.3, and repeatedly throughout the course.

While the biconditional should be easy to remember, it is harder to make sense of the conditional connective. Short of simply memorizing the truth table, here are two examples that might help.

Examples 2.7. 1. Suppose your professor says, "If the class earns a B average on the midterm, then I'll bring doughnuts." The only situation in which the teacher will have lied is if the class earns a B average but she fails to provide doughnuts.

2. ($F \implies T$ really can be true!) Let P be the proposition "$7 = 3$" and Q be "$0 = 0$." Since multiplication of both sides of an equation by zero is algebraically valid, we see that

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

   This argument is perfectly correct: the *implication* $P \implies Q$ is *true.* It (rightly!) makes us uncomfortable because the hypothesis is *false.*

   If we instead add 1 to each side of $7 = 3$, we'd obtain a example where $F \implies F$ is true.

Tautologies and Contradictions

Definition 2.8. 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.

Examples 2.9. 1. $P \wedge (\neg P)$ is a contradiction.

Regardless of the proposition P, it cannot be true at the same time as its negation!

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

2. $\big(P \wedge (P \implies Q)\big) \implies Q$ is a tautology.

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

The Converse and Contrapositive

Definition 2.10. The *converse* of $P \implies Q$ is the reversed implication $Q \implies P$.
The *contrapositive* of $P \implies Q$ is the implication $\neg Q \implies \neg P$.

It is vital to understand the distinction between these. In general, the truth status of the converse bears no relation to that of the original, though the contrapositive is much better behaved.

Theorem 2.11. *The contrapositive of an implication is logically equivalent to the original.*

Proof. Compute the truth table and observe that the third and sixth columns are identical:[3]

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

■

Example 2.12. Let P and Q be the following statements:

P: "Claudia is holding a peach." Q: "Claudia is holding a piece of fruit."

Since a peach is indeed a piece of fruit, the proposition $P \implies Q$ is *true*:

$P \implies Q$: "If Claudia is holding a peach, then she is holding a piece of fruit."

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

$Q \implies P$: "If Claudia is holding a piece of fruit, then she is holding a peach."

This is palpably false: Claudia could be holding an apple! However, in accordance with Theorem 2.11, the *contrapositive* is *true*:

$\neg Q \implies \neg P$: "If Claudia is *not* holding any fruit, then she is *not* holding a peach."

Negating Logical Expressions

Mathematics often requires us to negate propositions. What would you suspect to be the negation of a conditional $P \implies Q$? Is it enough to say "P doesn't imply Q"? But what does this mean?

We again rely on a truth table: to get the last column, recall that negation simply swaps T and F. Can we write this column in another way? Since there is only a single T in the final column, we see that we've proved the following.

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

Theorem 2.13. *$\neg(P \implies Q)$ is logically equivalent to $P \wedge \neg Q$ ("P and not Q").*

[3] Otherwise said, $(P \implies Q) \iff (\neg Q \implies \neg P)$ is a tautology.

Example 2.14. Consider the implication

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

Hopefully its negation is clear:

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

As in Example 2.7, it might help to think about what it means for the original statement to be *false*.

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.

Our final results in basic logic also involve negations; they are named for Augustus de Morgan, a famous 19th century logician.

Theorem 2.15 (de Morgan's laws). *Let P and Q be propositions.*

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

Proof. For the first law, observe that the fourth and seventh columns of the truth table are identical.

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

The second law is an exercise. ∎

Example 2.16. Consider the sentence:

I rode the subway *and* I had coffee.

To negate this using de Morgan's first law, we might write:

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

Subway	Coffee	Su and Co
T	T	T
T	F	F
F	T	F
F	F	F

This feels awkward in English because the negation encompasses three distinct possibilities. Note how the logical (inclusive) use of *or* includes the last row of the truth table: the possibility that you neither rode the subway nor had coffee.

As with Example 2.4, this is another advert for the use of logic: English simply isn't very helpful for precisely stating complex logical statements.

Aside: Algebraic Logic We can use truth tables to establish other laws of basic logic, for instance:

Double negation	$\neg(\neg P) \iff P$	
Commutativity	$P \wedge Q \iff Q \wedge P$	$P \vee Q \iff Q \vee P$
Associativity	$(P \wedge Q) \wedge R \iff P \wedge (Q \wedge R)$	$(P \vee Q) \vee R \iff P \vee (Q \vee R)$
Distributivity	$(P \wedge Q) \vee R \iff (P \vee R) \wedge (Q \vee R)$	$(P \vee Q) \wedge R \iff (P \wedge R) \vee (Q \wedge R)$

To make things more algebraic, we've replaced "is logically equivalent to" with a biconditional.[4]

Armed with such laws, one can often suitably manipulate logical expressions without laboriously creating truth tables. This is not the focus of this course, though you might find it fun!

For this course, it is probably not worth memorizing these laws. Your intuitive understanding of *and*, *or* and *not* mean you'll likely apply the laws correctly whenever necessary.

Exercises 2.1. A reading quiz and several questions with linked video solutions can be found online.

1. Express each statement in the form, "If . . . , then . . . " 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. Suppose "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 statement and think about each 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 $z \geq 3$.
3. Orange County is considering two competing transport plans: 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?" 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*)
4. Consider the proposition: "If the integer m is greater than 3, then $2m$ is not prime."
   (a) Rewrite the proposition using the word 'necessary.'
   (b) Rewrite the proposition using the word 'sufficient.'
   (c) Write the negation, converse and contrapositive of the proposition.
5. Suppose the following sentence is true: "If Amy likes art, then no-one likes history." What, if anything, can we conclude if we discover that someone likes history.

[4] Stating the laws in this fashion is to assert that each expression is a tautology (Definition 2.8). For instance, to claim that "$\neg(\neg P)$ is logically equivalent to P" is to assert that $\neg(\neg P) \iff P$ is a tautology.

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

7. Use truth tables to establish the following laws of logic:

   (a) Double negation: $\neg(\neg P) \iff P$.
   (b) Idempotent law: $P \wedge P \iff P$.
   (c) Absorption law: $P \wedge (P \vee Q) \iff P$.
   (d) Distributive law: $(P \wedge Q) \vee R \iff (P \vee R) \wedge (Q \vee R)$.

8. (a) Decide whether $(P \wedge \neg P) \implies Q$ is a tautology, a contradiction, or neither.
   (b) Explain why $\neg P \vee \neg Q$ is logically equivalent to $P \implies (P \wedge \neg Q)$.
   (c) Prove: $\big((P \wedge \neg Q) \implies F\big) \iff (P \implies Q)$ is a tautology. Here F represents a *contradiction.*

9. (a) Prove that the expressions $(P \implies Q) \wedge (Q \implies P)$ and $P \iff Q$ are logically equivalent.
   (b) Prove that $\big((P \implies Q) \wedge (Q \implies R)\big) \implies \big(P \implies R\big)$ is a tautology.

   Why do these make intuitive sense?

10. Use logical algebra (e.g., page 11) to show that $\big((P \vee Q) \wedge \neg P\big) \wedge \neg Q$ is a contradiction.

11. Do there exist propositions P, Q for which both $P \implies Q$ and its converse are *false*? Explain.

12. Your friend insists that the negation of the sentence "Mark and Mary have the same height" is "Mark or Mary do not have the same height." What is the correct negation? Where did your friend go wrong?

13. Suppose that the following statements are *true*:

    (a) Every octagon is magical.
    (b) If a polygon is not a rectangle, then is it not a square.
    (c) 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*)

14. The connective $\downarrow$ (the *Quine dagger*, *NOR*) is defined by the truth table:

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

    (a) Prove that $P \downarrow Q$ is logically equivalent to $\neg(P \vee Q)$.
    (b) Find a logical 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 $\downarrow$ which is logically equivalent to $P \wedge Q$.

15. (Just for fun) 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. When were they born and how old will they turn this year?

    Do your best to give a formal proof of your correctness.

2.2 Propositional Functions & Quantifiers

The majority of mathematical propositions are more complicated that those seen in Section 2.1. In particular, they typically involve *variables*, for instance

"x is an integer greater than 5."

Definition 2.17. A *propositional function* is a family of propositions which depend on one or more variables. The collection of permitted variables is the *domain*.

If P is a propositional function depending on a single variable x, then for each object a in the domain, $P(a)$ is a proposition. Typically $P(x)$ is true for some x and false for others.

Example 2.18. Consider the propositional function $P(x)$: "$x^2 > 4$" with domain the real numbers. Plainly $P(1)$ is false ("$1^2 > 4$") and $P(6)$ is true ("$6^2 > 4$").

In mathematics, propositional functions are often *quantified*. English contains various quantifiers (*all, some, many, few, several,* etc.), but in mathematics we are primarily concerned with just two.

Definition 2.19. The *universal quantifier* $\forall$ is read 'for all'. The *existential quantifier* $\exists$ is read 'there exists.' Given a propositional function $P(x)$, we define two new *quantified propositions*:

- "$\forall x, P(x)$" is true if and only if $P(x)$ is true for *every* x in its domain.
- "$\exists x, P(x)$" is true if and only if $P(x)$ is true for *at least one* x in its domain.

It is common to describe the domain when quantifying propositions by including a descriptor after the quantifier (*bounding the quantifier*—see below).

As with connectives, there are many ways to express quantified propositions both mathematically and in English. The use of symbolic quantifiers involves a trade-off: compact statements can improve clarity, but they are harder to read for the uninitiated, so consider your audience! While it is your choice whether to employ symbolic quantifiers in your own *writing*, it is essential that you know how to *read/recognize* them and that you can *translate* between various incarnations.

Example (2.18 cont.). To gain some practice with bounded quantifiers, we introduce the notation $x \in \mathbb{R}$: this means that x is a real number.

- "$\forall x \in \mathbb{R},\ x^2 > 4$" might be read, "The square of every real number is greater than 4."
  The quantified expression is *false* since $1^2 > 4$ is false: we call $x = 1$ a *counter-example*.
- "$\exists x \in \mathbb{R},\ x^2 > 4$" might be read, "There is a real number whose square is greater than 4."
  The quantified expression is *true* since $6^2 > 4$ (is true): we call $x = 6$ an *example*.

Due to their importance, it is worth defining these last concepts formally.

Definition 2.20. An *example* of $\exists x, P(x)$ is an element x_0 in the domain of P for which $P(x_0)$ is *true*. A *counter-example* to $\forall x, P(x)$ is an element x_0 in the domain of P for which $P(x_0)$ is *false*.

Universal Quantifiers and Connectives: Hidden Quantifiers Universally quantified statements are interchangeable with implications. Given a propositional function $Q(x)$, let $P(x)$ be the proposition "x lies in the domain of Q." Then

$\forall x, Q(x)$ is logically equivalent to $P(x) \implies Q(x)$

Connectives containing variables are therefore assumed to be universal. When written as a connective, the universal quantifier is typically *hidden*.[5]

Examples 2.21. 1. The universal statement "Every cat is neurotic," may also be written

If x is a cat, then x is neurotic.

2. Revisiting Example 2.18, we could rewrite "$\forall x \in \mathbb{R}, x^2 > 4$" as a connective

$x \in \mathbb{R} \implies x^2 > 4$ (If x is a real number, then $x^2 > 4$)

3. The following three sentences have identical meaning:

The square of an odd integer is odd. $\forall n$ odd, n^2 is odd. n odd $\implies n^2$ odd.

In only one of the sentences is the universal quantifier explicit. For even more variety, the third sentence can also be viewed as a universal statement about all *integers*; including the hidden quantifier in this case results in

$\forall n \in \mathbb{Z}$, n odd $\implies$ n^2 odd.

where the symbol $\mathbb{Z}$ represents the (set of) integers.

We've already seen that *disproving* a universal statement requires only that we supply a *counterexample*. While such might require some effort to find, often the resulting argument is very simple. By contrast, *proving* a universal statement is the same as proving a connective, an activity that is typically much more involved. We therefore largely postpone this to the next section. Regardless, a simple proof of our *oddness* claim should be easy to follow.

Proof of Example 2.21.3. If an integer n is odd, then it may be written in the form $n = 2k + 1$ for some integer k. But then

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

is plainly also odd. ∎

Similarly, *proving* an existential statement (by providing an *example*) is typically more straightforward than *disproving* such. To understand this duality, we need to understand how to *negate* quantified propositions.

[5] By contrast, the existential quantifier is never hidden: it is always explicitly written as a symbol ($\exists$), or as a phrase in English (*there is*, *there exists*, *some*, *at least one*, etc.).

Negating Quantified Propositions

To negate a proposition, we consider what it means for it to be *false.* We already understand what this means for a universal proposition:

"$\forall x, P(x)$" is false if and only if *there exists a counter-example.*

The negation of a universal statement is *existentially quantified*:

The negation of "$P(x)$ is *always* true" is "$P(x)$ is *sometimes* false."

Repeating this with $\neg P(x)$ results in a related observation:

The negation of "$P(x)$ is *always* false" is "$P(x)$ is *sometimes* true."

Theorem 2.22. *For any propositional function $P(x)$:*

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

Examples 2.23. 1. "Everyone owns a bicycle," has negation, "Someone does not own a bicycle." It is somewhat ugly, but we could write this symbolically:

$$\neg\big(\forall \text{ people } x,\ x \text{ owns a bicycle}\big) \iff \exists \text{ a person } x \text{ such that } x \text{ does not own a bicycle}$$

2. The quantified proposition[6] "$\exists x > 0$ such that $\sin x = 4$," has the form $\exists x,\ P(x)$. Its negation has the form $\forall x,\ \neg P(x)$. Explicitly:

$$\forall x > 0,\ \sin x \neq 4$$

Since the sine function satisfies the inequalities $-1 \leq \sin x \leq 1$, the original proposition is *false* and its negation *true.*

Warning! *Never* negate a quantifier's bounds: $\forall x \leq 0\ldots$ is completely wrong!

3. Be especially careful when negating connectives: after negation, a hidden quantifier $\forall x$ becomes *explicit.*

$\neg\big(P(x) \implies Q(x)\big)$ is logically equivalent to $\exists x, P(x) \wedge \neg Q(x)$

(a) (Example 2.21.3) The negation of "n odd $\implies n^2$ odd"is the (false) claim

$\exists n \in \mathbb{Z}$ with n odd *and* n^2 even.

(b) (Example 2.18) The negation of the false claim "$x \in \mathbb{R} \implies x^2 > 4$" is the true assertion

$\exists n \in \mathbb{R}$ for which $x^2 \leq 4$

[6] "$\exists x > 0$" indicates that the domain of the proposition "$\sin x = 4$" is the *positive* real numbers.

Multiple Quantifiers

A propositional function can have several variables, each of which may be quantified.

Examples 2.24. 1. The quantified proposition

$$\forall x > 0, \exists y > 0 \text{ such that } xy = 4$$

might be read, "Given any positive number, there is another such that their *product* is four." Hopefully you believe that this is *true*! Here is a simple argument which comes from viewing it as an implication, "If $x > 0$, then $\exists y > 0$ such that $xy = 4$."

Proof. Suppose we are given $x > 0$. Let $y = \frac{4}{x}$, then $xy = 4$, as required. ∎

Being clear about *domains* is critical. Suppose we modify the original proposition:

$$\forall x \in \mathbb{R}, \exists y \in \mathbb{R} \text{ such that } xy = 4 \tag{†}$$

Our proof now fails! The new statement (†) is *false*: indeed $x = 0$ provides a *counter-example.*

Disproof. Let $x = 0$. Since $xy = 0$ for any real number y, we cannot have $xy = 4$. ∎

Alternatively, we could *negate* (†): following Theorem 2.22, we switch the symbols $\forall \leftrightarrow \exists$ and negate the final proposition,[7]

$$\neg\big(\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, xy = 4\big) \iff \exists x \in \mathbb{R}, \forall y \in \mathbb{R},\ xy \neq 4 \tag{¬(†)}$$

Our *disproof* of (†) is really a *proof* of the negation: we provided the *example* $x = 0$, thus demonstrating the truth of a $\exists$-statement. Since the negation is true, the original (†) is false.

2. Order of quantifiers matters! The meaning of a sentence will likely change if we alter the order of quantification. This might also change the truth state of a proposition.

   (a) $\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, x^2 < y$

   Proof. Suppose a real number x is given. Let $y = x^2 + 1$, then $x^2 < y$, as required. ∎

   We proved this by viewing it as an implication, "If $x \in \mathbb{R}$, then $\exists y \in \mathbb{R}, x^2 < y$."

   (b) $\exists y \in \mathbb{R}, \forall x \in \mathbb{R}, x^2 < y$

   Disproof. We demonstrate the truth of the negation, "$\forall y, \exists x, x^2 \geq y$."
   Suppose a real number y is given. Let $x = \sqrt{|y|}$, then $x^2 = y \geq y$, as required. ∎

[7] Here is an abstract justification for this heuristic. Consider a propositional function $P(x)$: "$\exists y, Q(x, y)$," then

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

Putting it all together We finish with two examples you might have seen elsewhere. For this course, *you do not have to know what these statements mean,* though you do have to be able to *negate* them.

Examples 2.25. 1. Vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in the vector space $\mathbb{R}^3$ are *linearly independent* if

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

In a linear algebra course, the expression $\forall a, b, c \in \mathbb{R}$ would often be hidden. The negation of this statement, what it means for $\mathbf{x}, \mathbf{y}, \mathbf{z}$ to be *linearly dependent*, is

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

2. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *continuous at* $a \in \mathbb{R}$ if

$$\forall \epsilon > 0,\ \exists \delta > 0 \text{ such that } |x - a| < \delta \implies |f(x) - f(a)| < \epsilon$$

The negation, what it means for f to be *discontinuous at* $x = a$, is

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

The original statement contained a hidden quantifier $\forall x$ which became explicit upon negation.

Exercises 2.2. A self-test quiz and several worked questions can be found online.

1. Rewrite each sentence using quantifiers. Then write the negation (use 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. 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.
3. A friend claims that the sentence "$x^2 > 0 \implies x > 0$" has negation "$x^2 > 0$ and $x \leq 0$." Why is this incorrect? What is the correct negation?
4. Consider the quantified statement
$$\forall x, y, z \in \mathbb{R},\ (x-3)^2 + (y-2)^2 + (z-7)^2 > 0 \qquad (*)$$
   (a) Express $(*)$ in words.
   (b) Is $(*)$ true or false? Explain.
   (c) Express the negation of $(*)$ in symbols, and then in words.
   (d) Is the negation of $(*)$ true or false? Explain.
5. Suppose P, Q, R are propositional functions. Compute the negations of the following:
   (a) $\forall x, \exists y, P(x) \wedge Q(y)$ (b) $\forall x, \exists y, \forall z, R(x, y, z)$

6. Revisit Example 2.24.2. Decide whether each of the following is true or false:

   (a) $\exists x \in \mathbb{R}, \forall y \in \mathbb{R}, x^2 < y$ (b) $\forall y \in \mathbb{R}, \exists x \in \mathbb{R}, x^2 < y$

7. The following are statements about positive real numbers x, y. Which is true? Explain.

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

8. 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.

9. Prove or disprove the following statements.

   (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.

10. Consider the following proposition (*you do not have to know what is meant by a field*).

    All non-zero elements x in a field $\mathbb{F}$ have an inverse: some $y \in \mathbb{F}$ for which $xy = 1$.

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

11. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *decreasing* if:

    $$x \leq y \implies f(x) \geq f(y)$$

    (a) State what it means for f not to be decreasing (*where is the hidden quantifier?*)
    (b) Give an example to show that *not decreasing* and *increasing* do not mean the same thing.

12. Consider the proposition:

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

    (a) State the negation of the proposition.
    (b) Prove that the original proposition is *false.*
    (c) Suppose you rewrite the proposition:

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

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

13. (Hard) Let $(x_n) = (x_1, x_2, x_3, \ldots)$ denote a sequence of real numbers.

    "(x_n) *diverges to* ∞" means: $\forall M > 0$, $\exists N \in \mathbb{R}$ such that $n > N \implies x_n > M$
    "(x_n) *converges to* L" means: $\forall \epsilon > 0$, $\exists N \in \mathbb{R}$ such that $n > N \implies |x_n - L| < \epsilon$

    (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) (Challenge: non-examinable) Use the definitions to prove that the sequence defined by $x_n = n$ diverges to ∞, and that the sequence defined by $y_n = \frac{1}{n}$ converges to zero.

2.3 Methods of Proof

Everything thus far has been in service to what follows: to provide the language and logical fluency necessary to understand mathematical arguments and to begin to create your own. One can study foundational logic much more deeply, but that is not our purpose. The real work begins now.

In mathematics, a Theorem is[8] a justified assertion of the truth of an implication $P \implies Q$. A *proof*, for there can be many different approaches, is any logical argument which justifies the theorem. The first step in analyzing or strategizing a proof is to identify the hypothesis P and conclusion Q.

There are four standard methods of proof, though a longer argument might use several.

Direct Assume the hypothesis P and deduce the conclusion Q.[9] This structure should be intuitive, though it may help to revisit the truth table in Definition 2.5 and the tautology of Example 2.9.2.

Contrapositive Assume $\neg Q$ and deduce $\neg P$: a direct proof of the contrapositive $\neg Q \implies \neg P$. This approach works because the contrapositive is logically equivalent to $P \implies Q$ (Theorem 2.11).

Contradiction Assume P and $\neg Q$, and deduce a *contradiction*. By Theorem 2.13 and Exercise 2.1.8c, we see that $P \implies Q$ is true.

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

Each method has its advantages and disadvantages: the direct method typically has a simpler logical flow whereas the contrapositive/contradiction approaches are useful when the negations $\neg P$, $\neg Q$ are easier to work with than P, Q themselves. All methods are equally valid, and, as we'll see shortly, one can often prove a simple theorem using all three approaches!

As you work through this section, pay special attention to the logical structure—to encourage this, the mathematical level of this section is very low—and refer to the previous sections if the logical terminology feels unfamiliar. In particular, re-read *Planning and Writing Proofs* (page 4).

Direct Proofs

We begin by generalizing Example 2.21.3.

Theorem 2.26. *The product of any pair of odd integers is odd.*

To make sense of this, we first need to identify the logical structure by writing the theorem in terms of propositions and connectives. One way is to view the Theorem in the form $P(x, y) \implies Q(x, y)$:

- $P(x, y)$ is "x and y are both odd." This is our assumption, the hypothesis.
- $Q(x, y)$ is "The product of x and y is odd." This is what we wish to demonstrate, the conclusion.
- Both propositional functions are statements about *integers.* The Theorem is *universal* ("any pair"), and so contains a (hidden) quantifier $\forall x, y \in \mathbb{Z}$.

To perform a proof, we also need a clear understanding of the meaning of all necessary terms. To keep things simple, we'll take *integer* and *product* as understood and be explicit as to the meaning of *oddness.*

[8] It is sometimes awkward to fit a theorem into this format but it can always be done. Often all that is stated is the conclusion Q, in which case P would be the assertion "All mathematics we already know/assume to be true."

[9] From now on, to *assume* a proposition is to suppose its *truth.* To suppose P is false, we "assume/suppose $\neg P$."

A direct proof can be viewed as a proof sandwich, whose bread slices are the hypothesis and conclusion (P and Q): write these down as a first step. Next define any useful terms in the hypothesis. All that remains is to perform a simple calculation!

Proof. Let x and y be odd integers. (state hypothesis P)
There are integers k, l for which $x = 2k+1$ and $y = 2l+1$. Then, (definition of *odd*)

$$\begin{aligned} xy &= (2k+1)(2l+1) = 4kl + 2k + 2l + 1 \\ &= 2(2kl + k + l) + 1 \end{aligned}$$ (computation/algebra)

Since $2kl + k + l$ is an integer, we conclude that xy is odd. (state conclusion Q) ∎

To make the conclusion absolutely clear, we explicitly wrote xy in the form 2(integer)+1.

Insufficient Generality Before leaving this example, it is worth quickly highlighting the most common mistakes seen in such arguments.

Fake Proof 1. $x = 3$ and $y = 5$ are both odd, hence $xy = 15$ is odd.

This is an *example* of the theorem. Since the theorem is *universal*, a single example is not a proof.

Fake Proof 2. Let $x = 2k+1$ and $y = 2k+1$ be odd. Then

$$xy = (2k+1)(2k+1) = 2(2k^2 + 2k) + 1 \quad \text{is odd.}$$

This is still insufficiently general in that it only verifies a special case (x odd $\implies x^2$ odd). There is nothing wrong with trying out examples or sketching incomplete thoughts—indeed both are encouraged!—but you need to be aware of when your argument isn't sufficiently general.

For another simple direct proof, consider the sum of two consecutive integers.

Theorem 2.27. *The sum of any pair consecutive integers is odd.*

The theorem is again a universal claim of the form $P(x, y) \implies Q(x, y)$ about two integers:

- $P(x, y)$ is "x, y are consecutive integers."
- $Q(x, y)$ is "$x + y$ is odd."

The trick is to observe that, being consecutive, we may write y in terms of x. The proof sandwich is still visible, though it would be hard to write down the last sentence without already having settled on the trick, which is essentially the definition of "consecutive integers."

Proof. Suppose we are given two consecutive integers. Label the smaller of these x, then the other $x + 1$. But then their sum

$$x + (x + 1) = 2x + 1$$

is odd. ∎

Proof by Contrapositive

Here is another simple result about odd and even integers.

Theorem 2.28. *If the sum of two integers is odd, then they have opposite parity.*

The theorem is yet another universal statement $(\forall x, y \in \mathbb{Z})$ of the form $P(x,y) \implies Q(x,y)$:

$P(x,y)$: "$x+y$ is odd." $Q(x,y)$: "x, y have opposite parity."

Parity means *evenness or oddness*: the conclusion is that one of x, y is even and the other odd.

Attempting a direct proof results in an immediate difficulty:

Direct Proof? Suppose $x+y$ is odd. Then $x+y = 2k+1$ for some integer k...

We want to conclude something about x and y *individually*, but the direct approach lumps them together in the same algebraic expression.

A contrapositive argument suggests itself because the new hypothesis $\neg Q$, in treating x and y separately, gives us twice as much to start with.[10]

$\neg Q(x,y)$: "x, y have the same parity." $\neg P(x,y)$: "$x+y$ is even."

The minor remaining difficulty is that "same parity" encompasses *two* possibilities: x, y are either both even, or both odd. The proof therefore contains two cases.

Proof. Suppose x and y have the same parity. There are two cases. (state hypothesis $\neg Q$)

Case 1: Assume x and y are both even.
Write $x = 2k$ and $y = 2l$, for some integers k, l. (definition of *even*)
Then $x + y = 2(k+l)$ is even. (computation)

Case 2: Assume x and y are both odd.
Write $x = 2k+1$ and $y = 2l+1$ for some integers k, l. (definition of *odd*)
Then $x + y = 2(k+l+1)$ is even. (computation)

In both cases $x+y$ is even. (state conclusion $\neg P$) ∎

Again observe the proof sandwich and how the proof depending on little more than the definitions of even and odd.

When presenting a lengthier argument, consider orienting the reader by starting with the phrase, "We prove the contrapositive." In simple cases like the above this is unnecessary, since the logical structure should be completely clear without such assistance. It is also unnecessary to define and spell out the propositions P and Q or include any of the bracketed commentary. However, you should feel free to continue this practice if you think it would aid your explanation, of if you are nervous about your proof skills.[11]

[10] Warning! The contrapositive is still a *universal* statement: $\forall x, y \in \mathbb{Z}, \neg Q(x,y) \implies \neg P(x,y)$. We are not negating the whole theorem so do not convert $\forall$ to $\exists$!

[11] ...and want to guarantee some partial credit!

For another example of a contrapositive argument, we extend the first result of this section.

Theorem 2.29. *The product of two integers is odd* if and only if *both integers are odd.*

This has the form $P \iff Q$, which comprises *two theorems in one*: $P \implies Q$ and $Q \implies P$ (see Exercise 2.1.9a). A contrapositive argument for the ($\Rightarrow$) direction is again suggested because the right hand side treats the two integers separately.

Proof. ($\Rightarrow$) We prove the contrapositive. Let x, y be integers, at least one of which is even. Suppose, *without loss of generality*, that $x = 2k$ is even. Then $xy = 2ky$ is also even.

($\Leftarrow$) This is precisely Theorem 2.26, which we've already proved. ∎

Without loss of generality Often abbreviated to WLOG, this phrase is common in mathematics. Here it saves us from performing the almost identical argument assuming $y = 2l$ is even. WLOG is stated when a mathematician makes a choice which does not materially affect the argument.

Proof by Contradiction

To introduce contradiction proofs consider another simple result.[12]

Example 2.30. Let x be an integer. If $3x + 5$ is even, then $5x + 2$ is odd.

We could proceed directly according to the following sketch:

$$3x + 5 \text{ even} \implies 3x \text{ odd} \implies x \text{ odd} \implies 5x \text{ odd} \implies 5x + 2 \text{ odd} \qquad (*)$$

This isn't wrong! You should believe each implication; indeed we've proved most of them. However it would be nice not to rely on so many other results. A similar contrapositive approach (reverse arrows and negate propositions in $(*)$) would have the same weakness.

The advantage of a contradiction approach is that we have twice as much to work with: the hypothesis ($3x + 5$ even) *and* the negation of the conclusion ($5x + 2$ even).

Proof. Suppose both $3x + 5$ and $5x + 2$ are even. Then their sum is also even. However,

$$(3x + 5) + (5x + 2) = 8x + 7 = 2(4x + 3) + 1 \qquad (\dagger)$$

is odd. Contradiction (an integer cannot be both even and odd!). ∎

Remember to mention the word *contradiction* at the end, so the reader knows what you've done.

A nice side-effect of this approach is that it suggests an alternative *direct proof.*

Direct Proof. For any integer x, ($\dagger$) says that $3x + 5$ and $5x + 2$, in summing to an odd number, have *opposite parity.* ∎

The last argument in fact proves that $3x + 5$ is even *if and only if* $5x + 2$ is odd: the converse of our claim comes for free! Look back at $(*)$: you should believe that all the arrows are reversible.

[12] While what follows is a theorem, we'll typically reserve the word for results that are worth remembering in their own right. Examples like this are good for practice (change the numbers!), but are not individually very interesting.

Such variety is one of the things that makes proving theorems fun! While your choice of proof is largely a matter of personal taste, remember your audience. The final argument is very slick but might risk confusing a reader rather than empowering them.[13]

Three Proofs of the Same Result We finish this section with three proofs of the same result. All are based on the same factorization of a polynomial

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

and the well-known fact that $ab = 0 \iff a = 0$ or $b = 0$ (see Exercise 14). Since the mathematics is so simple, pay attention to and compare the *logical structures.*

Example 2.31. Let x be a real number. Then $x^3 + 2x^2 - 3x - 10 = 0 \implies x = 2$.

Direct Proof. Suppose $x^3 + 2x^2 - 3x - 10 = 0$. By factorization, $(x-2)[(x+2)^2+1] = 0$, so at least one of the factors must be zero. Since $(x+2)^2 + 1 \geq 1 > 0$, we conclude that $x - 2 = 0$, from which $x = 2$. ■

Contrapositive Proof. Suppose $x \neq 2$. Since $(x+2)^2 + 1 \geq 1 > 0$, we see that

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

■

Contradiction Proof. Suppose $x^3 + 2x^2 - 3x - 10 = 0$ 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 $(x+2)^2 + 1 = 0$. However, $(x+2)^2 + 1 \geq 1$ for all real numbers x, so we have a contradiction. ■

Exercises 2.3. A self-test quiz and several worked questions can be found online.

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.
2. For any given integers a, b, c, if a is even and b is odd, prove that $7a - ab + 12c + b^2 + 4$ is odd.
3. Prove that if n is an integer greater than 1, then $n! + 2$ is even.
   ($n! = n(n-1)(n-2)\cdots 1$ *is the* factorial *of the integer* n)
4. (a) Let $x \in \mathbb{Z}$. Prove that $5x + 3$ is even if and only if $7x - 2$ is odd.
   (b) Can you conclude anything about $7x - 2$ if $5x + 3$ is odd?

[13] The Hungarian mathematician Paul Erdős referred to simple, elegant proofs as 'from the Book,' as if the Almighty kept a tome of perfect proofs. As with all matters spiritual, one person's Book might be very different to another's…

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

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

   (a) Is the proposition true or false? Justify your answer. Is its converse true?
   (b) Repeat part (a) for the proposition $x^3 - 3x^2 - 2x + 6 = 0 \implies x \neq 3$.

6. 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. ∎

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

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

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

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

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

Proof. Suppose that m^2 is even and m is odd. Write $m = 2k + 1$ for some integer k. Then

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

is odd. Contradiction. ∎

9. Here is a 'proof' that every real number x equals zero. Find the mistake.

$$\begin{aligned} x = y &\implies x^2 = xy \implies x^2 - y^2 = xy - y^2 \\ &\implies (x-y)(x+y) = (x-y)y \\ &\implies x + y = y \\ &\implies x = 0 \end{aligned}$$

10. Prove or disprove: An integer n is even if and only if n^3 is even.

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

12. Let x and y be integers. Prove $x^2 + y^2$ is even **if and only if** x and y have the same parity.

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

14. Suppose $a, b \in \mathbb{R}$. Prove that $ab = 0 \iff a = 0$ or $b = 0$.

15. 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 Further Proofs & Strategies

The arguments in this section are slightly trickier and more representative of typical mathematical proofs. Some of these results are famous and worth knowing in their own right. We will also introduce *lemmas* and *corollaries*, which are used to break up the presentation of complex results.

Proving Universal Statements

Most of the results we've seen thus far have been universal statements. As discussed on page 14, any theorem $P(x) \implies Q(x)$ is implicitly universal, albeit with the quantifier $\forall x$ being typically hidden. Revisit Examples 2.24 on multiple quantifiers so see how these fit into our proof framework; here is another example.

Example 2.32. $\forall x \geq 0, \exists y < 0$ such that $x^3 < y^2$

View the claim as an implication "$x \geq 0 \implies (\exists y < 0$ such that $x^3 < y^2)$" and prove directly.

Proof. Suppose $x \geq 0$ is given. *Define* $y = -\sqrt{x^3} - 1$, then $y \leq -1 < 0$ and

$$y^2 = x^3 + 1 + 2\sqrt{x^3} \geq x^3 + 1 > x^3$$

■

Notice how y, in being existentially quantified *after* x, is allowed to depend on x. We'll revisit this shortly to see what happens when we reverse the quantifiers. Remember that you might need some scratch work to find a suitable y: you shouldn't (yet!) expect to create such an argument in one shot.

For a more involved example of a universal result, here is a famous inequality relating the arithmetic and geometric means of two numbers.

Theorem 2.33 (AM–GM inequality). *If x, y are non-negative real numbers, then*

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

with equality if and only if $x = y$.

If your faith is wavering, first try an example: e.g., $\frac{3+5}{2} = 4 \geq \sqrt{15}$. It should also be clear that both sides are equal whenever $y = x$. The logical structure $\forall x, y \geq 0, Q(x, y)$ can be viewed as an implication $P(x, y) \implies Q(x, y)$, where $P(x, y)$ is "$x, y \geq 0$." The real challenge is making sense of $Q(x, y)$. There are really two separate results here:

1. If $x, y \geq 0$, then $\frac{x+y}{2} \geq \sqrt{xy}$
2. If $x, y \geq 0$, then $\frac{x+y}{2} = \sqrt{xy} \iff x = y$

Concentrate on the first since it is simpler. The hypothesis $(x, y \geq 0)$ doesn't give us much to work with, so it seems sensible to play with the inequality and try to get rid of the ugly square-root:

$$\frac{x+y}{2} \geq \sqrt{xy} \implies (x+y)^2 \geq 4xy \implies x^2 - 2xy + y^2 \geq 0 \implies (x-y)^2 \geq 0$$

Now we have something we believe! The question is whether we can reverse the arrows. Only the first should give you any pause: it is here that we use the non-negativity of x, y.

Proof. Suppose $x, y \geq 0$. Multiply out a trivial inequality:

$$\begin{aligned}(x-y)^2 \geq 0 &\iff x^2 - 2xy + y^2 \geq 0 \iff x^2 + 2xy + y^2 \geq 4xy \\ &\iff (x+y)^2 \geq 4xy \\ &\iff \frac{x+y}{2} \geq \sqrt{xy}\end{aligned}$$

The square-root is well-defined because $x, y \geq 0$, and the inequality is preserved since the square-root function is *increasing*. For the second result, observe that the final inequality is an *equality* precisely when *all* the inequalities are equalities; this is if and only if $x = y$. ∎

The scratch work really helped us figure out how and where to apply the hypothesis. Notice also how the second result came almost for free! Result 1 only need the ($\Rightarrow$) in the proof, but the second result needs them all to be biconditional.

For variety, here is a contradiction proof incorporating the same calculations in a different order.

Contradiction Proof. Let $x, y \geq 0$ and suppose that $\frac{x+y}{2} < \sqrt{xy}$. Since $x + y \geq 0$, the second inequality holds if and only if $(x+y)^2 < 4xy$. Now multiply out and rearrange:

$$\begin{aligned}(x+y)^2 < 4xy &\iff x^2 + 2xy + y^2 < 4xy \\ &\iff x^2 - 2xy + y^2 < 0 \\ &\iff (x-y)^2 < 0\end{aligned}$$

Contradiction (squares of real numbers are non-negative). We conclude that $\frac{x+y}{2} \geq \sqrt{xy}$.
Now suppose that $\frac{x+y}{2} = \sqrt{xy}$. Following the biconditionals in the above argument, we see that equality holds if and only if $(x-y)^2 = 0$, from which we recover $x = y$. ∎

The AM–GM inequality in fact holds for any finite collection of non-negative numbers $x_1, \ldots, x_n$:

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \cdots x_n}$$

with equality if and only if all the x_i are equal. Proving this is a lot harder (see Exercise 14).

Disproving Existential Statements: Non-existence Proofs

Of the four possible combinations "prove/disprove a universal/existential statement," we've now tackled three; what remains is to consider how to prove that something does not, or cannot, exist. Recall our basic rule of negation:

$$\neg\big(\exists x, Q(x)\big) \iff \forall x, \neg Q(x)$$

To show that $\exists x, Q(x)$ is *false* is therefore to prove a *universal* statement.[14] As before (page 14), let $P(x)$ be the proposition "x lies in the domain of Q." We conclude that

$\neg\big(\exists x, Q(x)\big)$ is logically equivalent to $P(x) \implies \neg Q(x)$

[14] The notation $\nexists x, Q(x)$ is discouraged because it obscures this essential fact.

Once viewed as an implication, any of our proof strategies might be applicable. Contradiction and contrapositive arguments are particularly common however, since the right hand side is already a *negative* statement.

Example 2.34. The equation $x^{17} + 12x^3 + 13x + 3 = 0$ has no positive (real number) solutions.

Before seeing a proof, consider several ways in which this claim could be presented.

Non-existence ($\neg(\exists x, Q(x))$)	There are no $x > 0$ for which $x^{17} + 12x^3 + 13x + 3 = 0$.
Universal ($\forall x, \neg Q(x)$)	For all $x > 0$, we have $x^{17} + 12x^3 + 13x + 3 \neq 0$.
Direct ($P \Rightarrow \neg Q$)	If $x > 0$, then $x^{17} + 12x^3 + 13x + 3 \neq 0$.
Contrapositive ($Q \Rightarrow \neg P$)	If $x^{17} + 12x^3 + 13x + 3 = 0$, then $x \leq 0$.
Contradiction ($P \wedge Q$)	$x > 0$ and $x^{17} + 12x^3 + 13x + 3 = 0$ is *impossible.*

We present two very similar arguments based on the direct and contradiction structures.

Direct proof. Suppose that $x > 0$. But then $x^{17} + 12x^3 + 13x + 3 > 0$ since all terms are positive. We conclude that $x^{17} + 12x^3 + 13x + 3 \neq 0$. ∎

Contradiction proof. Assume that $x > 0$ satisfies $x^{17} + 12x^3 + 13x + 3 = 0$. Since all terms on the left hand side are positive, we have a contradiction. ∎

In practice, some version of one of these arguments would likely be given without any additional commentary. The reader is assumed to be familiar with the underlying logic without it being spelled out. Our discussion could be considered scratch work.

Subdividing Theorems: Lemmas & Corollaries

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. Often the intention is to improve the readability of a difficult/complex argument, but you may also wish to (de-)emphasize the relative importance of certain results. Mathematics does this by using *lemmas* and *corollaries.*

Lemma: A theorem whose importance you want to downplay or which will be used when proving a more significant result.

Corollary: A theorem which follows quickly once you understand another result, perhaps as a special case or by modifying the proof in some small way.

Presentation style varies hugely between authors and journals: some reserve *theorem* only for the most important results, with everything else presented as a lemma or corollary, while others never use these terms (or just call everything a *proposition*!). Regardless, lemmas and corollaries are useful to have in your toolkit if readability is your goal.

In preparation for our next, much more important, result, here is a simple lemma.

Lemma 2.35. *Suppose n is an integer. Then n^2 is even $\iff$ n is even.*

At this stage, you should be able to prove this yourself. This is really just a special case of Theorem 2.29: if you're completely unsure how to start, revisit that result, and the rest of Section 2.3.

Irrational Numbers

Since their definition is inherently negative, irrational numbers provide good examples of non-existence/contradiction arguments. They are also interesting in their own right!

Definition 2.36. A real number x is *rational* if it may be written in the form $x = \frac{m}{n}$ for some integers m, n. A real number is *irrational* if no such integers exist.

You likely know of a few irrational numbers ($\sqrt{2}, \pi, e$), but how do we *prove* that a given number is irrational? Our next result is very famous, with versions dating back at least to Aristotle (c. 340 BCE).

Theorem 2.37. *$\sqrt{2}$ is irrational.*

We must *disprove* the existence claim $\exists m, n \in \mathbb{Z}, \sqrt{2} = \frac{m}{n}$. As before, consider several restatements:

Non-existence	There are no integers m, n for which $\sqrt{2} = \frac{m}{n}$.
Universal	For all integers m, n, we have $\sqrt{2} \neq \frac{m}{n}$.
Direct	If $m, n \in \mathbb{Z}$, then $\sqrt{2} \neq \frac{m}{n}$.
Contrapositive	If $\sqrt{2} = \frac{m}{n}$, then m, n are not both integers.
Contradiction	$m, n \in \mathbb{Z}$ and $\sqrt{2} = \frac{m}{n}$ is *impossible.*

Can you see the obvious drawbacks of a direct or contrapositive approach? We prove by contradiction, while outsourcing a repeated step to the ($\Rightarrow$) direction of Lemma 2.35 to improve readability.

Proof. Suppose that $m, n \in \mathbb{Z}$ and that $\sqrt{2} = \frac{m}{n}$. *Without loss of generality*, assume that m, n have no common factors. Cross-multiply and square:

$$m^2 = 2n^2 \text{ is even} \implies m \text{ is even} \qquad \text{(Lemma 2.35)}$$

whence $m = 2k$ for some integer k. But then

$$2n^2 = m^2 = 4k^2 \implies n^2 = 2k^2 \text{ is even} \implies n \text{ is even} \qquad \text{(Lemma 2.35)}$$

We see that m and n have a common factor of 2. Contradiction. ∎

The major difficulty lies in the assumption that m, n have no common factors. In the same way that we can simplify $\frac{4}{6} = \frac{2}{3}$, our assumption is *without loss of generality* because it costs us nothing *once we assume $\sqrt{2} = \frac{m}{n}$ is rational.* It is crucial to appreciate that we aren't contradicting the assumption that m, n have no common factors, lest our calculation continue forever without resolution!

$$m^2 = 2n^2 \implies n^2 = 2k^2 \implies k^2 = 2l^2 \implies \cdots$$

The irrationality of $\sqrt{3}$, $\sqrt[3]{2}$, etc., can be proved similarly (π and e are *much* harder!). Now we have the theorem, it is easily applied to demonstrate the irrationality of many other numbers.

Example 2.38. Suppose that $\sqrt{2} - 5\sqrt{3} = x$ were rational: $\exists m, n \in \mathbb{Z}$ such that $x = \frac{m}{n}$. Then

$$75 = (5\sqrt{3})^2 = (\sqrt{2} - x)^2 = 2 + x^2 - 2\sqrt{2}x \implies \sqrt{2} = \frac{x^2 - 73}{2x} = \frac{m^2 - 73n^2}{2mn}$$

Otherwise said, $\sqrt{2}$ is *rational*: contradiction.

Non-constructive Existence Proofs

Every existence proof we've seen so far has been *constructive*: we exhibit/construct an *explicit example* x for which $Q(x)$ is true. Sometimes, however, this is expecting a bit too much. It is often far easier to show the existence of something *without* explicitly stating what it is. We present two famous examples of this situation.

Theorem 2.39. *There are irrational numbers a, b for which a^b is rational.*

Proof. Consider the number $x = (\sqrt{2})^{\sqrt{2}}$. There are two possibilities:

1. x is rational. Let $a = b = \sqrt{2}$.
2. x is irrational. Let $a = x$ and $b = \sqrt{2}$ and apply the usual exponential laws to see that

$$a^b = \left((\sqrt{2})^{\sqrt{2}}\right)^{\sqrt{2}} = (\sqrt{2})^{\sqrt{2}\cdot\sqrt{2}} = (\sqrt{2})^2 = 2$$

In either case, a, b are irrational and a^b rational. ∎

The proof is very sneaky: it does not provide an explicit example and does not tell us whether $(\sqrt{2})^{\sqrt{2}}$ is rational. In fact this number isn't rational, though demonstrating it is massively harder.[15]

We finish with a particularly famous example of a non-constructive existence proof found in Euclid's *Elements* (300 BCE), the most influential textbook of mathematical history. As ever, we need a solid definition before we try to prove anything.

Definition 2.40. 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, though it is not completely obvious, that every integer ≥ 2 is either prime or a product of primes (*composite*). In particular, every integer ≥ 2 is divisible by at least one prime. We now state Euclid's result, and prove it by contradiction.

Theorem 2.41 (*Elements*, Book IX, Prop. 20). *There are infinitely many prime numbers.*

Proof. Assume there are exactly n primes $p_1, \ldots, p_n$ and define the integer

$$\Pi := p_1 \cdots p_n + 1$$

Certainly Π is divisible by some prime p_i (in our list by assumption!), as is the product $p_1 \cdots p_n$. But then the difference

$$1 = \Pi - p_1 \cdots p_n$$

is divisible by p_i, contradicting the fact that $p_i \geq 2$. ∎

[15] If you're interested, look up the Gelfond–Schneider Theorem (1934), Hilbert's Seventh Problem, and what they say about *algebraic* and *transcendental numbers*. Such ideas are far beyond the level of this text!

Exercises 2.4. A self-test quiz and worked questions can be found online.

1. Prove or disprove:
   (a) There exist integers m and n such that $2m - 3n = 15$.
   (b) There exist integers m and n such that $6m - 3n = 11$.
2. Prove or disprove: There exists a line L in the plane such that, for all points A, B in the plane, we have that A, B lie on L.
3. Prove: For every positive integer n, the integer $n^2 + n + 3$ is odd and greater than or equal to 5.
4. Let p be an odd integer. Prove that the equation $x^2 - x - p = 0$ has no *integer* solutions.
5. Prove or disprove: $\sqrt{2} - \sqrt{7}$ is rational.
6. Prove or disprove the following conjectures about real numbers x, y.
   (a) If $3x + 5y$ is irrational, then at least one of x and y is irrational.
   (*Be careful! This isn't a logical 'and': what happens when you negate?*)
   (b) If x and y are rational, then $3x + 4xy + 2y$ is rational.
   (c) If x and y are irrational, then $3x + 4xy + 2y$ is irrational.
7. Prove by contradiction: if x and y are positive real numbers, then $\sqrt{x+y} \neq \sqrt{x} + \sqrt{y}$.
   How would you change things to make a *contrapositive* argument?
8. Prove that between any two distinct rational numbers there exists another rational number.
9. Consider the proposition:

   For any non-zero rational r and any irrational number t, the number rt is irrational.

   (a) Translate this statement into logic using quantifiers and propositional functions.
   (b) Prove the statement.
10. (a) An integer n is *not divisible by 3* if and only if $\exists k \in \mathbb{Z}$ for which $n = 3k + 1$ or $3k - 1$.
    Prove that if n^2 is divisible by 3, then n is divisible by 3.
    (b) Prove that $\sqrt{3}$ is irrational.
    (c) Prove that $\sqrt[3]{2}$ is irrational. (*Hint: revisit Exercise 2.3.10*)
11. (Recall Example 2.34) You are given the following facts:
    - All polynomials are continuous.
    - (Intermediate Value Theorem) If f is continuous on the interval $[a, b]$ and L lies between $f(a)$ and $f(b)$, then there is some x in the interval (a, b) for which $f(x) = L$.
    - If $f'(x) > 0$ on an interval, then f is an increasing function.

    Use these facts to give formal proofs of two claims that should be familiar from calculus:
    (a) $x^{17} + 12x^3 + 13x + 3 = 0$ has a solution x in the interval $(-1, 0)$.
    (b) $x^{17} + 12x^3 + 13x + 3 = 0$ has *exactly one* real number solution x.

12. The real numbers satisfy the *Archimedean property*:

    For any $x, y > 0$, there exists a positive integer n such that $nx > y$.

    (a) Use the Archimedean property to show that there are no positive real numbers which are less than $\frac{1}{n}$ for all positive integers n.

    (b) 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 argument? Give a correct proof.

    (c) Suppose $x < y$ are real numbers. Prove that there exists a positive integer n for which $n(y - x) > 1$.

    (d) (Hard) Prove: $\forall x, y \in \mathbb{R}$ with $x < y$, $\exists m, n \in \mathbb{Z}$ for which $nx < m < ny$. Hence conclude an extension of Exercise 8: between any two *real* numbers there exists a rational number.

13. Suppose $x, y, z \geq 0$ satisfy $x + y + z = 1$. Use the AM–GM inequality (two-variable or full version) to answer the following.

    (a) What is the largest possible value of xyz and when does it occur?

    (b) (Hard) Prove that $(1 - x)(1 - y)(1 - z) \geq 8xyz$.

14. (Hard) We prove the full AM–GM inequality.

    (a) When $n = 3$, try mimicking our earlier approach by cubing the desired inequality. Why does this seem unwise?

    (b) Prove that $x \leq e^{x-1}$ for all real numbers x, with equality if and only if $x = 1$.
    (*Hint: Consider $f(x) = e^{x-1} - x$ and apply a derivative test from calculus*)

    (c) Let $\mu = \frac{x_1 + x_2 + \cdots + x_n}{n}$ be the arithmetic mean. Apply part (b) to each expression $x = \frac{x_i}{\mu}$ to conclude that $x_1 \cdots x_n \leq \mu^n$ and hence complete the proof.

3 Divisibility and the Euclidean Algorithm

In this section we introduce *congruence,* which generalizes the notion of an integer's parity (evenness/oddness). The study of congruence is of fundamental importance to the sub-discipline of number theory, and provides some of the most straightforward examples of groups and rings. We will cover the basics in this section, returning in Chapter 7 for more formal observations.

3.1 Divisors, Remainders and Congruence

Definition 3.1. Let m, n be integers. The proposition $n \mid m$ is read "n *divides* m," and means

$$\exists k \in \mathbb{Z} \text{ such that } m = kn$$

We could also say that "n is a *divisor* of m," that "m is *divisible* by n," or that "m is a *multiple* of n."

The negated symbol $\nmid$ is read *does not divide.*

Examples 3.2. 1. We write $4 \mid 20$ since $20 = 4 \times 5$. The same equation also says that $5 \mid 20$.

2. The proposition $9 \nmid 7$ is read "*9 does not divide 7.*" It is shorthand for $\neg(9 \mid 7)$.

When integers do not divide, there is a *remainder* left over. Your study of remainders likely goes back to elementary school when you first learned division: for instance,[16]

$$33 \div 5 = 6 \,\mathsf{r}\, 3 \qquad \text{("6 remainder 3")}$$

An important foundational result says that unique remainders always exist.

Theorem 3.3 (Division Algorithm). *Suppose $m, n \in \mathbb{Z}$ with n positive. Then there exist* unique *integers q, r (the* quotient *and* remainder*) for which*

$$m = qn + r \quad \textit{and} \quad 0 \leq r < n$$

In elementary school language, $m \div n = q \,\mathsf{r}\, r$.

Examples 3.4. 1. Given $m = 23$ and $n = 7$, we have $23 = 3 \cdot 7 + 2$; that is $q = 3$ and $r = 2$.

2. If $m = -11$ and $n = 3$, then $-11 = (-4) \cdot 3 + 1$; that is $q = -4$ and $r = 1$.

3. The formula $m = 6q + 4$, where $q \in \mathbb{Z}$, describes all integers with remainder 4 on division by 6.

An *algorithm* is typically a computational process: if $m > 0$ one could view this as the repeated subtraction of n from m until the result $r = m - qn$ satisfies $0 \leq r < n$. A rigorous proof requires foundational ideas related to induction to which we will return in Chapter 5. For our current purposes, we just need to know that remainders exist. Indeed our next step is to find a way to compare remainders *without* explicitly invoking the division algorithm.

[16] The common meaning of *divide* is to apportion a quantity equally. Thus to divide 33 apples between 5 people, each person gets 6 apples and 3 are left over. In grade school mathematics, fractions come much later.

Definition 3.5. Let a, b and n be integers with n positive. The proposition

$$a \equiv b \pmod{n} \qquad \text{“}a \text{ is congruent to } b \text{ modulo } n\text{”}$$

means that a and b have the *same remainder* on division by n. The integer n is called the *modulus.*

Examples 3.6. Consider remainders modulo 3 (division by 3).

1. We write $7 \equiv 10 \pmod{3}$, since $7 = 2 \cdot 3 + 1$ and $13 = 4 \cdot 3 + 1$ have the same remainder ($r = 1$).
2. We write $6 \not\equiv 10 \pmod{3}$, since $6 = 2 \cdot 3 + 0$ and $17 = 5 \cdot 3 + 2$ have different remainders.

Calculating using the division algorithm is tedious. Our next result is crucial in that it permits the direct comparison of remainders. This can be treated as an equivalent definition of congruence.

Theorem 3.7. $a \equiv b \pmod{n} \iff n \mid (b - a) \iff b = a + kn$ *for some integer* k

Proof. The second biconditional is nothing more than an application of Definition 3.1:

$$\begin{aligned} n \mid (b-a) &\iff \exists k \in \mathbb{Z} \text{ such that } b - a = kn \\ &\iff b = a + kn \text{ for some integer } k \end{aligned}$$

Before presenting direct arguments for each direction of the first biconditional, it is helpful to introduce notation from the division algorithm:

$$\begin{aligned} & a = q_1 n + r_1 \qquad b = q_2 n + r_2 \qquad 0 \le r_1, r_2 < n \\ \implies & b - a = (q_2 - q_1)n + (r_2 - r_1) \end{aligned} \tag{$*$}$$

($\Rightarrow$) If $a \equiv b \pmod{n}$, then a, b have the same remainder $r_1 = r_2$. But then $(*)$ says that $n \mid (b - a)$.

($\Leftarrow$) Assume that $n \mid (b - a)$ so that $b - a = kn$ for some integer k. By $(*)$, we see that

$$r_2 - r_1 = (b - a) - (q_2 - q_1)n = (k - q_2 + q_1)n$$

is divisible by n. Since the remainders satisfy $0 \le r_1, r_2 < n$, we moreover see that

$$-n < r_2 - r_1 < n$$

The only possibility is $r_2 - r_1 = 0$. Otherwise said, a, b have the same remainder: $a \equiv b \pmod{n}$. ∎

If you're having trouble with the last step, think about an example! Suppose $n = 26$ and write $x = r_2 - r_1$. Hopefully you believe that $x = 0$ is the only *integer* satisfying the two conditions,

$$x \text{ is divisible by } 26 \quad \text{and} \quad -26 < x < 26$$

Since the result is abstract, it is good to recap the relationship between congruence and divisibility.

- Each $a \in \mathbb{Z}$ is congruent to *exactly one* of the integers $0, 1, 2, \ldots, n - 1$ modulo n: its *remainder*.
- a is divisible by n if and only if $a \equiv 0 \pmod{n}$.

Examples 3.8. 1. We describe all integers x which are congruent to 7 on division by 11:

$$x \equiv 7 \pmod{11} \iff 11 \mid (x-7) \iff x-7 = 11k \iff x = 11k+7$$

for some integer k.

2. To get more of a feel for the notation, consider the following conjectures:

   (a) $a \equiv 8 \pmod{6} \implies a \equiv 2 \pmod{3}$

   (b) $a \equiv 2 \pmod{3} \implies a \equiv 8 \pmod{6}$

   Conjecture (a) is true. If $a \equiv 8 \pmod{6}$, then $a = 6k+8$ for some integer k, from which

$$a = 6k+8 = 3(2k+2)+2 \implies a \equiv 2 \pmod{3}$$

   Conjecture (b) is false. The *counter-example* $a = 5$ disproves this (universal) claim:

$$5 \equiv 2 \pmod{3} \quad \text{and} \quad 5 \not\equiv 8 \pmod{6}$$

Modular Arithmetic

Remainders have a natural arithmetic that is very similar to that of the real numbers. We the same symbols; even the congruence symbol $\equiv$ looks a bit like an equals sign![17] Apart from being fun, *modular arithmetic* has many applications, including cryptography and data security: cell-phones and computers perform millions of these calculations every day! Here are the basic rules, which generalize of most of the results we proved in Section 2.3.

Theorem 3.9. *Suppose a, b, c, d are integers and that n is some modulus. Then,*

1. *If $a \equiv c$ and $b \equiv d$ modulo n, then*

$$a \pm b \equiv c \pm d \quad \text{and} \quad ab \equiv cd \pmod{n}$$

2. *The usual associative, commutative and distributive laws of arithmetic hold for congruences:*

$$a+(b+c) \equiv (a+b)+c \qquad a+b \equiv b+a \qquad a(b+c) \equiv ab+ac$$
$$a(bc) \equiv (ab)c \qquad ab \equiv ba$$

The theorem says that the operations 'take the remainder' and 'add/subtract/multiply' can be performed in any order or combination, the result will be the same.

Example 3.10. Find the remainder when $29+14$ is divided by 6. We do this in two ways:

(a) First find the sum 43, then compute its remainder: $43 \equiv 1 \pmod{6}$ since $6 \mid (43-1)$.

(b) Alternatively, we could find the remainder of each component and then add:

$$29+14 \equiv 5+2 \equiv 7 \equiv 1 \pmod{6}$$

[17] This is no accident. In Chapter 7 we'll see that congruence is an important example of an *equivalence relation:* a generalized notion of equality. Indeed, two integers are congruent if and only if something about them is equal: their *remainders*!

Proof. 1. We prove the multiplication rule. Suppose that $a \equiv c$ and $b \equiv d$. By Theorem 3.7, we have $c = a + kn$ and $d = b + ln$ for some integers k, l. Now compute:

$$cd - ab = (a + kn)(b + ln) - ab = (bk + al + kln)n$$

is divisible by n, whence $ab \equiv cd$. The addition/subtraction argument is almost identical.

2. The associative, commutative and distributive laws hold because $x = y \implies x \equiv y \pmod{n}$, regardless of n (equal numbers have the same remainder!). ∎

The ability to take remainders *before* adding and multiplying is remarkably powerful, and allows us rapidly to perform some surprising calculations.

Examples 3.11. 1. Find the remainder when 37^{423} is divided by 10.

The sheer size of 37^{423} makes this appear impossible at first glance.[18] Instead we think about the rules of arithmetic modulo 10. Since $37 \equiv 7 \equiv -3 \pmod{10}$, we see that

$$37 \cdot 37 \equiv (-3) \cdot (-3) \equiv 9 \equiv -1 \pmod{10}$$

This is more promising, for we can use it to simplify the original expression:

$$37^{423} \equiv \underbrace{(-3) \cdot (-3) \cdots (-3)}_{423 \text{ times}} \equiv \left((-3)^2\right)^{211}(-3) \equiv (-1)^{211}(-3) \equiv 3 \pmod{10}$$

2. Here we compute modulo $n = 6$:

$$7^9 + 14^3 \equiv 1^9 + 2^3 \equiv 1 + 8 \equiv 9 \equiv 3$$

It would have been madness to compute $7^9 + 14^3 = 40356351$ before finding the remainder!

3. Find the remainder when $124^{12} \cdot 65^{49}$ is divided by 11.

This needs several steps and simplifications. Since $124 = 11^2 + 3$ and $65 = 11 \cdot 6 - 1$, we write

$$\begin{aligned} 124^{12} \cdot 65^{49} &\equiv 3^{12} \cdot (-1)^{49} \equiv (3^3)^4 \cdot (-1) \\ &\equiv -(27^4) \equiv -(5^4) \\ &\equiv -(25^2) \equiv -(3^2) \equiv 2 \pmod{11} \end{aligned}$$

When performing these calculations:

- Replace each integer by something *small* with the same remainder: $37 \equiv -3 \pmod{10}$ is more helpful than $37 \equiv 7 \pmod{10}$, since powers of -3 are much easier to work with.
- The base of an exponential expression can be reduced, but *not* the exponent: $17^{23} \equiv 3^{23} \pmod{7}$ is correct, but $3^{23} \not\equiv 3^2 \pmod{7}$. Exponentiation is just shorthand for repeated multiplication.

[18] Using logarithms, a pocket calculator will tell you that $37^{423} \approx 2.2 \times 10^{663}$ has 663 digits! This is no help since what we want is the *units* digit, not its largest few significant figures.

Application: On what day were you born?

While we all know our *date* of birth, most of us do not know on which *day* of the week we were born. You can answer this question quite easily (perhaps in your head!) using modular arithmetic.

- Since $365 \equiv 1 \pmod 7$, a standard year advances the calendar one weekday.
- Each leap year[19] advances the calendar an additional day.

Can you figure the weekday today's date in your year of birth? Thinking about the length of each month modulo 7, you should also be able to find your birth*day*.

Example 3.12. Paul Revere was born January 1st, 1735, in Boston. Given that January 1st, 2024 was a Monday, find the weekday of Revere's birth.

The dates differ by 289 years, in which time there have been $\frac{288}{4} - 2 = 70$ leap years (not 1800 and 1900). The calendar has therefore advanced $289 + 70 \equiv 2$ weekdays: Revere was born on a Saturday.

Division and Congruence Equations

Division in modular arithmetic behaves in unexpected ways. We'll consider this further in Exercise 3.2.15, but for the present it is safest to convert congruences to statements about integers.

Examples 3.13. 1. Even when there is a common factor, dividing both sides is perilous. For instance

$$\begin{aligned} 42 \equiv 12 \pmod{10} &\implies \exists k \in \mathbb{Z},\ 42 - 12 = 10k \implies \exists k \in \mathbb{Z},\ 21 - 6 = 5k \\ &\implies 21 \equiv 6 \pmod 5 \end{aligned}$$

Division by 2 also divided the modulus! If we hadn't divided the modulus, the result would be *false*: $21 \not\equiv 6 \pmod{10}$.

2. Congruence equations are much harder to solve than standard equations. For instance, we cannot solve $2x \equiv 7 \pmod 9$ by division: $x \equiv \frac{7}{2}$ is meaningless since $\frac{7}{2}$ is not an integer!

   It won't always work, but in this case sneakily *multiplying* by 5 solves the problem:

$$2x \equiv 7 \implies 10x \equiv 35 \implies x \equiv 8 \pmod 9$$

Exercises 3.1. A reading quiz and several questions with linked video solutions can be found online.

1. Prove, using the definition of "divides," that $n \mid a$ and $n \mid b \implies n \mid (a+b)$.
2. Let a, b, c be integers. Prove or disprove each of the following claims:

   (a) $a \mid b$ and $b \mid c \implies a \mid c$ (b) $a = b \iff a \mid b$ and $b \mid a$

   (b) $a \mid b$ and $a \mid c \implies a \mid bc$ (d) $a \mid c$ and $b \mid c \implies ab \mid c$
3. List all integers x for which $x \equiv 3 \pmod 5$ and $-10 \le x \le 20$.
4. Use the division algorithm to prove that if p is an odd prime, then $p \equiv 1$ or $p \equiv 3 \pmod 4$.

[19] Leap years occur whenever the year is divisible by 4. Among centuries, *only* years divisible by 400 are leap years: thus 1900 wasn't a leap year but 2000 was. This is only practical back to the invention of the Gregorian calendar in the 1600s.

5. Prove the first part of Theorem 3.9: if $a \equiv c$ and $b \equiv d$, then $a + b \equiv c + d \pmod{n}$.

6. Find a positive integer n and integers a, b such that $a^2 \equiv b^2 \pmod{n}$ but $a \not\equiv b \pmod{n}$.

7. Check explicitly that $3^{23} \not\equiv 3^2 \pmod{7}$. (*Hint:* $3^3 = 27 \ldots$)

8. Compute the following remainders—use a calculator to help!
   (a) $12^9 + 19^{24}$ on division by 10.
   (b) 30^{10} on division by 13.
   (c) $17^{251} \cdot 23^{12} - 19^{41}$ on division by 5. (*Hint:* $17 \equiv 2$ *and* $2^2 \equiv -1 \pmod{5}$)
   (d) (Hard) $12^{10} + 2^{36} \cdot 18^{12}$ on division by 141. (*Hint: what nice number is close to 141?*)

9. Prove that $3 \mid (4^n - 1)$ for all positive integers n.

10. Let n be an integer. Prove that exactly one of n, $n + 2$ and $n + 4$ is divisible by 3.

11. (a) Let n be a positive integer. Prove that n is congruent to the sum of its digits modulo 9. (*Hint: e.g.* $345 = 3 \cdot 10^2 + \cdots$)
    (b) Is the integer 123456789 divisible by 9?

12. Describe all integers x which satisfy the congruence equations:
    (a) $3x \equiv 2 \pmod{8}$ (b) $7x \equiv 28 \pmod{42}$.

13. Abraham Lincoln was born February 12th, 1809. On what day of the week was this? (*Start by looking up the day for February 12th* this year)

14. Let n be an integer.
    (a) Prove: $n^2 \equiv 0$ or $1 \pmod{3}$. (*Hint: prove by cases*)
    (b) Prove: $n^2 \equiv 0$ or $1 \pmod{4}$.
    (c) Find all possible remainders of n^2 on division by 7.
    (d) Find all possible remainders of n^3 on division by 7.

15. Use some part(s) of Exercise 14 to prove the following.
    (a) $\sqrt{4m + 6}$ is not an integer, for any integer $m \geq -1$.
    (b) Any number which is simultaneously a square and a cube must be of the form $7k$ or $7k + 1$ for some integer k.

16. Let n be an integer ≥ 2 and consider numbers of the form $\underbrace{11 \cdots 11}_{n \text{ times}}$
    (a) Prove every such number can be written as $4k + 3$ for some $k \in \mathbb{Z}$.
    (b) Prove that no such number is a perfect square.

17. Fermat's Little Theorem states that if p is prime and $a \not\equiv 0 \pmod{p}$, 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) Verify that $341 \mid (2^{341} - 2)$. What does this say about the *converse* to part (b)?

3.2 Greatest Common Divisors and the Euclidean Algorithm

A basic goal for number theorists is to find *integer* solutions to equations. For instance:

> Are there any *integer points* on the line with equation $9x - 21y = 6$? That is, does the equation $9x - 21y = 6$ have any solutions, where both x, y are *integers*?

You might start by sketching both lines (lined graph paper will help). What do you observe? If there are integer points, do they seem to follow any pattern?

In this section we will see how to solve all such *linear Diophantine equations*.[20] The method introduces a famous procedure dating at least to Euclid's *Elements* (c. 300 BCE).

Definition 3.14. Let a and b be integers, not both zero. Their *greatest common divisor* $\gcd(a, b)$ is the largest (positive) divisor of both a and b. We say that a and b are *relatively prime* if $\gcd(a, b) = 1$.

Example 3.15. The positive divisors of $a = 60$ and $b = 90$ are listed in the table. The greatest common divisor $\gcd(60, 90) = 30$ is plainly the largest number common to both rows.

a	1	2	3	4	5	6	10	12	15	20	30	60
b	1	2	3	5	6	9	10	15	18	30	45	90

Listing divisors is very inefficient given large integers. This is where Euclid rides to the rescue.

Theorem 3.16 (Euclidean Algorithm). *Let $a > b$ be positive integers. We construct a decreasing sequence of integers $b = r_0 > r_1 > r_2 > \cdots$*

1. *Apply the division algorithm (Theorem 3.3):* $a = q_1 b + r_1$ *with* $0 \le r_1 < b$
2. *If $r_1 > 0$, apply the division algorithm again:* $b = q_2 r_1 + r_2$ *with* $0 \le r_2 < r_1$
3. *If $r_2 > 0$, apply again:* $r_1 = q_3 r_2 + r_3$ *with* $0 \le r_3 < r_2$
4. *While $r_i > 0$, keep repeating the division algorithm, dividing each r_{i+1} by r_i.*

The algorithm eventually terminates with a remainder of zero: some $r_{t+1} = 0$. The greatest common divisor is the last non-zero remainder: $\gcd(a, b) = r_t$.

Exercise 8 provides a proof. If either of a, b are negative just ignore the signs: for instance $\gcd(-4, 34) = \gcd(34, 4) = 2$.

Example 3.17. We compute $\gcd(1260, 750)$ using the Euclidean algorithm. Note how each line is a single instance of the division algorithm $a = qb + r$ and how the remainders move diagonally $\swarrow$ at each step. For this first example, we also summarize the data in a table.

$$\begin{aligned} 1260 &= 1 \times 750 + 510 \\ 750 &= 1 \times 510 + 240 \\ 510 &= 2 \times 240 + 30 \\ 240 &= 8 \times 30 + 0 \end{aligned}$$

a	q	b	r
1260	1	750	510
750	1	510	240
510	2	240	30
240	8	30	0

Theorem 3.16 says that $\gcd(1260, 750) = 30$, the last non-zero remainder.

[20] Equations with integer coefficients and solutions honor the number theorist Diophantus of Alexandria (3rd C. CE).

Reversing the Algorithm

To apply the Euclidean algorithm to our motivational problem of integer points on lines, we run it backwards. Start with the penultimate line of the algorithm and move upwards, substituting remainders one at a time. The result is an expression $\gcd(a,b) = ax + by$ for some integers x, y.

Example (3.17, cont). We find integers x, y such that $1260x + 750y = 30$ $(= \gcd(1260, 750))$. Start by expressing 30 using the third step of the algorithm and work upwards:

$$\begin{aligned} 30 &= 510 - 2 \times 240 && (3^{\text{rd}}/\text{penultimate line}) \\ &= 510 - 2(750 - 510) = 3 \times 510 - 2 \times 750 && (2^{\text{nd}} \text{ line}) \\ &= 3(1260 - 750) - 2 \times 750 && (1^{\text{st}} \text{ line}) \\ &= 3 \times 1260 - 5 \times 750 \end{aligned}$$

Rearranging, we see that $x = 3$ and $y = -5$ satisfy the equation $1260x + 750y = 30$. To simplify, divide everything by 30 to see that $(3, -5)$ is an integer point on the line $42x + 25y = 1$.

This process of reversing the algorithm works in general, yielding a very powerful result.

Corollary 3.18 (Bézout's Identity). *Given integers a, b, not both zero, $\exists x, y \in \mathbb{Z}$ such that*

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

If either a, b are negative, apply the algorithm to $|a|, |b|$ and correct the signs afterwards.

Corollary 3.19. *Suppose a, b, c are integers for which $\gcd(a,b) = 1$ and $a \mid bc$. Then $a \mid c$.*

Proof. By Bézout's identity, $\exists x, y \in \mathbb{Z}$ for which $ax + by = 1$, whence $acx + bcy = c$. The left hand side is now a multiple of a. ∎

Examples 3.20. 1. The claim "$a \mid c$ and $b \mid c \implies ab \mid c$" is, in general, *false* (e.g. $a = b = c = 2$). However: Suppose $c = 2m = 3n$ for some integers m, n. Since $\gcd(2,3) = 1$ and $2 \mid (3n)$, we see that $2 \mid n$. Thus $n = 2k$ for some $k \in \mathbb{Z}$ and $c = 6k$. We conclude:

$$2 \mid c \text{ and } 3 \mid c \implies 6 \mid c$$

Though we could have done this example without heavy machinery (use Theorem 2.29), Bézout proves that the general claim holds whenever a, b are coprime.

2. (Example 3.17, mk. III) We know that $(x_0, y_0) = (3, -5)$ solves the equation $1260x + 750y = 30$ (equivalently $42x + 25y = 1$). Suppose (x, y) is any other integer solution and observe that

$$42(x - x_0) + 25(y - y_0) = 1 - 1 = 0 \implies 42(x - x_0) = -25(y - y_0)$$

Since $\gcd(42, 25) = 1$, the corollary says $42 \mid (y - y_0)$. Write $y - y_0 = 42t$ for some $t \in \mathbb{Z}$, then

$$42(x - x_0) = -25 \cdot 42t \implies x - x_0 = -25t$$

We've therefore found *all* integer solutions to the original equation:

$$x = 3 - 25t, \; y = -5 + 42t \text{ where } t \in \mathbb{Z}$$

The method in the example works for all solvable linear Diophantine equations.

Theorem 3.21. *Let a, b, c be integers where a, b are not both zero, and let $d = \gcd(a, b)$.*

1. *The equation $ax + by = c$ has an integer solution (x, y) if and only if $d \mid c$.*
2. *If a solution exists, then there are infinitely many of them:*

$$x = x_0 - \frac{b}{d}t, \qquad y = y_0 + \frac{a}{d}t \tag{$*$}$$

where $t \in \mathbb{Z}$ and (x_0, y_0) is some fixed solution (say as supplied by the Euclidean algorithm).

Warning! If $c \neq d = \gcd(a, b)$, you will need to scale the integers obtained in Bézout's identity to find an initial solution (x_0, y_0): see below.

Example 3.22. We compute $\gcd(123, 78) = 3$: remainders are underlined for clarity.

$$\left.\begin{array}{l} \underline{123} = 1 \times \underline{78} + \underline{45} \\ \underline{78} = 1 \times \underline{45} + \underline{33} \\ \underline{45} = 1 \times \underline{33} + \underline{12} \\ \underline{33} = 2 \times \underline{12} + \underline{9} \\ \underline{12} = 1 \times \underline{9} + \underline{3} \\ \underline{9} = 3 \times \underline{3} + \underline{0} \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} \underline{3} = \underline{12} - \underline{9} \\ \quad = \underline{12} - (\underline{33} - 2 \times \underline{12}) = 3 \times \underline{12} - \underline{33} \\ \quad = 3(\underline{45} - \underline{33}) - \underline{33} = 3 \times \underline{45} - 4 \times \underline{33} \\ \quad = 3 \times \underline{45} - 4(\underline{78} - \underline{45}) = 7 \times \underline{45} - 4 \times \underline{78} \\ \quad = 7(\underline{123} - \underline{78}) - 4 \times \underline{78} \\ \quad = 7 \times \underline{123} - 11 \times \underline{78} \end{array}\right.$$

(a) Since $3 \nmid 8$, the equation $78x + 123y = 8$ has no integer solutions.

(b) By Bézout's identity, the equation $123x - 78y = -6$ has a particular solution

$$(x_0, y_0) = (-14, 22) \qquad \text{(Warning: Multiply } (7, -11) \text{ by } -2)$$

All integer solutions to the equation are then given by

$$(x, y) = (-14, 22) + \left(\tfrac{78}{3}, \tfrac{123}{3}\right) t = \big(-14 + 26t, 22 + 41t\big), \quad \text{where } t \in \mathbb{Z}$$

Linear Congruence Equations

We began to consider linear congruence equations in Example 3.13. The "points on lines" method also applies in this context. To see why, observe that

$$ax \equiv c \pmod{n} \iff \exists y \in \mathbb{Z} \text{ such that } ax = ny + c$$

The Theorem tells us when we can solve the right hand side, and how. To solve the congruence equation, we just need to extract the x-part.

Examples 3.23. 1. To solve $123x \equiv -6 \pmod{78}$ is to find a solution to the equation $123x = 78y - 6$. By the above, all solutions have the form $x = -14 + 26t$. Otherwise said

$$123x \equiv -6 \pmod{78} \implies x \equiv -14 \equiv 12 \pmod{26}$$

2. The congruence equation $4x \equiv 6 \pmod{20}$ has no solutions. If it did, then the linear equation $4x = 20y + 6$ would have integer solutions, but it doesn't: $\gcd(4, 20) = 4$ does not divide 6.

3. Here is a slightly different approach for $7x \equiv 11 \pmod{23}$. By applying the algorithm,

$$\gcd(23,7) = 1 = 7 \cdot 10 - 23 \cdot 3 \implies 7 \cdot 10 \equiv 1 \pmod{23}$$

The original equation may now be solved via *multiplication*:

$$7x \equiv 11 \implies x \equiv 10 \cdot 7x \equiv 110 \equiv 18 \pmod{23}$$

In this context, *division by 7* really means *multiplication*[21] *by 10*.

Exercises 3.2. A reading quiz and several questions with linked video solutions can be found online. Some of these questions are tricky, particularly the last few, and are included to give you a taste of upper-division number theory and abstract algebra.

1. Use the Euclidean algorithm to compute the greatest common divisors.
   (a) $\gcd(20,12)$ (b) $\gcd(100,36)$ (c) $\gcd(207,496)$
2. For each part of Exercise 1, find integers x, y which satisfy Bézout's identity $\gcd(a,b) = ax + by$.
3. Describe all the integer points on the line $9x - 21y = 6$ using Theorem 3.21.
4. (a) Use Theorem 3.21 to show that there are no integer points on the line $4x + 6y = 1$.
   (b) Give an elementary proof (without using the Theorem) of part (a)?
5. Find all integer points (x, y) on the following lines, or show that none exist.
   (a) $16x - 33y = 2$ (b) $122x + 36y = 3$
   (c) $303x + 204y = 6$ (d) $324x - 204y = -12$
6. (a) Show that there exists no integer x such that $3x \equiv 5 \pmod{6}$.
   (b) Find all solutions x to the congruence equation $12x \equiv 1 \pmod{17}$.
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 93 candies, can you be sure how much candy each person took?
8. We complete the proof of the Euclidean algorithm (Theorem 3.16).
   (a) Suppose $a > b > r_1 > r_2 \cdots$ is a sequence of non-negative integers. Why must there be only *finitely many* terms? This shows that the algorithm terminates with some $r_{t+1} = 0$.
   (b) Suppose that a, b, q, r are integers satisfying $a = qb + r$. Prove that $\gcd(a,b) = \gcd(b,r)$. (*You cannot use Bézout's identity to do this, since Bézout is a corollary of the algorithm!*)
   (c) Argue that $\gcd(a,b) = r_t$.
9. Suppose $m \neq 0$. What is $\gcd(m,0)$? Why? Why is Bézout's identity trivial in this situation?
10. Use Bézout's identity to prove that if $k \mid a$ and $k \mid b$, then $k \mid \gcd(a,b)$.

[21] In future studies, you'll refer to 10 as the *multiplicative inverse of 7 in the ring* $\mathbb{Z}_{23}$, and write $7^{-1} = 10$ (see Exercise 18).

11. Prove: $\gcd(m,n) = 1 \iff \exists x, y \in \mathbb{Z}$ such that $mx + ny = 1$.
(*Hint: One direction follows from Bézout's identity, but the other…*)

12. Use Example 3.20.1 to prove the following.
    (a) Let $n \geq 3$ be an odd number. Show that $n \equiv 1 \pmod 3 \implies n \equiv 1 \pmod 6$.
    (b) $\frac{1}{6}n(n+1)(2n+1)$ is an integer.

13. Let $a, b, p \in \mathbb{Z}$. Recall Definition 2.40: the only positive divisors of a prime are itself and 1.
    (a) Suppose p is prime and that $p \mid ab$. Prove that $p \mid a$ or $p \mid b$.
    (*Hint: if $p \nmid a$, use Corollary 3.19*)
    (b) Suppose $p \geq 2$ is an integer with the property that $\forall a, b \in \mathbb{Z}$, $p \mid ab \implies p \mid a$ or $p \mid b$. Prove that p is prime.
    (*Hint: prove the contrapositive*)

14. (Hard) Show that if a is relatively prime to both b and c then it is relatively prime to bc.
(*Hint: suppose $d \mid a$ and $d \mid bc$ and try to prove that $d = \pm 1$*)

15. The general rule for congruence division is as follows:

$$ac \equiv bc \pmod n \implies a \equiv b \pmod{\tfrac{n}{\gcd(c,n)}}$$

    (a) Use the rule to find all solutions to the congruence equation $22x \equiv 66 \pmod{77}$.
    (b) We prove the rule. Let $d = \gcd(c, n)$.
        i. Explain why $\gcd(\frac{c}{d}, \frac{n}{d}) = 1$
        ii. If $ac \equiv bc \pmod n$, prove that $\frac{n}{d} \mid (a - b)$.

16. (a) Prove part 1 of Theorem 3.21 using Bézout's identity.
    (b) Prove part 2 by mimicking the method in Example 3.17, (mk. III).

17. (Hard) Apply the discussion of the Euclidean algorithm and linear equations to the following.
    (a) Complete the table.
    If n is a positive integer, make a conjecture for the value of $\gcd(2n, n+1)$ and prove it.

n	1	2	3	4	5	6
$\gcd(2n, n+1)$						

    (b) Show that $\gcd(5n + 2, 12n + 5) = 1$ for every integer n.

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. 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) 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.

4 Sets and Functions

Sets are the fundamental building blocks of mathematics, providing the language necessary for describing mathematical objects and for grouping objects together according to shared characteristics. Understanding how to read and effectively employ set notation is our primary focus. The mathematical discipline of *set theory* is, however, far more ambitious than this. Set theorists define all basic mathematical objects—*numbers, addition, functions,* etc.—purely in terms of sets, an impractical approach for most working mathematicians, most of the time.[22] We will only scratch the surface of this. Indeed long before one can accept that such an approach has its place in mathematics, a significant level of familiarity with sets and their basic operations is necessary.

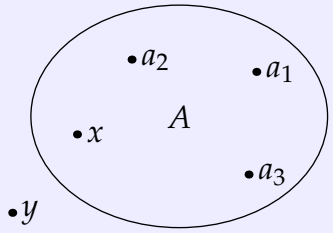
4.1 Set Notation and Subsets

Without any attempt to define the meaning of *object,* we start with a naïve definition.

Definition 4.1. A *set* is a collection of objects, its *elements* or *members.*

The notation $x \in A$ is read "x is an *element/member* of the set A," or more often simply "x is in A." Otherwise said, x is an object in the collection labelled A.

If y is a member of some other set, but not of A, we write $y \notin A$ ("y is not in A").



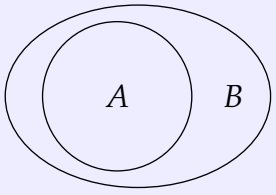
As in the definition, it is typical to use upper-case letters ($A, B, C, \ldots$) for abstract sets and lower-case letters for their elements.

Venn diagrams are useful for visualizing abstract sets. A set is represented by a region in the plane, with elements depicted by dots. The diagram in the definition represents a set A comprising at least four elements a_1, a_2, a_3 and x. The element y does not lie in A.

Example 4.2. Let A be the set of (names of) US states. Then Michigan $\in A$ and Saskatchewan $\notin A$.

Definition 4.3. Let A and B be sets.

1. Sets are *equal,* written $A = B$, if they have precisely the same elements.
2. A is a *subset* of B, written $A \subseteq B$, if every element of A is also an element of B.
3. A is a *proper subset* of B if $A \subseteq B$ and $A \neq B$. To stress this, we'd write $A \subsetneq B$. The Venn diagram on the right represents a proper subset.



The following observations are simply translations of the definition:

1. Equality: $A = B \iff A \subseteq B$ and $B \subseteq A$.
2. Subset: $A \subseteq B \iff \big(x \in A \implies x \in B\big) \iff \big(\forall x \in A, x \in B\big)$
3. Not a subset: $A \nsubseteq B \iff \exists x \in A$ for which $x \notin B$.

[22] Within *axiomatic set theory* it can take over 100 pages to justify writing $1 + 1 = 2$. The difficulty is that rigorous definitions—using sets—are first required of the notions *one, two, equals* and *add*...

Roster & Set-Builder Notation

Roster notation is the most basic way to describe the elements of a set: simply *list* the elements in any order between curly brackets { , }.

Example 4.4. Let A be the set of (real number) solutions to the equation $2x^2 - 7x + 3 = 0$ and let B be the set of *integer* solutions to the same equation. Since the polynomial factorizes as $(2x-1)(x-3)$, we see that $A = \{3, \frac{1}{2}\}$ and $B = \{3\}$. We could also write $A = \{\frac{1}{2}, 3\}$ since *order doesn't matter in roster notation.* Moreover:

- $A \not\subseteq B$ since $\frac{1}{2} \in A$ and $\frac{1}{2} \notin B$.
- $B \subseteq A$ since 3 (the only element of B) lies in A. Indeed $B \subsetneq A$ is a proper subset since $A \neq B$.

Roster notation is ideal for small sets, but is of limited utility when trying to describe large sets. This is where our second notation rides to the rescue.

Set-builder notation describes the elements of a set using some common property. Suppose $\mathcal{U}$ is some (already understood) set and $P(x)$ is a propositional function with domain $\mathcal{U}$, then

$$A := \{x \in \mathcal{U} : P(x)\} \qquad (\text{"}A \text{ is the set of } x \text{ in } \mathcal{U} \text{ such that } P(x)\text{"})$$

defines a set A as *the subset* of $\mathcal{U}$ all of whose elements x satisfy the property $P(x)$. A vertical separator $|$ is often used instead of a colon: we'll use both, though it might be essential in a given context to use one rather than the other for clarity.

Examples 4.5. 1. Continuing Example 4.4, recall that $\mathbb{R}$ represents the set of real numbers and $\mathbb{Z}$ the set of integers. In set-builder notation, our solution sets may be written

$$A = \{x \in \mathbb{R} : 2x^2 - 7x + 3 = 0\}, \qquad B = \{x \in \mathbb{Z} : 2x^2 - 7x + 3 = 0\}$$

In this case the qualifying proposition $P(x)$ is "$2x^2 - 7x + 3 = 0$."

We can also express the fact that B is a subset of A in this notation (this time with a vertical separator),

$$B = \{x \in A \mid x \in \mathbb{Z}\} \qquad (\text{"the set of elements } x \text{ in } A \text{ such that } x \text{ is an integer"})$$

2. Let $X = \{2, 4, 6\}$ and $Y = \{1, 2, 5, 6\}$. There are many options for how to write these in set-builder notation. For instance:

$$X = \{n \in \mathbb{Z} : \tfrac{1}{2}n \in \{1, 2, 3\}\}, \qquad Y = \{n \in \mathbb{Z} \mid 1 \leq n \leq 6 \text{ and } n \neq 3, 4\}$$

We now practice the opposite skill by converting five sets from set-builder to roster notation.

$$S_1 = \{x \in X : x \text{ is divisible by } 4\} = \{4\} \qquad S_2 = \{y \in Y : y \text{ is odd}\} = \{1, 5\}$$
$$S_3 = \{x \in X \mid x \in Y\} = \{2, 6\} \qquad S_4 = \{x \in X : x \notin Y\} = \{4\}$$
$$S_5 = \{y \in Y \mid y \text{ is odd and } y - 1 \in X\} = \{5\}$$

Can you find alternative descriptions in set-builder notation for the sets $S_1, \ldots, S_5$ above? Take your time getting used to this notation: the ability to translate between various descriptions of a set is *crucial* to reading mathematics!

3. We use the set $C = \{0, 1, 2, 3, \ldots, 24\}$ to describe $D = \{n \in \mathbb{Z} : n^2 - 3 \in C\}$ in roster notation. Start by expanding the criterion for membership in D:

$$n^2 - 3 \in C \iff n^2 \in \{3, 4, 5, \ldots, 25, 26, 27\}$$

Since n must be an integer, it follows that $D = \{\pm 2, \pm 3, \pm 4, \pm 5\}$.

4. To express $E = \{0, 2, 6, 12, \ldots\}$ in set-builder notation we might spot a pattern and decide that

$$E = \{n \in \mathbb{Z} : n = m(m+1) \text{ for some integer } m \geq 0\}$$

The problem is that we cannot guarantee our correctness! Perhaps the correct formula is

$$n = m(m+1) + m(m-2)(m-6)(m-12)$$

In the first case the next term in the sequence is $4 \cdot 5 = 20$, whereas in the second case it is $20 + 128 = 148$. For larger sets, the clarity afforded by set-builder notation is essential!

Common Sets of Numbers

We've used some of this notation already, and much of the rest should be familiar.

Natural numbers	$\mathbb{N} = \{1, 2, 3, 4, \ldots\}$ is the set of *positive* integers.
Integers	$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.
Rational numbers	$\mathbb{Q} = \{\frac{m}{n} : m \in \mathbb{Z} \text{ and } n \in \mathbb{N}\} = \{\frac{a}{b} \mid a, b \in \mathbb{Z} \text{ and } b \neq 0\}$.
Real numbers	$\mathbb{R}$. Even a rudimentary definition is too involved for this text.[23]
Complex numbers	$\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}$ where $i^2 = -1$. We won't use these.

Examples 4.6. 1. For instance: $7 \in \mathbb{N}$, $\pi \in \mathbb{R}$, $-\frac{7}{9} \notin \mathbb{Z}$, $\sqrt{2} \notin \mathbb{Q}$ and $3 + \sqrt{5}i \in \mathbb{C}$.

2. The basic symbols can be decorated to make natural modifications. For example:
   - $\mathbb{N}_0 = \{0, 1, 2, 3, 4, \ldots\} = \mathbb{Z}_0^+ = \{x \in \mathbb{Z} : x \geq 0\}$. Also called the *whole numbers* ($\mathbb{W}$).
   - $\mathbb{Z}_{\geq 5} = \{5, 6, 7, 8, \ldots\} = \{x \in \mathbb{Z} : x \geq 5\}$ denotes the integers greater than or equal to 5.
   - $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$ is the set of positive real numbers.
   - $4\mathbb{Z} = \{\ldots, -8, -4, 0, 4, 8, 12, \ldots\} = \{x \in \mathbb{Z} : 4 \mid x\}$ is the set[24] of integer multiples of 4. This notation can be used for non-integer multiples, e.g. $\pi\mathbb{Z} = \{\ldots, -\pi, 0, \pi, 2\pi, \ldots\}$.
   - $2\mathbb{Z} + 1 = \{x \in \mathbb{Z} : x \equiv 1 \pmod 2\}$ is the set of odd integers.

3. *Intervals* are the most commonly encountered subsets of the real numbers. For instance:
   - $[1, \pi] = \{x \in \mathbb{R} \mid 1 \leq x \leq \pi\}$ is a *closed* interval
   - $[-4, 7.21) = \{x \in \mathbb{R} \mid -4 \leq x < 7.21\}$ is a *half-open* interval.
   - $(-\infty, \sqrt{2}) = \{x \in \mathbb{R} \mid x < \sqrt{2}\}$ is an *infinite (open)* interval.

[23] We simply assume the reader is comfortable with the idea of the real line where number corresponds to *length*. A rigorous development of $\mathbb{R}$ is a matter for an upper-division *analysis* course.

[24] Be careful!—the colon is the "such that" separator while $\mid$ denotes the property "4 divides x."

In view of the natural subset relationships $\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R} \subsetneq \mathbb{C}$, we consider a simple result.

Lemma 4.7 (Transitivity of Subset). *Suppose $A \subseteq B$ and $B \subseteq C$. Then $A \subseteq C$.*

Proof. Think back to the criteria following Definition 4.3. Suppose $A \subseteq B$ and $B \subseteq C$. Then

$$x \in A \overset{(A \subseteq B)}{\Longrightarrow} x \in B \overset{(B \subseteq C)}{\Longrightarrow} x \in C$$

We conclude that $A \subseteq C$. ∎

Compare this to Exercise 2.1.9b: if we translate each subset relation into an implication, the proof structure is $(x \in A \Rightarrow x \in B) \wedge (x \in B \Rightarrow x \in C) \Longrightarrow (x \in A \Rightarrow x \in C)$. This is typical of basic results about sets: after translation, the theorem reduces to one of the standard rules of logic.

Cardinality and the Empty Set

It is helpful to introduce some terminology to describe the *size* of a set.

Definition 4.8. A *finite set* contains only a finite number of elements: this number is its *cardinality*, written $|A|$. A set with infinitely many elements is said to be an *infinite set*.
The symbol $\emptyset$ denotes the *empty set*: a set containing no elements (cardinality zero: $|\emptyset| = 0$).

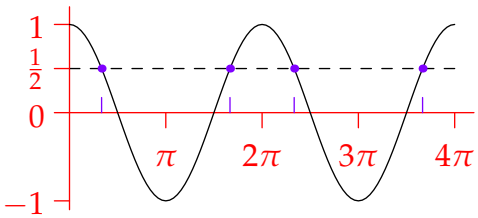
Examples 4.9. 1. Let $A = \{1, 3, \pi, \sqrt{2}, 103\}$, then $|A| = 5$.

2. Let $B = \{4, \{1,2\}, \{3\}\}$. The elements of B are 4, $\{1,2\}$ and $\{3\}$, therefore $|B| = 3$. It doesn't matter that the *element* $\{1,2\} \in B$ is also a set!

3. Recall some basic trigonometry:

$$\left\{x \in [0, 4\pi] : \cos x = \frac{1}{2}\right\} = \left\{\frac{\pi}{3}, \frac{5\pi}{3}, \frac{7\pi}{3}, \frac{11\pi}{3}\right\}$$

has cardinality 4.



4. There are many representations of the empty set: for example

$$\emptyset = \{x \in \mathbb{R} : x^2 = -1\} = \{x \in \mathbb{N} : x^2 + 3x + 2 = 0\} = \{n \in \mathbb{N} : n < 0\}$$

In general, if X is any set and $P(x)$ is false for all $x \in X$, then[25] $\emptyset = \{x \in X : P(x)\}$.

Cardinality is a very simple concept for finite sets; if B is finite, so is any subset, and we have

$$A \subseteq B \implies |A| \leq |B|$$

For infinite sets, cardinality is more subtle. We'll return to this issue and uncover some of the bizarre and fun consequences of infinite cardinalities in Chapter 8.

[25] In some formalizations of set theory, the existence of the empty set is an axiom: an assumption made without proof. Provided one accepts that set-builder notation always defines a set—this is itself an axiom!—and that at least one set X exists, the empty set may be *defined* as in the example: a suitable property $P(x)$ might be something like "$x \notin \{x\}$."

We finish with a couple of simple results regarding the empty set.

Lemma 4.10. *Let A be a set.*

1. *If $|A| = 0$, then $A = \emptyset$. The empty set is* the unique set *with zero cardinality.*
2. *$\emptyset \subseteq A$ and $A \subseteq A$*

Proof. Think about the claim $\emptyset \subseteq A$: by the observations following Definition 4.3, this means

$$x \in \emptyset \implies x \in A$$

This is true (for any set A!) since there are no elements x satisfying the hypothesis.[26]

1. Suppose A has cardinality zero. Repeating and combining with the above observation, we see that $\emptyset \subseteq A$ and $A \subseteq \emptyset$. We conclude that $A = \emptyset$.
2. We already know that $\emptyset \subseteq A$. For the second part, simply observe that $x \in A \implies x \in A$.

■

Exercises 4.1. A reading quiz and several questions with linked video solutions can be found online.

1. Describe the following sets in roster notation: that is, list their elements.
   (a) $\{x \in \mathbb{N} : x^2 \leq 3x\}$ (b) $\{n \in \{0,1,2,3,\ldots,19\} : n+3 \equiv 5 \pmod 4\}$
   (c) $\{n \in \{-2,-1,0,1,\ldots,23\} : 4 \mid n^2\}$ (d) $\{x \in \frac{1}{2}\mathbb{Z} : 0 \leq x \leq 4 \text{ and } 4x^2 \in 2\mathbb{Z}+1\}$
   (e) $\{y \in \mathbb{R} : y = x^2 \text{ for some } x \in \mathbb{R} \text{ with } x^2 - 3x + 2 = 0\}$
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. Each of the following sets of real numbers is a single interval. Determine the interval.
   (a) $\{x \in \mathbb{R} : x > 3 \text{ and } x \leq 17\}$ (b) $\{x \in \mathbb{R} : x \not\leq 3 \text{ or } x \leq 17\}$
   (c) $\{x^2 \in \mathbb{R} : x \neq 0\}$ (d) $\{x \in \mathbb{R}^- : x^2 \geq 16 \text{ and } x^3 \leq 27\}$
4. Is the set $\{x \in \mathbb{Z} : -1 \leq x < 43\}$ finite or infinite? If finite, what is its cardinality?
5. What is the cardinality of the set $\Big\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\Big\}$? What are its elements?
6. Let $A = \emptyset$, $B = \{A\}$, $C = \{\{A\}\}$ and $D = \{A, \{0\}, \{0,1\}\}$.
   Answer the following true or false:
   (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) $\{1\} \in D$
7. List all the *proper* subsets of $\{1,2,3\}$.

[26] If $P(x)$ is always false, then $(\forall x)\ P(x) \implies Q(x)$ is true. This is called a *vacuous* (empty) theorem.

8. Let A, B, C, D be the following sets:

$$A = \{-4, 1, 2, 4, 10\} \qquad B = \{m \in \mathbb{Z} : |m| \leq 12\} \quad (\textit{absolute value of } m)$$
$$C = \{n \in \mathbb{Z} : n^2 \equiv 1 \pmod 3\} \qquad D = \{t \in \mathbb{Z} : t^2 + 3 \in [4, 20)\}$$

Of the 12 subset relations $A \subseteq B$, $A \subseteq C, \ldots, D \subseteq C$, which are true and which false?

9. Let $A = \{1, 2, \{1, 2\}, \{3\}\}$ and $B = \{1, 2\}$. Answer the following true or false:

(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) $\emptyset \in A$

10. Let $A = \{0, 2, 4, 6, 8, 10\}$. Write the set $B = \{X \subseteq A : |X| = 2\}$ in roster notation.

11. (a) Suppose $A \subseteq B \subseteq C \subseteq A$. Show that $A = B = C$.
(b) Is it possible for sets A, B, C to satisfy $A \subsetneq B \subseteq C \subseteq A$? Why/why not?

12. Let $A = \{1,2,3,4\}$, and let $B = \{\{x, y\} : x, y \in A\}$.

(a) Describe B in roster notation (*what happens when $x = y$?*).
(b) Find the cardinalities of the following sets:

$$C = \Big\{\{x, \{y\}\} : x, y \in A\Big\} \quad \text{and} \quad D = \Big\{\Big\{\{x, \{y\}\} : x, y \in A\Big\}\Big\}$$

13. 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.

14. For which real numbers $x > 0$ do we have $[0, x] \subsetneq [0, x^2]$? Prove your assertion.

15. Let $m, n \in \mathbb{N}$. Prove: $m\mathbb{Z} \subseteq n\mathbb{Z} \iff n \mid m$.

16. Given $A \subseteq \mathbb{Z}$ and $x \in \mathbb{Z}$, we say that x is A-mirrored if and only if $-x \in A$. Also define

$$M_A := \{x \in \mathbb{Z} : x \text{ is } A\text{-mirrored}\}$$

(a) What does it mean for x *not* to be A-mirrored?
(b) Find M_B given $B = \{0, 1, -6, -7, 7, 100\}$.
(c) Assume that $A \subseteq \mathbb{Z}$ is closed under addition: 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 $A = M_A$?

17. Define the set $[1] = \{x \in \mathbb{Z} : x \equiv 1 \pmod 5\}$.

(a) Describe the set $[1]$ in roster notation.
(b) Compute the set $M_{[1]}$, as defined in Exercise 16. Is $M_{[1]}$ equal to $[1]$?
(c) Now consider the set $[10] = \{x \in \mathbb{Z} : x \equiv 10 \pmod 5\}$. Are the sets $[10]$ and $M_{[10]}$ equal? Prove or disprove.

18. Consider the set $A = \{a, b, c, d\}$.

(a) Of each cardinality 0, 1, 2, 3 and 4, how many subsets has A? Is there a pattern?
(b) Completely expand the polynomial $(1 + x)^4$. What do you notice about the coefficients?

4.2 Unions, Intersections and Complements

In this section we construct new sets from old, modeled on the logical concepts of *and*, *or*, and *not*.

Definition 4.11. Suppose A and B are sets.

1. The *union* of A and B is the set of elements in A or B:

$$A \cup B := \{x : x \in A \text{ or } x \in B\} \qquad (x \in A \cup B \iff x \in A \text{ or } x \in B)$$

2. The *intersection* of A and B is the set of elements in both A and B:

$$A \cap B := \{x : x \in A \text{ and } x \in B\} \qquad (x \in A \cap B \iff x \in A \text{ and } x \in B)$$

   We say that A and B are *disjoint* if $A \cap B = \emptyset$.

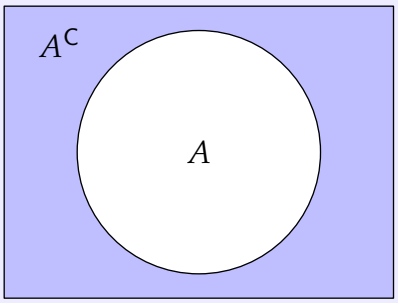
3. The *complement* of A is the set of all elements not in A (with respect to some *universal set*[27] $\mathcal{U}$):

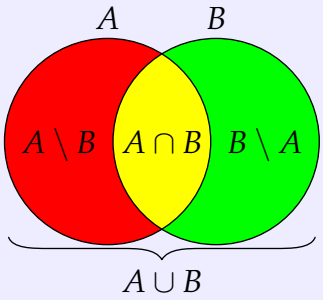
$$A^{\mathsf{C}} := \{x \in \mathcal{U} : x \notin A\} \qquad (x \in A^{\mathsf{C}} \iff x \notin A \text{ (and } x \in \mathcal{U}))$$

4. The *complement of A relative B* is the set of elements in B which are not in A:

$$B \setminus A = \{x \in B : x \notin A\} = B \cap A^{\mathsf{C}} \qquad (x \in B \setminus A \iff x \in B \text{ and } x \notin A)$$

   This can be read "B minus A" (some authors write $B - A$). Similarly $A^{\mathsf{C}} = \mathcal{U} \setminus A$, etc.





In the first Venn diagram, the outer box depicts the universal set $\mathcal{U}$. Though it doesn't constitute a proof, the second diagram certainly suggests that

$$A = (A \setminus B) \cup (A \cap B) \quad \text{and} \quad B = (B \setminus A) \cup (A \cap B)$$

Observe the notational similarity with logic: $\cup$ looks a bit like $\vee$ (OR); $\cap$ like $\wedge$ (AND).

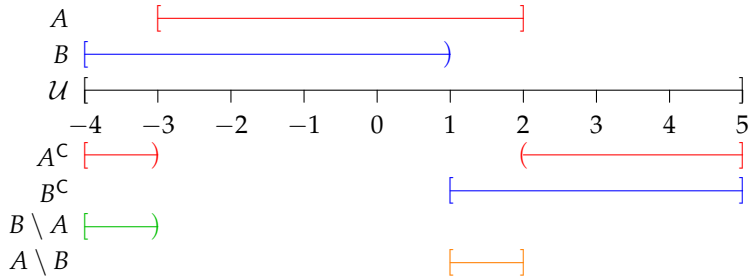
Examples 4.12. 1. Let $\mathcal{U} = \{1,2,3,4,5\}$, $A = \{1,2,3\}$, and $B = \{2,3,4\}$. Then

$$A^{\mathsf{C}} = \{4,5\} \qquad B^{\mathsf{C}} = \{1,5\} \qquad B \setminus A = \{4\} \qquad A \setminus B = \{1\}$$
$$A \cup B = \{1,2,3,4\} \qquad A \cap B = \{2,3\} \qquad A \cap B^{\mathsf{C}} = \{1\} \qquad A^{\mathsf{C}} \cup B^{\mathsf{C}} = \{1,4,5\}$$

[27] This is needed for complements since x has to live somewhere: should, say, $\{1\}^{\mathsf{C}}$ be the set of all *integers* except 1, *real numbers* except 1, etc.?! In many contexts the universal set is naturally assumed: $\mathcal{U} = \mathbb{R}$ makes sense if you are doing calculus, $\mathcal{U} = \mathbb{Z}$ if doing modular arithmetic. The universal set is not needed for the rest of the definition (that the union is a set is typically an axiom, while intersections and relative complements are automatically subsets of pre-existing sets).

2. Using interval notation, let $\mathcal{U} = [-4,5]$, $A = [-3,2]$, and $B = [-4,1)$. Then

$$A^{\mathsf{C}} = [-4,-3) \cup (2,5]$$
$$B^{\mathsf{C}} = [1,5]$$
$$A \setminus B = [1,2]$$
$$B \setminus A = [-4,-3)$$
$$A \cup B = [-4,2]$$
$$A \cap B = [-3,1)$$



While you should be comfortable with these just from the picture, for practice it is worth trying algebraic arguments. In particular, note how A^{C} relies on de Morgan's law (Theorem 2.15):

$$\begin{aligned} x \in A^{\mathsf{C}} &\iff x \notin A \iff \neg(x \in A) \iff \neg(-3 \le x \text{ and } x \le 2) \\ &\iff x < -3 \text{ or } x > 2 && \text{(de Morgan)} \\ &\iff x \in [-4,-3) \cup (2,5] && \text{(remember that } x \in \mathcal{U} \text{ always!)} \end{aligned}$$

3. Let $A = (-\infty, 3)$ and $B = [-2, \infty)$ in interval notation. Then $A \cup B = \mathbb{R}$ and $A \cap B = [-2,3)$.

For the remainder of this section, we summarize the basic rules of set algebra.

Theorem 4.13 (Union/intersection rules). *Let A, B, C be sets. Then:*

1. $\emptyset \cup A = A$ *and* $\emptyset \cap A = \emptyset$
2. $A \cap B \subseteq A \subseteq A \cup B$
3. $A \cup B = B \cup A$ *and* $A \cap B = B \cap A$
4. $A \cup (B \cup C) = (A \cup B) \cup C$ *and* $A \cap (B \cap C) = (A \cap B) \cap C$
5. $A \cup A = A \cap A = A$
6. $A \subseteq B \implies A \cup C \subseteq B \cup C$ *and* $A \cap C \subseteq B \cap C$

If you don't believe a result, try *visualizing* it using a Venn diagram. The basic proof strategy is the same for all parts: convert each statement into propositions (Definition 4.11 parentheses) and use what you know from basic logic (e.g., Theorem 2.15 and page 11). We prove part 2 and half of part 6, leaving some of the rest to the Exercises.

Proof. 2. There are two results here: $A \cap B \subseteq A$ and $A \subseteq A \cup B$. We prove separately, with some commentary on the side.

(a) Suppose $x \in A \cap B$. (Goal: want to prove $x \in A \cap B \Rightarrow x \in A$)
Then $x \in A$ and $x \in B$. (Definition of intersection)
Plainly $x \in A$. We conclude that $A \cap B \subseteq A$ (Definition of subset)

(b) Suppose $y \in A$. (Goal: prove $y \in A \Rightarrow y \in A \cup B$)
Then "$y \in A$ or $y \in B$" is true, whence $y \in A \cup B$. (Definition of union/or)
We conclude that $A \subseteq A \cup B$.

6. (first half) Suppose $A \subseteq B$. We wish to prove that $x \in A \cup B \implies x \in A \cup C$. However,

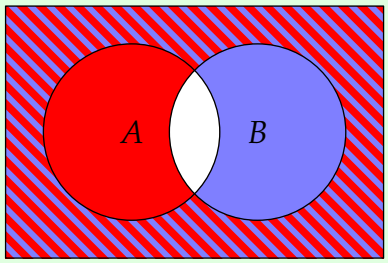
$$\begin{aligned} x \in A \cup C &\implies x \in A \text{ or } x \in C && \text{(definition of union)} \\ &\implies x \in B \text{ or } x \in C && \text{(since } A \subseteq B\text{)} \\ &\implies x \in B \cup C \end{aligned}$$

■

The next batch of rules describe how complements interact with other set operations: parts 1 and 2 are *de Morgan's laws for sets*; unsurprisingly, their proofs depend on the corresponding laws of logic.

Theorem 4.14 (Complement rules). *Let A, B be sets. Then:*

1. $(A \cap B)^C = A^C \cup B^C$ *(see picture)*
2. $(A \cup B)^C = A^C \cap B^C$
3. $(A^C)^C = A$
4. $A \subseteq B \iff B^C \subseteq A^C$



Proof. We prove only part 1. As before, the natural approach is to restate the result using propositions.

$$\begin{aligned} x \in (A \cap B)^C &\iff \neg(x \in A \cap B) \iff \neg(x \in A \text{ and } x \in B) \\ &\iff \neg(x \in A) \text{ or } \neg(x \in B) && \text{(de Morgan's first law of logic)} \\ &\iff x \in A^C \text{ or } x \in B^C \\ &\iff x \in A^C \cup B^C \end{aligned}$$

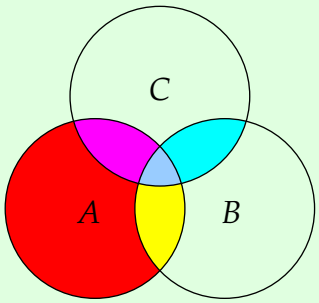
■

Our final pair of results describe the interaction of unions and intersections.

Theorem 4.15 (Distributive laws). *For any sets A, B, C:*

1. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
2. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

The Venn diagram illustrates the second result: think about adding the colored regions.



Proof. We prove only the first result.

$$\begin{aligned} x \in A \cap (B \cup C) &\iff x \in A \text{ and } x \in B \cup C \\ &\iff x \in A \text{ and } (x \in B \text{ or } x \in C) \\ &\iff (x \in A \text{ and } x \in B) \text{ or } (x \in A \text{ and } x \in C) && \text{(distributive law, page 11)} \\ &\iff x \in A \cap B \text{ or } x \in A \cap C \\ &\iff x \in (A \cap B) \cup (A \cap C) \end{aligned}$$

■

Exercises 4.2. A reading quiz and several questions with linked video solutions can be found online.

1. Describe each set simply as you can: e.g.,

$$\{x \in \mathbb{R} : x^2 < 9 \text{ and } x^3 < 8\} = (-3,3) \cap (-\infty, 2) = (-3,2)$$

   (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) $\{y \in \mathbb{R} : \exists x \in \mathbb{R} \text{ with } y = x^2 \text{ and } x \neq 1\}$
   (d) $\{z \in \mathbb{Z} : z^2 \text{ is even and } z^3 \text{ is odd}\}$
   (e) $\{y \in 3\mathbb{Z} + 2 : y^2 \equiv 1 \pmod 3\}$

2. Let $A = \{1,3,5,7,9\}$, $B = \{1,4,7,10\}$ and $\mathcal{U} = \{1,2,\ldots,10\}$. What are the following sets?

   (a) $A \cap B$ (b) $A \cup B$ (c) $B \setminus A$ (d) A^{C}
   (e) $(A \setminus B)^{\mathsf{C}}$ (f) $A^{\mathsf{C}} \cap B^{\mathsf{C}}$ (g) $(A \cup B) \setminus (A \cap B)$

3. Give formal proofs of the following parts of Theorems 4.13, 4.14 and 4.15.

   (a) $\emptyset \cap A = \emptyset$ (b) $A \cap (B \cap C) = (A \cap B) \cap C$
   (c) $(A^{\mathsf{C}})^{\mathsf{C}} = A$ (d) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
   (e) $A \subseteq B \iff B^{\mathsf{C}} \subseteq A^{\mathsf{C}}$

4. By showing that each side is a subset of the other, give a formal proof of the set identity

$$A = (A \setminus B) \cup (A \cap B)$$

   Now repeat your argument using only results from set algebra (Theorems 4.14 and 4.15).

5. Prove the identity $A \cup B = A \iff B \subseteq A$ for any sets A, B.

6. Prove the identities for any sets A, B, C:

   (a) $(A \cap B \cap C)^{\mathsf{C}} = A^{\mathsf{C}} \cup B^{\mathsf{C}} \cup C^{\mathsf{C}}$
   (b) $(A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$

7. Prove or disprove the following conjectures (*Hint: revisit Section 2.4*).

   (a) $\exists x \in \mathbb{R} \setminus \mathbb{Q}$ such that $x^2 \in \mathbb{Q}$. (b) $\forall x \in \mathbb{R} \setminus \mathbb{Q}$ we have $x^2 \in \mathbb{Q}$.

8. Let $A \subseteq \mathbb{R}$, and let $x \in \mathbb{R}$. We say that x is *far away* from the set A if and only if:

$$\exists d > 0 \text{ such that } A \cap [x - d, x] = \emptyset$$

   If this does not happen, we say that x is *close to* A.

   (a) Draw a picture of a set A and elements x, y such that x is *far away* from and y is *close to* A.
   (b) State the meaning of "x is close to A" (negate "x is far away from A").
   (c) Let $A = \{1,2,3\}$.
      i. Show that $x = 4$ is *far away* from A using the definition.
      ii. Let $A = \{1,2,3\}$. Show that $x = 1$ is *close* to A.
   (d) For general $A \subseteq \mathbb{R}$, show that if $x \in A$, then x is *close* to A.
   (e) Let $A = (a,b)$ be a bounded interval. Is the end-point *a far away* from A? What about b?

4.3 Introduction to Functions

Sets become a lot more useful and interesting once you start transforming their elements! This is accomplished using *functions*. In this section we introduce some basic concepts and notation, much of which should be familiar. A formal definition will be given in Chapter 7, but for the present the following will suffice.

Definition 4.16. Let A, B be sets. A *function* $f : A \to B$ is a rule assigning to each input $a \in A$ a single output $b \in B$, typically denoted $f(a)$. Various sets are associated to f:

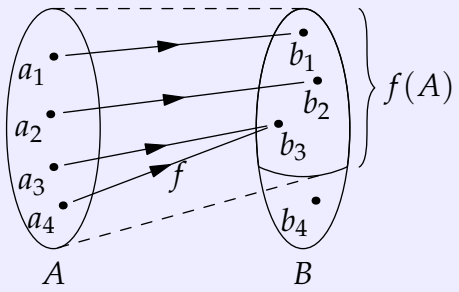
Domain: $\mathrm{dom}(f) = A$ is the set of inputs to the function.

Codomain: $\mathrm{codom}(f) = B$ is the set of potential outputs.

Image of a subset $U \subseteq A$: the set of outputs given inputs in U

$$f(U) := \{f(u) \in B : u \in U\}$$

Range: $\mathrm{range}(f) = f(A) = \{f(a) \in B : a \in A\}$ is the set of realized outputs.



Inverse image (or *pre-image*) of a subset $V \subseteq B$: the set of inputs which are mapped into V

$$f^{-1}(V) := \{a \in A : f(a) \in V\}$$

The rule defining a function can be described using arrow notation $f : a \mapsto b$.

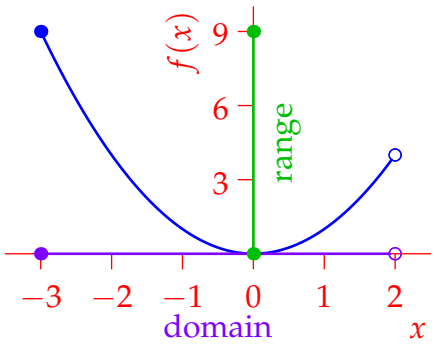
Examples 4.17. 1. Functions whose codomain is (a subset of) the real numbers $\mathbb{R}$ are often graphed: the domain and range are found by projecting the graph onto the two axes.

For instance if $f : [-3, 2) \to \mathbb{R}$ is the square function

$$f : x \mapsto x^2 \qquad \text{(equivalently } f(x) = x^2\text{)}$$

then $\mathrm{dom}(f) = [-3, 2)$ and $\mathrm{range}(f) = [0, 9]$. We could also calculate other images/pre-images, for example,

$$f([-1, 2)) = \{x^2 : -1 \le x < 2\} = [0, 4)$$
$$f^{-1}((-10, 2]) = \{x \in [-3, 2) : -10 < x^2 \le 2\}$$
$$= [-\sqrt{2}, \sqrt{2}]$$



2. Define $f : \mathbb{Z} \to \{0, 1, 2\}$ by $f : n \mapsto n^2 \pmod 3$. The table shows a few examples (remember $\mathrm{dom}(f) = \mathbb{Z}$ is infinite!).

n	0	1	2	3	4	5	6	7
$f(n)$	0	1	1	0	1	1	0	1

Exercise 3.1.14 confirms what the examples suggest, that $\mathrm{range}(f) = \{0, 1\}$. We also compute a single inverse image (revisit the previous two sections if you're unsure about the notation):

$$f^{-1}(\{1\}) = \{x \in \mathbb{Z} : x^2 \equiv 1 \pmod 3\} = \{x \in \mathbb{Z} : x \equiv 1 \text{ or } 2 \pmod 3\}$$
$$= (3\mathbb{Z} + 1) \cup (3\mathbb{Z} + 2)$$

3. Let $A = \{0,1,2,\ldots,7\}$ be the set of remainders modulo 8 and define two functions $f, g : A \to A$:

n	0	1	2	3	4	5	6	7
$f(n)$	0	3	6	1	4	7	2	5
$g(n)$	0	6	4	2	0	6	4	2

$$f(n) = 3n \pmod 8 \qquad g(n) = 6n \pmod 8$$

Since the domain has only 8 elements, the table describes everything. Observe that

$$\begin{aligned} &\text{range}(f) = A && f(\{1,5\}) = \{3,7\} && f^{-1}(\{1,2,3,4\}) = \{1,3,4,6\} \\ &\text{range}(g) = \{0,2,4,6\} && g(\{1,5\}) = \{6\} && g^{-1}(\{1,2,3,4\}) = \{2,3,6,7\} \end{aligned}$$

4. Let $A = \{0,1,2,3,4\}$ and let $B = \{\text{two-element subsets of } A\}$. Define

$$f : A \to B : a \mapsto \{a, a+1 \pmod 5\}$$

where we take the remainder modulo 5. You should be able to convince yourself that

$$\text{range}(f) = \{\{0,1\},\{1,2\},\{2,3\},\{3,4\},\{4,0\}\}$$
$$f(\{1,4\}) = \{f(1), f(4)\} = \{\{1,2\},\{4,0\}\} \quad \text{and} \quad f^{-1}\{\{2,4\},\{4,0\}\} = \{4\}$$

Injections, Surjections and Invertibility

We turn our attention to perhaps the most important properties a function can possess.

Definition 4.18. Let $f : A \to B$ be a function. We say that f is:

1. *Injective* (*1–1*, an *injection*) if it never outputs the same value twice. Equivalently,[28]

$$f(a_1) = f(a_2) \implies a_1 = a_2 \qquad (\text{“}\forall a_1, a_2 \in A\text{” is typically hidden})$$

2. *Surjective* (*onto*, a *surjection*) if every possible output is realized: $B = \text{range}(f)$. Equivalently,[28]

$$\forall b \in B,\ \exists a \in A \text{ such that } f(a) = b$$

3. *Bijective* (*invertible*, a *bijection*) if it is both injective and surjective. Equivalently

$$\forall b \in B,\ \exists \text{ a } \textit{unique } a \in A \text{ such that } f(a) = b$$

When f is bijective, its *inverse function* is $f^{-1} : B \to A : b \mapsto a$.

Since the definitions of injectivity and surjectivity are both universal statements, it suffices to provide *counter-examples* to show that a function is *not injective* or *not surjective.* Indeed:

f *not injective*:	$\exists a_1 \neq a_2 \in A$ such that $f(a_1) = f(a_2)$
f *not surjective*:	$\exists b \in B$ such that $\forall a \in A,\ f(a) \neq b$

[28] For injectivity this is the contrapositive: if f never takes the same value twice, then $a_1 \neq a_2 \implies f(a_1) \neq f(a_2)$. For surjectivity, the quantified statement expresses $B \subseteq \text{range}(f)$. The inclusion $\text{range}(f) \subseteq B$ is true for *any* function.

Examples (4.17, cont.). We briefly revisit our previous examples.

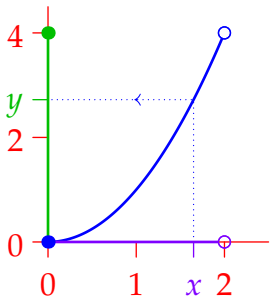
1. Let $f : [-3,2) \to \mathbb{R} : x \mapsto x^2$.
   - f is non-injective: $f(1) = f(-1)$ provides a counter-example. $(a_1 = 1 = -a_2)$
   - f is non-surjective: there is no $x \in [-3,2)$ for which $x^2 = -5$. $(b = -5)$

   We can obtain a related injective function by shrinking the domain, for instance $g : [0,2) \to \mathbb{R} : x \mapsto x^2$. Indeed

   $$g(x_1) = g(x_2) \implies x_1^2 = x_2^2 \implies x_1 = x_2$$

   since $x_1, x_2 \in [0,2)$ are non-negative. By also shrinking the codomain, we get a surjective (now *bijective*) function: $h : [0,2) \to [0,4) : x \mapsto x^2$.

   Proof of surjectivity: Given $y \in [0,4)$, let $x = \sqrt{y}$, then $y = h(x)$.



2. $f : \mathbb{Z} \to \{0,1,2\} : n \mapsto n^2 \pmod 3$ is neither injective nor surjective.
   - f is non-injective: for instance $f(1) = f(2)$.
   - f is non-surjective: $\text{range}(f) = \{0,1\} \neq \{0,1,2\} = \text{codom}(f)$.

3. Given $f, g : A \to A$ where $A = \{0,1,\ldots,7\}$ as in the table:

n	0	1	2	3	4	5	6	7
$f(n)$	0	3	6	1	4	7	2	5
$g(n)$	0	6	4	2	0	6	4	2

   - f is bijective: all elements of $\text{codom}(f)$ appear exactly once in the f-row.
   - g is non-injective: e.g., $g(0) = g(4)$.
   - g is non-surjective: e.g., $1 \notin \text{range}(g) = \{0,2,4,6\}$.

4. Let $A = \{0,1,2,3,4\}$ and $f : A \to \{\text{two-element subsets of } A\} : a \mapsto \{a, a+1 \pmod 5\}$
   - f is injective: suppose $a_1, a_2 \in A$, then

     $$f(a_1) = f(a_2) \implies \{a_1, a_1+1 \pmod 5\} = \{a_2, a_2+1 \pmod 5\} \implies a_1 = a_2$$

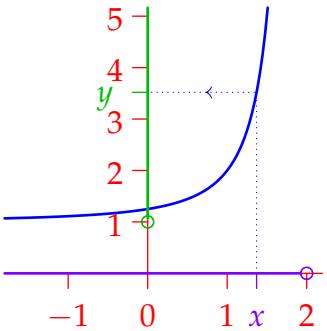
   - f is not surjective: e.g., $\{1,3\} \notin \text{range}(f)$.

You should have seen the approach of the next example in other classes.

Example 4.19. We show that $f : (-\infty, 2) \to (1, \infty) : x \mapsto 1 + \frac{1}{(x-2)^2}$ is bijective by *computing its inverse function*. Just solve for x in terms of y:

$$y = 1 + \frac{1}{(x-2)^2} \implies (x-2)^2 = \frac{1}{y-1}$$
$$\implies f^{-1}(y) = x = 2 - \frac{1}{\sqrt{y-1}}$$

The sign of the square-root was chosen so that $x \in \text{dom}(f) = (-\infty, 2)$.



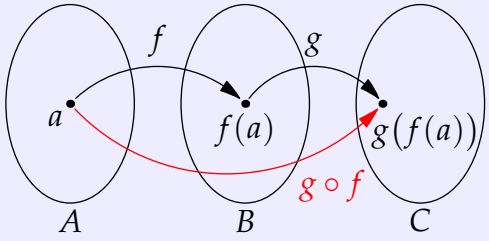
Composition of Functions

We consider how injectivity and surjectivity interact with composition of functions.

Definition 4.20. Given functions $f : A \to B$ and $g : B \to C$, their *composition* is the function

$$g \circ f : A \to C : a \mapsto g\big(f(a)\big)$$

Note the order: to compute $(g \circ f)(a)$, first apply f, then g.



In practice, some restriction of domains might be required in order to define a composition.

Example 4.21. If $f(x) = x^2$ and $g(x) = \frac{1}{x-1}$, then

$$(g \circ f)(x) = \frac{1}{x^2 - 1} \quad \text{and} \quad (f \circ g)(x) = \frac{1}{(x-1)^2}$$

Even though ± 1 are legitimate inputs for f, $\operatorname{dom}(g \circ f) = \mathbb{R} \setminus \{\pm 1\}$ is implied so as to prevent division by zero.

Theorem 4.22. *Let $f : A \to B$ and $g : B \to C$ be functions. Then:*

1. *If f and g are injective, then $g \circ f$ is injective.*
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. Suppose f and g are injective and let $a_1, a_2 \in A$. Then

$$\begin{aligned} (g \circ f)(a_1) = (g \circ f)(a_2) &\implies g\big(f(a_1)\big) = g\big(f(a_2)\big) \\ &\implies f(a_1) = f(a_2) && \text{(since } g \text{ is injective)} \\ &\implies a_1 = a_2 && \text{(since } f \text{ is injective)} \end{aligned}$$

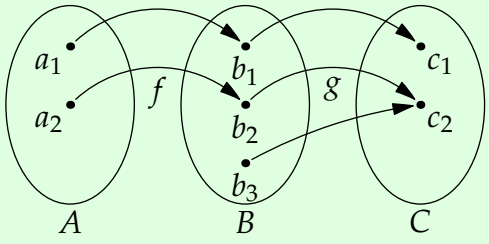
That is, $g \circ f$ is injective. We leave part 2 to the exercises. ∎

Somewhat surprisingly, the converse of this theorem is *false*. If a composition is injective or surjective, *only one* of the original functions is required also to be.

Theorem 4.23. *Suppose $f : A \to B$ and $g : B \to C$.*

1. *If $g \circ f$ is injective, then f is injective.*
2. *If $g \circ f$ is surjective, then g is surjective.*

The picture illustrates what can happen: f is only injective, *g is* only surjective, *but $g \circ f$ is* bijective.



Example 4.24. Here is a formulaic version of the picture in the theorem. Make sure you're comfortable with the definitions and draw pictures or graphs to help make sense of what's going on.

$$f : [0,2] \to [-4,4] : x \mapsto x^2 \qquad \text{(injective only)}$$
$$g : [-4,4] \to [0,16] : x \mapsto x^2 \qquad \text{(surjective only)}$$
$$g \circ f : [0,2] \to [0,16] : x \mapsto x^4 \qquad \text{(bijective!)}$$

Proof. This time we leave part 1 for the Exercises. Let $c \in C$ and assume $g \circ f$ is surjective. But then

$$\exists a \in A \text{ such that } c = (g \circ f)(a) = g\big(f(a)\big)$$

Otherwise said, $\exists b (= f(a)) \in B$ for which $c = g(b)$: that is, g is surjective. ∎

Functions and Cardinality

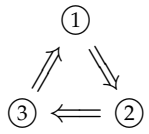
Injectivity and surjectivity are intimately tied to the notion of cardinality. In Chapter 8, we will use such functions to *define* cardinality for infinite sets. For the present we stick to finite sets.

Theorem 4.25. *Let A and B be finite sets. The following are equivalent:*

1. $|A| \leq |B|$
2. $\exists f : A \to B$ *injective*
3. $\exists g : B \to A$ *surjective*

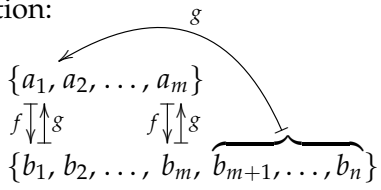
Moreover, $|A| = |B| \iff \exists f : A \to B$ *bijective.*

The theorem asserts that *any one* of the three numbered statements is true if and only if *all* are. It might appear that six arguments are required but, by proving in a circle, we only need three: for instance ① ⇒ ③ holds because ① ⇒ ② and ② ⇒ ③.
The proof is very abstract, but if you focus on the picture it should make sense.



Proof. Suppose $|A| = m$, $|B| = n$ and label the elements in roster notation:

$$A = \{a_1, a_2, \ldots, a_m\} \qquad B = \{b_1, b_2, \ldots, b_n\}$$



① ⇒ ② If $m \leq n$, *define* $f : A \to B$ by $f(a_k) = b_k$ as in the picture. This is injective since the $b_1, \ldots, b_m$ are distinct.

② ⇒ ③ Suppose $f : A \to B$ is injective. Without loss of generality, label the elements of B such that $b_k = f(a_k)$ for $1 \leq k \leq m$. *Define* the surjective function $g : B \to A$ as in the picture:[29]

$$g(b_k) := \begin{cases} a_k & \text{if } k \leq m \\ a_1 & \text{if } k > m \end{cases}$$

③ ⇒ ① Suppose $g : B \to A$ is surjective. Without loss of generality, label the elements of B such that $a_k = g(b_k)$ for $1 \leq k \leq m$. Since the b_k must be distinct, we see that $n \geq m$.

When $m = n$, the constructed functions are plainly *bijections* with $f^{-1} = g$. ∎

[29] The elements $b_{m+1}, \ldots, b_n$ could be mapped *anywhere*, we choose a_1 for simplicity.

Exercises 4.3. A reading quiz and several questions with linked video solutions can be found online.

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 $f(x) = -\sqrt{x^2+9}$.

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 it is injective.

3. (a) Find a set A so that the function $f : A \to \mathbb{R} : x \mapsto \sin x$ is injective.
   (b) Find a set B so that the function $f : \mathbb{R} \to B : x \mapsto \sin x$ is surjective.

4. A function $f : \mathbb{R} \to \mathbb{R}$ is *even* if and only if $\forall x \in \mathbb{R},\ f(-x) = f(x)$.
   (a) Prove that $f : x \mapsto x^2$ is even.
   (b) By negating the definition, state what it means for a function *not to be even.*
   (c) Give an example of a function that is *not* even: prove it.
   (d) Prove or disprove: for every f, $g : \mathbb{R} \to \mathbb{R}$ even, the composition $h = f \circ g$ is even.

5. The picture in Definition 4.16 illustrates a function
   $$f : \{a_1, a_2, a_3, a_4\} \to \{b_1, b_2, b_3, b_4\}$$

a	a_1	a_2	a_3	a_4
$f(a)$	b_1	b_2	b_3	b_3

   State the following:
   (a) range(f) (b) $f(\{a_1, a_4\})$ (c) $f^{-1}(\{b_3\})$ (d) $f^{-1}(\{b_4\})$

6. (a) Let $A = \{a,b,c\}$, $B = \{1,2,3,4\}$ and $f : A \to B$ be the function defined by $f(a) = f(c) = 1$ and $f(b) = 3$. State the following:
      i. $f^{-1}(\{1\})$ (b) $f^{-1}(\{3\})$ (c) $f^{-1}(\{1,3\})$ (d) $f^{-1}(\{2,4\})$
   (b) Let $g : [-1,\infty) \to \mathbb{R} : x \mapsto x^2 + 2x + 1$. Compute $g^{-1}((0,2])$.
   (c) Let $h : \mathbb{R} \to \mathbb{R} : x \mapsto \sin x$. Find $h^{-1}(\{-1,1\})$.

7. Define $f : (-\infty, 0] \to \mathbb{R}$ and $g : [0,\infty) \to \mathbb{R}$ by
   $$f(x) = x^2, \qquad g(x) = \begin{cases} \frac{x}{1-x} & x < 1 \\ 1 - x & x \geq 1 \end{cases}$$
   Is $g \circ f : (-\infty, 0] \to \mathbb{R}$ surjective? Justify your answer.

8. Recall Example 4.17.3. Consider the nine functions $f_k : A \to A : x \mapsto kx \pmod{10}$, where $k = 1, 2, \ldots, 9$. Find the range of each f_k. Can you find a relationship between k and the cardinality of range(f_k)?

9. 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 $f(x) = e^x$. We conclude that f is surjective. ∎

10. (a) Show there is a bijection between $\mathbb{Z}$ and $2\mathbb{Z}$.
    (b) Let S be the set of all circles in the plane which are centered at the origin. Find a bijection between S and $\mathbb{R}^+$.
    (c) Let A, B be *finite* sets. If $A \subsetneq B$, is it possible for there to be a bijection between A and B?

11. Prove that the composition of two surjective functions is surjective.

12. Suppose that $g \circ f$ is injective. Prove that f is injective.

13. Let $f : A \to B$. Prove the following:
    (a) f is injective if and only if $\forall b \in B$, $f^{-1}(\{b\})$ has *at most* one element.
    (b) f is surjective if and only if $\forall b \in B$, $f^{-1}(\{b\})$ has *at least* one element.

14. Prove that functional composition is associative. That is, if $f : A \to B$, $g : B \to C$, and $h : C \to D$ are functions, then for all $a \in A$ we have

$$\big(f \circ (g \circ h)\big)(a) = \big((f \circ g) \circ h\big)(a)$$

15. Following Theorem 4.22, the composition of bijective functions f, g is itself bijective. Give a *brief* explanation as to why $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

16. Let $f : A \to B$ and $X \subseteq A$. Fill in the blanks to complete a proof of the following facts:
    (a) $X \subseteq f^{-1}\big(f(X)\big)$. (b) If f is injective, then $X = f^{-1}\big(f(X)\big)$.

    Proof. (a) $x \in X \implies f(x) \in$ ____. Let $y = f(x)$, then $x \in$ _____ $\subseteq f^{-1}\big(f(X)\big)$.

    (b) $a \in f^{-1}\big(f(X)\big) \implies$ _____, whence $\exists x \in X$ with $f(a) = f(x)$. By injectivity, ____, whence $a \in X$. We conclude that _____. Combine with part (a) for the result. ∎

17. Let $f : A \to B$ be a function and let $Y \subseteq B$. Prove the following facts:
    (a) $f\big(f^{-1}(Y)\big) \subseteq Y$. (b) If f is surjective, then $f\big(f^{-1}(Y)\big) = Y$.

18. (Hard) Let $f : A \to B$ be a function and $X, Y \subseteq A$.
    (a) Prove that $f(X \cap Y) \subseteq f(X) \cap f(Y)$.
    (b) If f is injective, prove that $f(X \cap Y) = f(X) \cap f(Y)$.
    (c) If $f(X \cap Y) = f(X) \cap f(Y)$ for *all* $X, Y \subseteq A$, prove that f is injective.

5 Mathematical Induction and Well-ordering

In Section 2.3 we discussed three methods of proof: direct, contrapositive, and contradiction. The fourth standard method, *induction*, has a very different flavor. Before giving a formal definition, we consider how the need for induction arguments often arises.

5.1 Iterative Processes & Proof by Induction

Recursive processes are very common in mathematics and its applications: an unknown x_n satisfies a simple recurrence relation $x_{n+1} = f(x_n)$ where an initial value x_1 is given. The typical approach to such problems is to *hypothesize* a general formula $x_n = g(n)$ (spot a pattern!), and then *prove* the validity of the formula. Induction is the proof method often employed in such situations. To get us started, we investigate a famous game.

The Tower of Hanoi

Circular disks of decreasing radius are stacked on three pegs. Disks are moved one at a time from the top of a stack onto an empty peg or on top of a larger disk. If we start with 10 disks on the first peg, how many moves are required to transfer all disks to another peg?

To get a feel for the problem, try playing the game with smaller numbers of disks. Suppose n disks require r_n moves. Then:

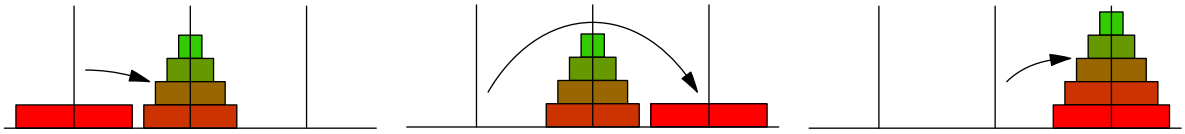
- $r_1 = 1$ since there is only one disk to move!
- $r_2 = 3$ as shown in the picture.

More experimentation will hopefully convince you that $r_3 = 7$, at which point you might be ready to hypothesize a general formula—if not, experiment more!

Conjecture 5.1. *The Tower of Hanoi with n disks requires $r_n = 2^n - 1$ moves.*

To make progress, consider a stack of $n+1$ disks. To move the largest disk, *all other disks must be stacked on a single peg.* Moving $n+1$ disks is therefore a three-step process:

1. Move the smallest n disks to another peg.
2. Move the largest disk.
3. Move the remaining disks on top of the largest.



The upshot is that r_n satisfies a *recurrence relation*: $r_{n+1} = r_n + 1 + r_n = 2r_n + 1$.

We are now in a position to prove our conjecture.

Proof. Certainly the formula $r_n = 2^n - 1$ holds when $n = 1$ disk (1 disk requires $r_1 = 1$ move).
Now suppose that n disks require $r_n = 2^n - 1$ moves, where $n \in \mathbb{N}$ is some fixed number. Then $n+1$ disks require

$$r_{n+1} = 2r_n + 1 = 2(2^n - 1) + 1 = 2^{n+1} - 1$$

moves $(*)$. Since n was arbitrary, we have in fact proved an *infinite collection of implications*:

$$r_1 = 2^1 - 1 \implies r_2 = 2^2 - 1 \implies r_3 = 2^3 - 1 \implies \cdots \implies r_n = 2^n - 1 \implies \cdots$$

Since the first of these propositions ($r_1 = 2^1 - 1$) is true, we conclude that $r_n = 2^n - 1$ for all $n \in \mathbb{N}$. ∎

To answer the original question, 10 disk require $r_{10} = 2^{10} - 1 = 1023$ moves; at one move per second this would take 17 minutes, 3 seconds.

Proof by Induction

The Tower of Hanoi argument is an example of *proof by induction.* This method is invoked when we want to prove a sequence of propositions $P(1)$, $P(2)$, $P(3)$, ..., one for each natural number. The abstract structure of an induction proof consists of two separate arguments:

(*Base case*) Prove that $P(1)$ is true.

(*Induction step*) Prove, for each $n \in \mathbb{N}$, that $P(n) \implies P(n+1)$. During this phase, $P(n)$ is termed the *induction hypothesis.*

The result is an infinite chain of implications:

$$P(1) \implies P(2) \implies P(3) \implies P(4) \implies P(5) \implies \cdots$$

Since $P(1)$ is true (base case), *all* the remaining propositions $P(2)$, $P(3)$, $P(4)$, ... are also true.
Re-read the Tower of Hanoi proof; can you separate the base case and induction step? Our presentation was informal since we were only introducing the concept. For a formal argument, a few modifications should be made.

- *Set-up* the proof: orient the reader by stating, "We prove by induction." It might also be helpful to spell out the propositions $P(n)$ and to tell the reader what variable (n) controls the induction.
- Label the *base case* and *induction step* so the reader knows what you are doing!
- After the induction step is complete, state your *conclusion.* In the above we would replace everything after $(*)$ with, "By mathematical induction, $r^n = 2^n - 1$ for all $n \in \mathbb{N}$."

Here is a straightforward and famous result, where we write the proof in our new language.

Theorem 5.2. *The sum of the first n positive integers is* $\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$

You should be familiar with summation notation $\sum_{k=1}^{n} k = 1 + 2 + 3 + \cdots + n$: if not *ask*!

Proof. We prove by induction. For each $n \in \mathbb{N}$, let $P(n)$ be the proposition $\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$.

Base case ($n = 1$): $\sum_{k=1}^{1} k = 1 = \frac{1}{2}1(1+1)$, says that $P(1)$ is true.

Induction Step: Fix $n \in \mathbb{N}$, and assume that $P(n)$ is true *for this n*. We compute the sum of the first $n+1$ positive integers and use our induction hypothesis $P(n)$ to simplify:

$$\begin{aligned}\sum_{k=1}^{n+1} k &= \sum_{k=1}^{n} k + (n+1) = \frac{1}{2}n(n+1) + (n+1) && \text{(induction hypothesis)}\\ &= \left(1 + \frac{1}{2}n\right)(n+1) = \frac{1}{2}(n+2)(n+1)\\ &= \frac{1}{2}(n+1)\big[(n+1)+1\big]\end{aligned}$$

Therefore $P(n+1)$ is true.

By mathematical induction, we conclude that $P(n)$ is true for all $n \in \mathbb{N}$. That is

$$\forall n \in \mathbb{N}, \quad \sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$$

■

Note how we grouped $\frac{1}{2}(n+1)\big[(n+1)+1\big]$ so that it is obviously the right hand side of $P(n+1)$. We present several more examples in a similar vein, but done a little faster. As is typical, we don't explicitly introduce the symbol $P(n)$, though you should feel free to continue doing this if you find it helpful. Aim for a layout like these in your formal writing.

Example 5.3. We prove by induction that, for all $n \in \mathbb{N}$,

$$2 + 5 + 8 + \cdots + (3n-1) = \frac{1}{2}n(3n+1) \qquad (*)$$

Base case ($n = 1$): The proposition $(*)$ is trivially true: $2 = \frac{1}{2} \cdot 1 \cdot (3 \cdot 1 + 1)$.

Induction Step: Fix $n \in \mathbb{N}$ and assume $(*)$ holds for this value of n. Then

$$\begin{aligned}2 + 5 + \cdots + (3n-1) + [3(n+1)-1] &= \frac{1}{2}n(3n+1) + 3n + 2\\ &= \frac{1}{2}(3n^2 + 7n + 4) = \frac{1}{2}(n+1)(3n+4)\\ &= \frac{1}{2}(n+1)\big[3(n+1)+1\big]\end{aligned}$$

which is the required proposition for $n+1$.

By mathematical induction $(*)$ holds for all $n \in \mathbb{N}$.

For brevity we labelled the desired proposition (what we'd have called $P(n)$) by $(*)$ so it could be referenced. The structure is similar to Theorem 5.2: since the goal is to evaluate a sum, the induction step is little more than adding the same thing $(3n+2)$ to both sides of the induction hypothesis. In fact, the example could have been proved directly by invoking Theorem 5.2—can you see how?

Our next two examples are a little harder, requiring more creativity to invoke the induction hypothesis. Both could alternatively be proved directly using modular arithmetic (Chapter 3).

Examples 5.4. 1. We prove by induction that $\forall n \in \mathbb{N}$, the integer $17^n - 4^n$ is divisible by 13.

Base case ($n = 1$): Plainly $17^1 - 4^1$ is divisible by 13.

Induction Step: Fix $n \in \mathbb{N}$ and assume that $17^n - 4^n = 13k$ for some $k \in \mathbb{Z}$. Then

$$\begin{aligned} 17^{n+1} - 4^{n+1} &= 17\big(17^n\big) - 4^{n+1} = 17\big(13k + 4^n\big) - 4^{n+1} \qquad \text{(induction hypothesis)} \\ &= 17 \cdot 13k + (17-4)4^n = 13\big(17k + 4^n\big) \end{aligned}$$

is divisible by 13.

By mathematical induction, $17^n - 4^n$ is divisible by 13 for all $n \in \mathbb{N}$.

2. We prove by induction: if $n \in \mathbb{N}$, then $n(n+1)(2n+1)$ is divisible by 6.

Base case ($n = 1$): The proposition reads $1 \cdot (1+1) \cdot (2 \cdot 1 + 1) = 6$, which is divisible by 6.

Induction Step: Fix $n \in \mathbb{N}$ and assume that $n(n+1)(2n+1) = 6k$ for some $k \in \mathbb{Z}$. Then

$$\begin{aligned} (n+1)(n+2)\big[2(n+1)+1\big] - 6k &= (n+1)\big[(n+2)(2n+3) - n(2n+1)\big] \\ &= (n+1)\big(2n^2 + 7n + 6 - (2n^2 + n)\big) = 6(n+1)^2 \end{aligned}$$

from which

$$(n+1)\big[(n+1)+1\big]\big[2(n+1)+1\big] = 6\big((n+1)^2 + k\big)$$

is divisible by 6.

By mathematical induction, $n(n-1)(2n-1)$ is divisible by 6 for all $n \in \mathbb{N}$.

Scratch work is your friend! Unless things are very simple, you should build an induction argument by starting with some scratch work for the hard part: the *induction step*. If you're stuck, explicitly stating the propositions $P(n)$ and $P(n+1)$ might help you see how to manipulate one into the other. Here are the propositions for Example 5.4.1:

$$\begin{aligned} P(n)&: \quad 17^n - 4^n = 13k \text{ for some integer } k \\ P(n+1)&: \quad 17^{n+1} - 4^{n+1} = 13l \text{ for some integer } l \end{aligned}$$

Since 17^n is common to both, you could try multiplying both sides of $P(n)$ by 17; if you re-read the example, you'll see that this is essentially the induction step! For Example 5.4.2, you might try multiplying out the cubic expressions

$$n(n+1)(2n+1) \text{ and } (n+1)(n+2)(2n+3)$$

and comparing coefficients. Since the leading term in both is n^3, the *difference* is quadratic and therefore much easier to think about...

Remember that scratch work isn't a proof; while your scratch work may make perfect sense to you, if a reader cannot follow it without assistance, then it isn't a proof! It is essential, once you think

you understand the induction step, to lay out the entire proof cleanly: *set-up, base case, induction step, conclusion.* As an example of what happens when you don't, here is a typical attempt at Example 5.3 by a student who is new to induction:

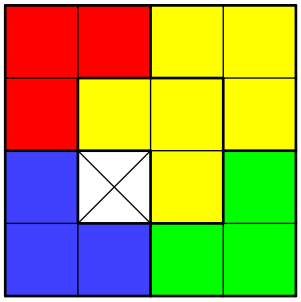
$$P(n+1) = 2+5+\cdots+(3n-1)+[3(n+1)-1] = \frac{1}{2}(n+1)\big[3(n+1)+1\big]$$
$$\frac{1}{2}n(3n+1)+(3n+2) = \frac{1}{2}(n+1)(3n+4)$$
$$3n^2+n+6n+4 = 3n^2+7n+4$$

Is this convincing to you? While there are many issues,[30] the student's work isn't without merit, since the required calculation is present (left side of 1^{st} line $=$ left side of 2^{nd}). While helpful as scratch work, a substantial re-write is needed to make this convincing.

We finish this section with a trickier example of this thinking at work.

Example 5.5. An *L-shaped tromino* is an arrangement of three squares in an "L" shape. We claim:

> If any single square is removed from a $2^n \times 2^n$ square gird, then the remaining grid may be tiled by L-shaped trominos.



The claim has the form $\forall n \in \mathbb{N}, P(n)$, but note that $P(n)$ is itself universal. The picture shows one of the *sixteen* possible examples when $n = 2$. To get an idea of how to structure the induction step, think how you might use 2×2 grids to analyse a 4×4 grid. The picture shows how!

- Whatever square we remove, one quarter of the $2^2 \times 2^2$ grid is a $2^1 \times 2^1$ grid with one square removed: this is tilable.
- Place a single tromino (yellow in the picture) so that one of its squares lies in each remaining quadrant. What's left of each quadrant is a $2^1 \times 2^1$ grid with one missing square: again tilable.

This scratch work is really an argument $P(1) \implies P(2)$! It remains only to formalize this intuition into a general proof. We proceed by induction on n.

Base case ($n = 1$): If a single square is removed from a 2×2 grid, the three remaining squares form single L-shaped tromino.

Induction step: Fix $n \in \mathbb{N}$ and assume that after removing *any* square from any $2^n \times 2^n$ grid, the remainder is tilable. Now take any $2^{n+1} \times 2^{n+1}$ grid and remove a square.

- By the induction hypothesis, the $2^n \times 2^n$ quadrant with the removed square is tilable.
- Place a single tromino in the center so that one of its squares lies in each remaining quadrant. What's left of each quadrant is a $2^n \times 2^n$ grid with one missing square: again tilable by the induction hypothesis.

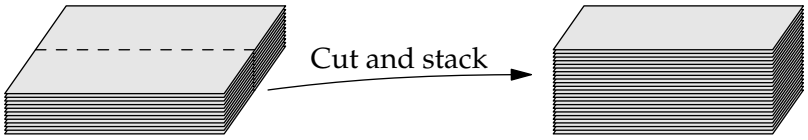
By induction, we conclude that every $2^n \times 2^n$ grid is tilable by trominos after any square is removed.

[30]
- $P(n)$ There is no *set-up*, *base case* or *conclusion* and the word *induction* is missing. The argument needs some English!
- $P(n)$ has not been *defined*. If you don't define it, don't write it!
- $P(n+1)$ is a *proposition*: it cannot *equal* a number! Replacing "$P(n+1) =$" with "$P(n+1) \iff$" would correct this.
- There are no conditional connectives to indicate the logical flow. Moreover, read top to bottom, the argument is essentially $P(n) \wedge P(n+1) \implies T$, rather than the correct induction step $P(n) \implies P(n+1)$.

Exercises 5.1. A reading quiz and several questions with linked video solutions can be found online.

1. Suppose you can move one disk on the Tower of Hanoi per second.
   (a) One of the oldest versions of the problem has monks transferring a tower of 64 disks. Roughly how many years would this take?
   (b) In a realistic human lifetime, how large a tower could be moved?

2. Imagine you cut a large large piece of paper in half and stack the two pieces on top of each other. You then repeat the process, cutting all sheets in half and making a single taller stack.



   If a single sheet of paper has thickness 0.1 mm, how many times would you have to repeat the cut-and-stack process until the stack of paper reached to the sun? ($\approx$ 150 million kilometers). *Prove* that you are correct.

3. A room contains n people. Everybody wants to shake everyone else's hand (but not their own).
   (a) Suppose n people require h_n handshakes. If person $n+1$ 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 by induction.

4. (a) In Example 5.4.2, what is the proposition $P(n+1)$?
   (b) In the induction step of Example 5.4.2, explain why it would be incorrect to write

$$\begin{aligned} P(n+1) - P(n) &= (n+1)\big[(n+2)(2n+3) - n(2n+1)\big] \\ &= (n+1)(2n^2+7n+6-2n^2-n) \\ &= 6(n+1)^2 \end{aligned}$$

   (c) Extend the Example by proving, by induction, that $\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$.

5. Prove by induction that for each natural number n, we have $\sum_{k=0}^{n} 2^k = 2^{n+1} - 1$.

6. Consider the statement: 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 $\sum_{k=1}^{4} k^3$?
   (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) Proof the statement by induction.

7. (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 $n \geq 1$.

8. Prove *by induction* that for every $n \in \mathbb{N}$ we have $n \equiv 5$ or $n \equiv 6$ or $n \equiv 7 \pmod 3$.

9. Prove by induction that, for all $n \in \mathbb{N}$,

$$1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n(n+1) = \frac{1}{3} n(n+1)(n+2)$$

10. (a) Show, by induction, that for all $n \in \mathbb{N}$, the number 4 divides the integer $11^n - 7^n$.
    (b) More generally, prove by induction that $(a-b) \mid (a^n - b^n)$ for any $a, b, n \in \mathbb{N}$.

11. (a) Find a formula for the sum of the first n odd natural numbers. Prove your assertion.
    (b) Use Theorem 5.2 to give an alternative direct proof of your formula.

12. Find the error in the following "proof" of the statement, "All cats have the same color fur."

> *Proof.* Let $P(n)$ be the proposition, "Any set of n cats have the same color fur." We prove by induction on n.
>
> *Base case* ($n = 1$): Any cat has the same color fur as itself.
>
> *Induction step*: Fix $n \in \mathbb{N}$ and assume $P(n)$. Take any set $S = \{C_1, C_2, \ldots, C_{n+1}\}$ of $n+1$ cats. The set $S \setminus \{C_1\}$ has n cats; by the induction hypothesis all have the same color fur. Again by the induction hypothesis, all cats in $S \setminus \{C_2\}$ have the same color fur. Combining these observations, we see that all cats in S have the same color fur. Since S was arbitrary, we see that $P(n+1)$ holds.
>
> By induction, $P(n)$ is true for all $n \in \mathbb{N}$, which establishes the claim. ∎

13. Use induction, the product rule, and the fact that $\frac{\mathrm{d}}{\mathrm{d}x} x = 1$ to prove the power law from calculus:

$$\forall n \in \mathbb{N},\ \frac{\mathrm{d}}{\mathrm{d}x} x^n = n x^{n-1}$$

14. A (real) *polynomial* of degree n is a function $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, whose *coefficients* a_k are real numbers and where $a_n \neq 0$.

    (a) Prove: for all $n \in \mathbb{N}$,

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{x^2} = p_n(x) e^{x^2}$$

    where $p_n(x)$ is some polynomial of degree n.

    (b) (Hard) Let $p(x)$ be a polynomial of degree $n \geq 1$. Show p has at most n roots.
    (*Hint: induct on the degree n*)

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 Well-ordering and the Principle of Mathematical Induction

In this section we think more carefully about the logic behind induction, by tying it to a fundamental property of the natural numbers.

Definition 5.6. A non-empty set of real numbers A is *well-ordered* if every non-empty subset of A contains a minimum element.

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 equivalently as follows, where in the second line we expand what is meant by a *minimum*:

- $\forall B \subseteq A$ such that $B \neq \emptyset$, $\min B$ exists.
- $\forall B \subseteq A$, $B \neq \emptyset \implies \exists b \in B$ such that $\forall x \in B$, $b \leq x$.

To show that A is *not* well-ordered, we need only exhibit a non-empty subset B with *no minimum.*

Examples 5.7. 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? It might be easier to think about why any *finite* set $A = \{a_1, \ldots, a_n\} \subseteq \mathbb{R}$ is well-ordered...

2. The interval $A = [3, 10)$ is not well-ordered. Indeed $B = (3, 4)$ is a non-empty subset which has no minimum element. While you should believe this, let's prove it anyway!

   We need to prove that $\forall b \in B, \exists x \in B$ with $x < b$. Given any $b \in (3, 4)$, observe that $x := \frac{b+3}{2}$ satisfies

   $$3 < x < b < 4 \quad \text{from which} \quad x \in B \text{ and } x < b$$

   You could also argue by contradiction (if $b \in B$ is $\min B$, then...).

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.

You might suspect (wrongly!) that every well-ordered set is finite. That the natural numbers form a well-ordered *infinite* set is, for us, an axiom: [31] a foundational claim that forms part of our basic conception of the natural numbers.

Axiom 5.8 (Well-ordering Principle). $\mathbb{N}$ is well-ordered.

Also known as the *least natural number principle*, the well-ordering principle is used repeatedly in mathematics. In fact we've already done so in this text! Consider the set of positive remainders generated by the Euclidean algorithm (Theorem 3.16) applied to $a > b \in \mathbb{N}$:

$$\{\ldots, r_2, r_1, b, a\} \subseteq \mathbb{N}$$

Well-ordering guarantees that this set has a minimal element r_t (which turns out to be $\gcd(a, b)$); this is essentially the argument for Exercise 3.2.8(a).

[31] There are many ways to define the natural numbers. Typically well-ordering is either an axiom (essentially part of the definition) or a theorem. Compare with Exercise 11 for an alternative approach.

Armed with this axiom, we can justify the method of proof by induction.

Theorem 5.9 (Principle of Mathematical Induction). *For each $n \in \mathbb{N}$, let $P(n)$ be a proposition. Additionally make the two standard assumptions for induction:*

Base case: $P(1)$ *is true*

Induction step: $\forall n \in \mathbb{N},\ P(n) \implies P(n+1)$

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Before attempting a proof, consider how the theorem could be written as a pure implication:

$$P(1) \wedge \big(\forall n \in \mathbb{N}, P(n) \implies P(n+1)\big) \implies \big(\forall n \in \mathbb{N}, P(n)\big)$$

This helps us select a proof strategy: a direct approach seems hard since the conclusion is universal; a contrapositive approach requires an ugly negation of the hypothesis; a proof by contradiction seems most sensible since negation of the conclusion is straightforward.

Proof. We argue by contradiction. Assume the base case, the induction step, and that $\exists n \in \mathbb{N}$ for which $P(n)$ is *false*. The set of natural numbers

$$S := \{k \in \mathbb{N} : P(k) \text{ is false}\}$$

is non-empty (since $n \in S$). Well-ordering guarantees that $s := \min S$ exists. Observe:

- $s \in S \implies P(s)$ is *false*.
- The *base case* tells us that $s \neq 1$, whence $s - 1 \in \mathbb{N}$.
- $s - 1 < \min S \implies P(s-1)$ is *true*.
- The *induction step* ($P(s-1) \implies P(s)$) tells us that $P(s)$ is *true*.

Contradiction: $P(s)$ cannot be both true and false! ∎

Now we have the proof, it is straightforward to extend the principle of induction. For any integer m (positive, negative or zero), the set

$$\mathbb{Z}_{\geq m} = \{n \in \mathbb{Z} : n \geq m\} = \{m, m+1, m+2, m+3, \ldots\}$$

is also well-ordered. By changing the base case to $P(m)$ and replacing $\mathbb{N}$ with $\mathbb{Z}_{\geq m}$, we immediately obtain the proof of a more general principle of induction.

Corollary 5.10 (Induction with base case m). *Fix an integer m. For each integer $n \geq m$, let $P(n)$ be a proposition. Suppose:*

Base case: $P(m)$ *is true*

Induction step: $\forall n \in \mathbb{Z}_{\geq m},\ P(n) \implies P(n+1)$

Then $P(n)$ is true for all $n \in \mathbb{Z}_{\geq m}$.

The intuitive concept is exactly as before, just with a different base case!

$$P(m) \implies P(m+1) \implies P(m+2) \implies P(m+3) \implies \cdots$$

Examples 5.11. 1. For all integers $n \geq 2$, we have[32]

$$\sum_{k=2}^{n} \frac{1}{k(k-1)} = 1 - \frac{1}{n} \tag{$*$}$$

Base case ($n = 2$): When $n = 2$, $(*)$ reads $\sum_{i=2}^{2} \frac{1}{i(i-1)} = \frac{1}{2} = 1 - \frac{1}{2}$.

Induction step: Assume that $(*)$ is true for some fixed $n \geq 2$. Then

$$\begin{aligned} \sum_{i=2}^{n+1} \frac{1}{k(k-1)} &= \sum_{i=2}^{n} \frac{1}{k(k-1)} + \frac{1}{(n+1)n} = 1 - \frac{1}{n} + \frac{1}{n(n+1)} \quad \text{(induction hypothesis)} \\ &= 1 - \frac{(n+1)-1}{n(n+1)} = 1 - \frac{1}{n+1} \end{aligned}$$

which is exactly $(*)$ when n is replaced by $n+1$.

By induction $(*)$ holds for all integers $n \geq 2$.

2. For all integers $n \geq 4$, we claim that $3^n > n^3$.

We really need a different base case: when $n = 3$, the claim $3^3 > 3^3$ is false! As is often the case, it helps to do some scratch work first. The induction hypothesis allows us to see that

$$3^{n+1} = 3 \cdot 3^n > 3n^3$$

The proof of the induction step therefore hinges on being able to show that $3n^3 \geq (n+1)^3$. There are many ways to convince yourself of this: for instance

$$3n^3 \geq (n+1)^3 \iff 3 \geq \left(\frac{n+1}{n}\right)^3 = \left(1 + \frac{1}{n}\right)^3 \tag{$\dagger$}$$

The right side *decreases* as n increases; since $n \geq 4$, the right side is at most $\left(\frac{5}{4}\right)^3 = \frac{125}{64} < 2$, whence $(*)$ holds for all $n \geq 4$.

We now prove the original claim by induction.

Base case ($n = 4$): Observe that $3^4 = 81 > 64 = 4^3$.

Induction step: Fix $n \in \mathbb{Z}_{\geq 4}$ and suppose that $3^n > n^3$. By $(\dagger)$, we see that

$$3^{n+1} = 3 \cdot 3^n > 3n^3 \geq (n+1)^3$$

By induction, we conclude that $3^n > n^3$ whenever $n \in \mathbb{Z}_{\geq 4}$.

---

[32] You might have encountered this example as a *telescoping series* in calculus:

$$\sum_{k=2}^{n} \frac{1}{k(k-1)} = \frac{1}{2 \cdot 1} + \frac{1}{3 \cdot 2} + \cdots + \frac{1}{n(n-1)} = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) = 1 - \frac{1}{n}$$

In calculus, you'd take the limit as $n \to \infty$ to obtain the infinite series $\sum_{k=2}^{\infty} \frac{1}{k(k-1)} = 1$

As a final example, we prove an extended version of de Morgan's law for sets (Theorem 4.14(a)).

Example 5.12. For any collection of sets $A_1, \dots, A_n$ where $n \geq 2$, we have

$$(A_1 \cap \cdots \cap A_n)^{\mathsf{C}} = {A_1}^{\mathsf{C}} \cup \cdots \cup {A_n}^{\mathsf{C}} \tag{‡}$$

Base case ($n = 2$): $A_1 \cap {A_2}^{\mathsf{C}} = {A_1}^{\mathsf{C}} \cup {A_2}^{\mathsf{C}}$ is precisely the standard de Morgan identity.

Induction step: Fix $n \in \mathbb{N}_{\geq 2}$ and suppose (‡) holds for *all* collections of n sets. Given a collection of $n+1$ sets, we see that

$$\begin{aligned}
(A_1 \cap \cdots \cap A_n \cap A_{n+1})^{\mathsf{C}} &= \big((A_1 \cap \cdots \cap A_n) \cap A_{n+1}\big)^{\mathsf{C}} \\
&= (A_1 \cap \cdots \cap A_n)^{\mathsf{C}} \cup {A_{n+1}}^{\mathsf{C}} && \text{(de Morgan again!)} \\
&= {A_1}^{\mathsf{C}} \cup \cdots \cup {A_n}^{\mathsf{C}} \cup {A_{n+1}}^{\mathsf{C}} && \text{(induction hypothesis)}
\end{aligned}$$

By induction the claim (‡) holds for any collection of n sets.

We could have approached the argument as a standard induction with base case $n = 1$. Instead we deliberately chose the base case $n = 2$, both to avoid confusion (the trivial ${A_1}^{\mathsf{C}} = {A_1}^{\mathsf{C}}$ isn't helpful or interesting) and to highlight the importance of de Morgan's law for two sets to the entire argument.

Proof by Minimal Counter-example

Sometimes authors present induction arguments as contradiction proofs in line with Theorem 5.9: the element $s = \min S$ is known as the *minimal counter-example.* Here are two variations on this idea; the first is a straight translation of an induction where the base case and induction step are clear.

Examples 5.13. 1. We prove: for all $n \in \mathbb{N}_0$, $\sum_{k=0}^{n} 2^k = 2^{n+1} - 1$.

Suppose to the contrary and let $s \geq 0$ be the minimal counter-example.

- Since $\sum_{k=0}^{0} 2^k = 2^0 = 1 = 2^{0+1} - 1$, we see that $s \neq 0$. (*base case*)
- But then

$$\sum_{k=0}^{s} 2^k = 2^s + \sum_{k=0}^{s-1} 2^k = 2^s + 2^s - 1 = 2^{s+1} - 1 \qquad \text{(\textit{induction step})}$$

contradicts the fact that s is a counter-example!

2. We re-prove Theorem 2.37: $\sqrt{2} \notin \mathbb{Q}$. Suppose $\sqrt{2}$ is rational and consider the set

$$S = \{x \in \mathbb{N} : \exists y \in \mathbb{N} \text{ with } x^2 = 2y^2\}$$

Let $s = \min S$ and $t \in \mathbb{N}$ be such that $s^2 = 2t^2$: plainly $t < s$. Since s^2 even $\implies s$ even, we can write $s = 2k$, from which

$$4k^2 = 2t^2 \implies n^2 = 2k^2 \implies t \in S$$

This contradicts the minimality of m.

This approach is often used in number theory in the guise of Fermat's *method of infinite descent.*

Aside: Well-ordering more generally Definition 5.6 is a weak version of a much deeper concept. Informally, to *well-order* a set means to list its elements in some order so that every non-empty subset has an initial element *with respect to that order.*
For instance, the set of negative integers $\mathbb{Z}^- = \{\ldots, -4, -3, -2, -1\}$ is *not* well-ordered with respect to the standard ordering of the integers, but is well-ordered with respect to the *reverse* ordering

$$\mathbb{Z}^- = \{-1, -2, -3, -4, \ldots\}$$

The principle of mathematical induction is easily modified to accommodate theorems of the form $\forall n \in \mathbb{Z}^-,\ P(n)$: the base case is $P(-1)$ and the induction step justifies the chain

$$P(-1) \implies P(-2) \implies P(-3) \implies \cdots$$

All the infinite well-ordered sets we've thus far seen have "looked like" the natural numbers, however more esoteric examples exist. For instance, the following well-ordered set looks like two copies of the natural numbers, one following the other:

$$A = \left\{0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots, 1, \frac{3}{2}, \frac{5}{3}, \frac{7}{4}, \frac{9}{5}, \ldots\right\} = \left\{1 - \frac{1}{n} : n \in \mathbb{N}\right\} \cup \left\{2 - \frac{1}{n} : n \in \mathbb{N}\right\}$$

Every non-empty subset of A really does have a minimum! It is possible to modify induction to apply to propositions indexed by well-ordered sets like this, though an extra step is required to deal with *limit elements* (like $1 \in A$) with no immediate predecessor. If your further studies include set theory, you'll likely spend much time considering well-orders and their associated *ordinals.*

Exercises 5.2. A reading quiz and several questions with linked video solutions can be found online.

1. Prove that the interval $(-2, 5]$ has no minimum element.
2. Prove that every finite set of real numbers is well-ordered.
3. (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 $n \geq 5$.
4. 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) Prove the result by induction.
5. Prove the geometric series formula: if $r \neq 1$ and $n \in \mathbb{N}_0$, then $\sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r}$
6. For all integers $n \geq 3$, prove that $\sum_{k=3}^{n} \frac{1}{k(k-2)} = \frac{3}{4} - \frac{2n-1}{2n(n-1)}$
7. Prove: for any $n \in \mathbb{N}$, $\sum_{i=1}^{n} \frac{1}{i^2} < 2$
   (*Hint: prove the stronger fact that* $\sum_{i=1}^{n} \frac{1}{i^2} < 2 - \frac{1}{n}$ *for all* $n \geq 2$)

8. The set $A_3 = \{1, 2, 3\}$ satisfies the property that the sum of its elements ($1 + 2 + 3 = 6$) is divisible by every element of A_3.

   (a) Use induction to prove that for any $n \geq 3$, there is a set A_n of n natural numbers such that the sum of its elements is divisible by every element of A_n.

   (b) Prove by contradiction that no set of *two* natural numbers satisfies this property.

9. Suppose that $x^2 + 4y^2 = 3z^2$ has a solution (x, y, z) where all three are *positive integers.*

   (a) By considering remainders modulo 3, prove that $3 \mid z$. Thus create a new solution (X, Y, Z) in positive integers, where $Z < z$.

   (b) Use the method of minimal counter-example to prove that $x^2 + 4y^2 = 3z^2$ has no solutions where $x, y, z \in \mathbb{N}$.

10. We use the fact that $\mathbb{N}_0$ is well-ordered to prove the division algorithm (Theorem 3.3).

    If $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, then $\exists$ unique $q, r \in \mathbb{Z}$ such that $m = qn + r$ and $0 \leq r < n$.

    Given m, n, define

    $$S = \mathbb{N}_0 \cap (m + n\mathbb{Z}) = \{k \in \mathbb{N}_0 : k = m - qn \text{ for some } q \in \mathbb{Z}\}$$

    (a) (Existence) Show that S is a *non-empty* subset of $\mathbb{N}_0$. By well-ordering, *define* $r := \min S$. Prove that $0 \leq r < n$.

    (b) (Uniqueness) Suppose two pairs of integers (q_1, r_1) and (q_2, r_2) satisfy $m = q_i n + r_i$ and $0 \leq r_1, r_2 < n$. Prove that $r_1 = r_2$.

11. (Hard) We consider a version of Peano's axioms for the natural numbers.

    i. (*Initial element*) $1 \in \mathbb{N}$

    ii. (*Successor function*) $f(n) = n + 1$ is a function $f : \mathbb{N} \to \mathbb{N}$

    iii. (*No predecessor of the initial element*) $1 \notin \text{range}(f)$

    iv. (*Unique predecessor/order*) f is injective: $m + 1 = n + 1 \implies m = n$

    v. (*Induction*) Any subset $A \subseteq \mathbb{N}$ with the following properties *equals* $\mathbb{N}$:

    $$1 \in A \quad \text{and} \quad \forall a \in A,\ a + 1 \in A$$

    (a) Replace $\mathbb{N}$ with $\mathbb{Z}$ in each axiom. Which are true and which false?

    (b) Let $T = \{(m, n) : m, n \in \mathbb{N}\}$ be the set of all ordered pairs of natural numbers.

    i. Let $f : T \to T$ be the function $f(m, n) = (m + 1, n)$. Letting the pair $(1, 1)$ play the role of 1, and f the successor function, decide which of Peano's axioms are satisfied by T.

    ii. Repeat the question for the same initial element and

    $$f : T \to T : (m, n) \mapsto \begin{cases} (m - 1, n + 1) & \text{if } m \geq 2 \\ (m + n, 1) & \text{if } m = 1 \end{cases}$$

    (c) Prove that $\text{range}(f) = \mathbb{N} \setminus \{1\}$: every element except 1 is the successor of something. (*Hint: let $A = \{1\} \cup \text{range}(f)$ in the induction axiom*)

    (d) Prove that $\mathbb{N}$, as defined by Peano, is well-ordered (with respect to $x < x + 1$, etc.).

5.3 Strong Induction

The principle of mathematical induction (Theorem 5.9) is often known as *weak* induction. The primary difference with *strong* induction is that the induction step assumes more than one proposition.

Theorem 5.14 (Principle of Strong Induction). *Let $l \geq m$ be fixed integers and suppose $P(n)$ is a proposition, one for each $n \in \mathbb{Z}_{\geq m}$. Suppose:*

Base case(s): $P(m), P(m+1), \ldots, P(l)$ *are true*

Induction step: $\forall n \geq l,\ \big(P(m) \wedge P(m+1) \wedge \cdots \wedge P(n)\big) \implies P(n+1)$

Then $P(n)$ is true for all $n \geq m$.

Exercise 6 provides a proof by showing that strong and weak induction are equivalent. We instead concentrate on a few examples. The main additional challenge of strong induction lies in determining how many base cases are required and in identifying precisely how to phrase the induction hypothesis: in practice one rarely needs to use all the propositions $P(m), \ldots, P(n)$ explicitly.

Example 5.15 (Fibonacci numbers). This famous sequence $(f_n)_{n=1}^{\infty} = (1, 1, 2, 3, 5, 8, 13, 21, \ldots)$ is defined by the 2^{nd}-order recurrence relation

$$\begin{cases} f_{n+1} = f_n + f_{n-1} & \text{if } n \geq 2 \\ f_1 = f_2 = 1 \end{cases}$$

While the Fibonacci sequence seems to be increasing, it also appears to be less than doubling at each step, suggesting the claim

$$\forall n \in \mathbb{N},\ f_n < 2^n$$

We prove this using strong induction. *Two base cases* are suggested since the sequence is defined by two initial conditions ($f_1 = f_2 = 1$): in the language of the Theorem, $m = 1$ and $l = 2$. Moreover, the fact that each term from f_3 onwards is the sum of its *two predecessors* suggests that the induction step requires the explicit use of two propositions.

Proof. *Base cases* ($n = 1, 2$): $f_1 = 1 < 2^1$ and $f_2 = 1 < 2^2$.

Induction step: Fix $n \geq 2$ and suppose[33] that $f_{n-1} < 2^{n-1}$ and $f_n < 2^n$. Then

$$f_{n+1} = f_n + f_{n-1} < 2^n + 2^{n-1} < 2^n + 2^n = 2^{n+1}$$

By induction, $f_n < 2^n$ for all $n \in \mathbb{N}$. ∎

The Fibonacci numbers satisfy many identities which can often be established by induction (e.g., Exercises 3 & 4).

[33] To follow Theorem 5.14 precisely, we would assume that $f_k < 2^k$ for *all* $k \leq n$. Feel free to write this if you like, though our phrasing is more typical. Since we only make explicit use of two cases in the induction step, it is clearer to state these concretely rather than introducing a new variable k.

Before seeing further examples, it is instructive to consider why we really needed strong induction to prove our Fibonacci example. Here are two broken attempts to prove the claim by weak induction.

Broken Proof A. Base case ($n = 1$): $f_1 = 1 < 2^1$.

Induction step: Fix $n \geq 2$ and suppose that $f_n < 2^n$. Then[34]

$$f_{n+1} = f_n + f_{n-1} < 2^n + ????$$

What is the problem? The induction hypothesis assumes $f_n < 2^n$, but not anything about f_{n-1}: we are stuck! Let's correct this flaw by making the induction hypothesis as in the correct proof.

Broken Proof B. Base case ($n = 1$): $f_1 = 1 < 2^1$.

Induction step: Fix $n \geq 2$ and suppose that $f_{n-1} < 2^{n-1}$ and $f_n < 2^n$. Then

$$f_{n+1} = f_n + f_{n-1} < 2^n + 2^{n-1} < 2^n + 2^n = 2^{n+1}$$

By induction, $f_n < 2^n$ for all $n \in \mathbb{N}$.

Where is the problem now? Consider the first instance, $n = 2$, in which the induction step is invoked:

$$f_3 = f_2 + f_1 < 2^2 + 2^1$$

We haven't proved enough base cases to get us started: the single base case establishes $f_1 < 2^1$, but not $f_2 < 2^2$. The induction step correctly establishes the chain of implications

$$P(1) \wedge P(2) \implies P(3), \quad P(3) \wedge P(4) \implies P(5), \quad P(4) \wedge P(5) \implies P(6), \ldots$$

but we can only get the process started if we prove *both* base cases $P(1)$ *and* $P(2)$.
The moral here is to try the induction step as scratch work. Your attempt should tell you *if* you need strong induction and *how many* base cases are required.

Example 5.16. A sequence of integers $(a_n)_{n=0}^{\infty}$ is defined by

$$\begin{cases} a_{n+1} = 5a_n - 6a_{n-1} & \text{if } n \geq 1 \\ a_0 = 0,\ a_1 = 1 \end{cases}$$

We prove by induction that $a_n = 3^n - 2^n$ for all $n \in \mathbb{N}_0$.

Proof. Base cases ($n = 0, 1$): $a_0 = 0 = 3^0 - 2^0$ and $a_1 = 1 = 3^1 - 2^1$.

Induction step: Fix $n \geq 1$ 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 induction, $a_n = 3^n - 2^n$ for all $n \in \mathbb{N}_0$. ∎

[34] The induction step requires $n \geq 2$ so as to invoke the recurrence relation: $f_{n+1} = f_n + f_{n-1}$ is meaningless when $n = 1$ ($f_{n-1} = f_0$ doesn't exist).

For another sequential induction example in the same vein, see Exercise 5.3 where *three* base cases are required, and the induction step explicitly uses *three* propositions.

To see strong induction in all its glory, where the induction step requires *all* previous propositions, we prove the existence part of the Fundamental Theorem of Arithmetic, which states that all integers ≥ 2 can be (uniquely) expressed as a product of primes: e.g., $3564 = 2^2 \times 3^4 \times 11$.

Theorem 5.17. *Every integer $n \geq 2$ is either prime or a product of primes.*

This provides the missing piece in our discussion of Euclid's Theorem (2.41) on the existence of infinitely many primes. First recall Definition 2.40: $p \in \mathbb{N}_{\geq 2}$ is *prime* if and only if its only positive divisors are itself and 1. A non-prime $q \in \mathbb{N}_{\geq 2}$ is said to be *composite*: $\exists a, b \in \mathbb{N}_{\geq 2}$ such that $q = ab$.

Proof. We prove by induction.

Base case ($n = 2$): The only positive divisors of 2 are itself and 1, hence 2 is prime.

Induction step: Fix $n \geq 2$ and suppose 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. Certainly $n + 1$ is divisible by a prime!
- $n + 1$ is composite. Then $n + 1 = ab$ for some natural numbers $a, b \geq 2$. Clearly $a, b \leq n$; by the induction hypothesis, *both* are prime or the product of primes. Therefore $n + 1$ is also the product of primes.

By induction we see that all natural numbers $n \geq 2$ are either prime, or a product of primes. ∎

Think carefully about why *only one* base case is required!

Exercises 5.3. A reading quiz and several questions with linked video solutions can be found online.

1. Define sequence $(b_n)_{n=1}^{\infty}$ as follows:

$$\begin{cases} b_{n+1} = 2b_n + b_{n-1} & \text{if } n \geq 2 \\ b_1 = 3,\ b_2 = 6 \end{cases}$$

   Prove by induction that b_n is divisible by 3 for all $n \in \mathbb{N}$.

2. Define a sequence $(c_n)_{n=0}^{\infty}$:

$$\begin{cases} c_{n+1} = \frac{49}{8} c_n - \frac{225}{8} c_{n-2} & \text{if } n \geq 2 \\ c_0 = 0,\ c_1 = 2,\ c_2 = 16 \end{cases}$$

   Prove by induction (use three base cases!) that $c_n = 5^n - 3^n$ for all $n \in \mathbb{N}_0$.

3. Let f_n be the n^{th} Fibonacci number (Example 5.15). Prove the following by induction $\forall n \in \mathbb{N}$:

   (a) $\sum_{k=1}^{n} f_k^2 = f_n f_{n+1}$ (b) $f_n \geq \left(\frac{3}{2}\right)^{n-2}$

   (*Hints: Weak induction is good enough for (a); why?*)

4. Extending Exercise 3(b), prove *Binet's formula* for the n^{th} Fibonacci number:

$$f_n = \frac{1}{\sqrt{5}}(\phi^n - \hat{\phi}^n) \quad \text{where} \quad \phi = \frac{1}{2}(1+\sqrt{5}) \text{ and } \hat{\phi} = \frac{1}{2}(1-\sqrt{5})$$

(ϕ *is the famous* golden ratio*:* $\phi, \hat{\phi}$ *are the solutions to the quadratic equation* $x^2 = x + 1$)

5. Prove by induction that every $n \in \mathbb{N}$ can be written in the form

$$n = 2^{r_1} + 2^{r_2} + \cdots + 2^{r_\ell}$$

for some $\ell \in \mathbb{N}$ and *distinct* integers $r_1, r_2, \ldots r_\ell \geq 0$.

(*Hints: try proving in the style of Theorem 5.17; consider the cases when* $n+1$ *is even/odd separately*)

6. Prove the principle of strong induction (Theorem 5.14) by applying *weak induction* to a new family of propositions $Q(n)$ via:

$$Q(n) \iff P(m) \wedge P(m+1) \wedge \cdots \wedge P(n)$$

7. Consider the proof of Theorem 5.17.

   (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 $n+1 = 45$?

   (c) 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.

8. In this question we use the alternative definition of prime (Exercise 3.2.13).[35]

   An integer $p \geq 2$ is *prime* if and only if $\forall a, b \in \mathbb{N},\ p \mid ab \implies p \mid a$ or $p \mid b$.

   Let p be prime, let $n \in \mathbb{N}$, and let $a_1, \ldots, a_n$ be natural numbers such that $p \mid a_1 a_2 \cdots a_n$. Prove by induction that,

   $$p \mid a_i \text{ for some } i \in \{1, 2, \ldots, n\}$$

   (*Hint:* $n = 1$ *isn't really part of the induction, but you can treat it as a base case*)

9. The *Fundamental Theorem of Arithmetic* states that every integer $n \geq 2$ can be written as a product of prime factors in a *unique* way (up to reordering of the prime factors). In other words,

   i. $n = p_1 p_2 \cdots p_k$ for some primes $p_1, p_2, \ldots, p_k$, and,

   ii. 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.

   Part i. is Theorem 5.17. Using Exercise 8, or otherwise, supply a proof of part ii.

[35] This is the strict definition of *prime,* while Definition 2.40 is the definition of *irreducible.* For the integers, Exercise 3.2.13 says that these concepts are synonymous.

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 co-ordinates* (x, y), an *ordered pair* where each co-ordinate (x, y) is a member of the set $\mathbb{R}$. The same approach can be used for any sets.

Definition 6.1. The *Cartesian product* of sets A and B is the set of ordered pairs

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$$

Otherwise said: $(a, b) \in A \times B \iff a \in A$ and $b \in B$.

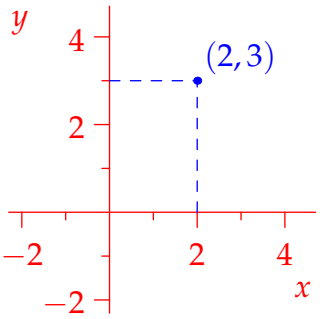
Examples 6.2. 1. The Cartesian product of the real line $\mathbb{R}$ with itself is the usual xy-plane.

As you've seen in other classes, rather than writing $\mathbb{R} \times \mathbb{R}$ which is unwieldy, we denote this set

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$$

More generally, the set of n-tuples of real numbers is[36]

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{R}\} = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ times}}$$



2. If $A = \{1, 2, 3\}$ and $B = \{\alpha, \beta\}$, then the Cartesian product of A and B is

$$A \times B = \{(1, \alpha),\ (1, \beta),\ (2, \alpha),\ (2, \beta),\ (3, \alpha),\ (3, \beta)\}$$

The order of terms in an ordered pair really matters: the Cartesian product with the roles reversed is

$$B \times A = \{(\alpha, 1),\ (\beta, 1),\ (\alpha, 2),\ (\beta, 2),\ (\alpha, 3),\ (\beta, 3)\}$$

While $A \times B \neq B \times A$, note that both Cartesian products have *six* $(= 3 \cdot 2 = 2 \cdot 3)$ elements.

3. The menu in a restaurant might be summarized set-theoretically:

$$\text{Mains} = \{\text{fish, steak, eggplant, pasta}\}, \qquad \text{Sides} = \{\text{asparagus, salad, potatoes}\}$$

The Cartesian product Mains $\times$ Sides is the set of all possible meals consisting of one main and one side. It should be obvious that there are $4 \times 3 = 12$ possible meal choices.

The last two examples illustrate one of the simplest properties of Cartesian products, in a result which indeed justifies the very use of the word *product*!

[36] Strictly this should be defined inductively, e.g., $\mathbb{R}^3 := \mathbb{R}^2 \times \mathbb{R} = \{((x, y), z) : x, y, z \in \mathbb{R}\}$, but this is very tedious!

Theorem 6.3. *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

$$A \times B = \left\{ \begin{array}{cccc} (a_1, b_1), & (a_1, b_2), & (a_1, b_3), & \cdots \quad (a_1, b_n), \\ (a_2, b_1), & (a_2, b_2), & (a_2, b_3), & \cdots \quad (a_2, b_n), \\ \vdots & \vdots & \vdots & \vdots \\ (a_m, b_1), & (a_m, b_2), & (a_m, b_3), & \cdots \quad (a_m, b_n) \end{array} \right\}$$

Every element of $A \times B$ is listed exactly once. There are m rows and n columns, so $|A \times B| = mn$. ∎

Set Identities These may be established as we've done previously (Section 4): convert everything into propositions regarding elements of sets and use basic logic. If you're feeling more confident, you might also be able to invoke previously established rules of set algebra.

Examples 6.4. 1. Consider the complement of a Cartesian product $A \times B$. If you had to guess an expression for $(A \times B)^C$, you might mistakenly think it is $A^C \times B^C$. Let us think more carefully:

$$\begin{aligned} (x, y) \in (A \times B)^C &\iff \neg((x, y) \in A \times B) && \text{(definition of complement)} \\ &\iff \neg(x \in A \text{ and } y \in B) && \text{(definition of } A \times B\text{)} \\ &\iff x \notin A \text{ or } y \notin B && \text{(de Morgan (logic))} \end{aligned}$$

By contrast, $(x, y) \in A^C \times B^C \iff x \notin A$ and $x \notin B$, so $(A \times B)^C \neq A^C \times B^C$. Indeed the complement of a Cartesian product is *not a Cartesian product!* For more on this, see Exercise 5.

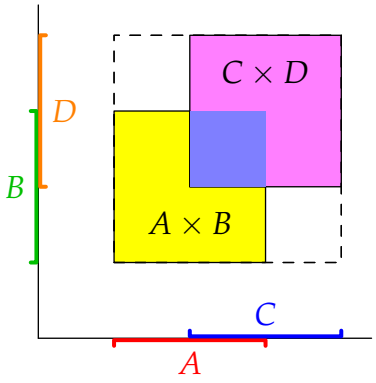
2. Let A, B, C, D be any sets. We prove that $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.

$$\begin{aligned} (x, y) \in (A \times B) \cup (C \times D) &\implies (x, y) \in A \times B \text{ or } (x, y) \in C \times D \\ &\implies (x \in A \text{ and } y \in B) \text{ or } (x \in C \text{ and } y \in D) \\ &\implies (x \in A \text{ or } x \in C) \text{ and } (y \in B \text{ or } y \in D) \\ &\implies x \in A \cup C \text{ and } y \in B \cup D \\ &\implies (x, y) \in (A \cup C) \times (B \cup D) \end{aligned}$$

All implications except the red arrow are in fact *biconditional.* To convince yourself of its truth, first consider what must be true about x, then do the same for y. The picture provides a visualization where A, B, C, D are intervals of real numbers:

- $(A \times B) \cup (C \times D)$ is the solid shaded region.
- $(A \cup C) \times (B \cup D)$ is the larger dashed square.

If $x \in C \setminus A$ and $y \in B \setminus D$, then $(x, y) \in (A \cup C) \times (B \cup D)$ but $(x, y) \notin (A \times B) \cup (C \times D)$, so we do not, in general, expect these sets to be equal.



Exercises 6.1. A reading quiz and several questions with linked video solutions can be found online.

1. (a) Suppose that $A = \{1, 2\}$ and $B = \{3, 4, 5\}$. State the set $A \times B$ in roster notation.
   (b) Sketch both $A \times B$ and $B \times A$ using dots in $\mathbb{R}^2$. What do you observe about your pictures?
   (c) If A, B, C are any sets, we may define

   $$A \times B \times C = \big\{(a, b, c) : a \in A,\ b \in B,\ c \in C\big\}$$

   If $C = \{6, 7\}$ and A, B are as above, state the set $A \times B \times C$ in roster notation.

2. Rewrite the condition $(x, y) \in (A^{\mathsf{C}} \cup B) \times (C \setminus D)$ in terms of (some of) the propositions

   $$x \in A, \quad x \notin A, \quad x \in B, \quad x \notin B, \quad y \in C, \quad y \notin C, \quad y \in D, \quad y \notin D$$

3. Consider the intervals $A = [2, 5]$ and $B = (0, 4)$ of real numbers.
   (a) Express the set $(A \setminus B)^{\mathsf{C}}$ in interval notation, as a disjoint union of intervals.
   (b) Draw a picture of the set $(A \setminus B)^{\mathsf{C}} \times (B \setminus A)$.

   (*Be careful: In this problem $B = (0, 4)$ is an **interval** (a subset of $\mathbb{R}$), not a **point** in $\mathbb{R}^2$!*)

4. A *straight line* in $\mathbb{R}^2$ is a subset of the form

   $$A_{a,b,c} = \big\{(x, y) : ax + by = c\big\}, \quad \text{for some constants } a, b, c, \text{ with } a, b \text{ not both zero}$$

   (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.

5. Draw a picture, similar to that in Example 6.4.2, which illustrates the fact that

   $$(A \times B)^{\mathsf{C}} = (A^{\mathsf{C}} \times B^{\mathsf{C}}) \cup (A^{\mathsf{C}} \times B) \cup (A \times B^{\mathsf{C}})$$

   Now give a rigorous proof of the claim.

6. Let $A = [1, 3]$, $B = [2, 4]$ and $C = [2, 3]$. Prove or disprove:

   $$(A \times B) \cap (B \times A) = C \times C$$

   (*Hint: Draw the sets $A \times B$, $B \times A$ and $C \times C$ in the Cartesian plane*)

7. Prove that $A \cap B = \emptyset \iff (A \times B) \cap (B \times A) = \emptyset$.
   (*Hint: try the previous question first*)

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) $A \times (B \setminus C) = (A \times B) \setminus (A \times C)$

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 any sets A, B, C, D, prove that

$$(A \cup C) \times (B \cup D) = (A \times B) \cup (A \times D) \cup (C \times B) \cup (C \times D)$$

10. Prove by induction. For all $n \in \mathbb{N}$, if $A_1, \ldots, A_n$ are finite sets, then

$$|A_1 \times \cdots \times A_n| = |A_1| \cdots |A_n|$$

11. (a) Suppose $|A| = 3$, 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) (Hard) More generally, suppose $|A| = m$, $|B| = n$ and $|A \cap B| = c$. What can you say about the cardinalities $|(A \times B) \cap (B \times A)|$ and $|(A \times B) \cup (B \times A)|$?

12. 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 *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 that there is a unique function $h : Z \to A \times B$ such that $\rho_1 = \pi_1 \circ h$ and $\rho_2 = \pi_2 \circ h$.

13. Let $E \subseteq \mathbb{N} \times \mathbb{N}$ be the smallest subset satisfying 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) (Hard!) Show that $E = \{(a, b) \in \mathbb{N} \times \mathbb{N} : \gcd(a, b) = 1\}$.
(*Hint: what do the generating rules have to do with the Euclidean algorithm?*)

14. (Hard) A strict set-theoretic definition requires that ordered pair (a, b) be defined as a set for instance $(a, b) := \{a, \{a, b\}\}$. We *prove* that $(a, b) = (c, d) \iff a = c$ and $b = d$.

(a) The *regularity axiom* of set theory says there is no set a for which $a \in a$. Use this to prove that the cardinality of $\{a, \{a, b\}\}$ is two.

(b) Prove that $(a, b) = (c, d) \implies (a = c \text{ and } b = d)$ or $(a = \{c, d\} \text{ and } c = \{a, b\})$

(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 $b = d$.

6.2 Power Sets

Thus far we have used the operations of subset, complement, union, intersection and Cartesian product to build new sets from old. There is essentially only one further method available.

Definition 6.5. Let A be a set. Its *power set* $\mathcal{P}(A)$ is the set of all subsets of A:

$$\mathcal{P}(A) = \{B : B \subseteq A\}$$

Otherwise said: $B \in \mathcal{P}(A) \iff B \subseteq A$.

Examples 6.6. 1. The set $A = \{1,3,7\}$ has the following subsets:

0-element subsets: $\emptyset$
1-element subsets: $\{1\}$, $\{3\}$, $\{7\}$
2-element subsets: $\{1,3\}$, $\{1,7\}$, $\{3,7\}$
3-element subsets: $\{1,3,7\}$

Gathering these together yields the power set:

$$\mathcal{P}(A) = \{\emptyset, \{1\}, \{3\}, \{7\}, \{1,3\}, \{1,7\}, \{3,7\}, \{1,3,7\}\}$$

The power set therefore has *eight* elements. Be absolutely certain you understand the difference between $\in$ and $\subseteq$. Here are eight propositions; which are true and which false?[37]

(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) $\{1\} \subseteq \mathcal{P}(A)$

2. The set $B = \left\{1, \{\{2\},3\}\right\}$ has precisely *two* elements, namely 1 and $\{\{2\},3\}$. As before, we gather the subsets of B in a table:

0-element subsets: $\emptyset$
1-element subsets: $\{1\}$, $\left\{\{\{2\},3\}\right\}$
2-element subsets: $\left\{1, \{\{2\},3\}\right\}$

Remember that to make a subset out of a single element you surround the element with braces:

$$1 \in B \implies \{1\} \subseteq B \implies \{1\} \in \mathcal{P}(B)$$
$$\{\{2\},3\} \in B \implies \left\{\{\{2\},3\}\right\} \subseteq B \implies \left\{\{\{2\},3\}\right\} \in \mathcal{P}(B)$$

Using different-sized braces is essential here! The power set of B has *four* elements:

$$\mathcal{P}(B) = \left\{\emptyset, \{1\}, \left\{\{\{2\},3\}\right\}, \left\{1, \{\{2\},3\}\right\}\right\}$$

As a further exercise in making careful use of notation, we prove a simple theorem.

[37] Only (a), (d), and (g) are true. Make sure you understand why!

Theorem 6.7. *If $A \subseteq B$, then $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.*

Proof. Suppose $A \subseteq B$ and let $C \in \mathcal{P}(A)$. We must show that $C \in \mathcal{P}(B)$.

$$\begin{aligned} C \in \mathcal{P}(A) &\implies C \subseteq A && \text{(definition of power set)} \\ &\implies C \subseteq B && \text{(since } C \subseteq A \text{ and } A \subseteq B\text{)} \\ &\implies C \in \mathcal{P}(B) && \text{(definition of power set)} \end{aligned}$$

We conclude that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$. ∎

The converse to this is also true: $\mathcal{P}(A) \subseteq \mathcal{P}(B) \implies A \subseteq B$. Try proving it yourself.

Cardinality and Power Sets

Let's investigate the relationship between the cardinality of a set and its power set. 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've seen this pattern before: we are looking at the first few lines of Pascal's Triangle!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 a general result.

Theorem 6.8. *Let A be a finite set. Then $|\mathcal{P}(A)| = 2^{|A|}$.*

Conjuring a proof may seem daunting given how little we know about A; only its *cardinality*. By introducing a variable n for the cardinality and rephrasing the theorem

$$\forall n \in \mathbb{N}_0,\ |A| = n \implies |\mathcal{P}(A)| = 2^n$$

induction seems like a sensible approach. But what might the induction step look like? The basic idea is to view a set with $n+1$ elements as the disjoint union of a set with n elements and a single-element set. It is instructive to see an example of the strategy before writing the proof.

Example 6.9. Let $B = \{1, 2, 3\}$. Delete $3 \in B$ to create a smaller set

$$A = B \setminus \{3\} = \{1, 2\}$$

In the table, the subsets of $Y \subseteq B$ are split into two groups depending on whether $3 \in Y$. Each subset $Y \subseteq B$ either has the form X or $X \cup \{3\}$ where $X \subseteq A$.

Plainly B has twice the number of subsets of A; two for each subset $X \subseteq A$.

Subsets of B	
X	$X \cup \{3\}$
$\emptyset$	$\{3\}$
$\{1\}$	$\{1,3\}$
$\{2\}$	$\{2,3\}$
$\{1,2\}$	$\{1,2,3\}$

This method of pairing is exactly mirrored in the induction step of our formal proof.

Proof. We prove by induction on the cardinality of A. For each $n \in \mathbb{N}_0$, consider the proposition

$$\text{Every set } A \text{ with cardinality } n \text{ has } |\mathcal{P}(A)| = 2^n \tag{$*$}$$

Base case ($n = 0$): If $n = 0$, then $A = \emptyset$ (Lemma 4.10). But then $\mathcal{P}(A) = \{\emptyset\} \implies |\mathcal{P}(A)| = 1 = 2^0$.

Induction step: Fix $n \in \mathbb{N}_0$ and assume $(*)$ is true for this n. Let B be *any* set with $n+1$ elements. Choose an element $b \in B$ and define $A = B \setminus \{b\}$. The subsets of B may be separated into two types:

1. Subsets $X \subseteq B$ which do not contain b.
2. Subsets $Y \subseteq B$ which contain b.

In the first case, X is 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 there are 2^n subsets $X \subseteq A$, whence

$$|\mathcal{P}(B)| = 2\,|\mathcal{P}(A)| = 2^{n+1}$$

By induction, $(*)$ holds for all $n \in \mathbb{N}_0$. ∎

Exercise 7 gives an alternative proof.

Example 6.10. You might erroneously expect the sets $\mathcal{P}(A \times B)$ and $\mathcal{P}(A) \times \mathcal{P}(B)$ to be the same. Here is a simple counter-example to convince you otherwise!
Let $A = \{a\}$ and $B = \{b, c\}$. Think about cardinalities:

$$|\mathcal{P}(A \times B)| = 2^{|A \times B|} = 2^{|A||B|} = 2^2 = 4$$
$$|\mathcal{P}(A) \times \mathcal{P}(B)| = |\mathcal{P}(A)|\,|\mathcal{P}(B)| = 2^{|A|}2^{|B|} = 2^{|A|+|B|} = 2^3 = 8$$

Since the cardinalities are different, the sets cannot be equal: $\mathcal{P}(A \times B) \neq \mathcal{P}(A) \times \mathcal{P}(B)$. But what about *subset*? Might the smaller set be a subset of the larger? Again the answer is no, as can be seen by computing the sets explicitly.

$$A \times B = \{(a,b),(a,c)\} \implies \mathcal{P}(A \times B) = \{\emptyset,\ \{(a,b)\},\ \{(a,c)\},\ \{(a,b),(a,c)\}\}$$

The elements of $\mathcal{P}(A \times B)$ are *sets of ordered pairs.* By contrast, the elements of $\mathcal{P}(A) \times \mathcal{P}(B)$ are *ordered pairs of sets*:

$$\begin{aligned}\mathcal{P}(A) \times \mathcal{P}(B) &= \{\emptyset, \{a\}\} \times \{\emptyset, \{b\}, \{c\}, \{b,c\}\} \\ &= \Big\{(\emptyset,\emptyset), (\emptyset,\{b\}), (\emptyset,\{c\}), (\emptyset,\{b,c\}), (\{a\},\emptyset), (\{a\},\{b\}), (\{a\},\{c\}), (\{a\},\{b,c\})\Big\}\end{aligned}$$

The elements of the two sets have completely different types, so there is no way that one could be a subset of the other!

Exercises 6.2. A reading quiz and several questions with linked video solutions can be found online.

1. For each A, find $\mathcal{P}(A)$ and $|\mathcal{P}(A)|$.

   (a) $A = \{1,2\}$ (b) $A = \{1,2,3\}$ (c) $A = \big\{(1,2),(2,3)\big\}$

   (d) $A = \big\{\emptyset, 1, \{a\}\big\}$ (e) $A = \Big\{\big\{1,2\big\}, 3, \big\{4,\{5\}\big\}\Big\}$ (f) $A = \Big\{(1,2),\ 3,\ \big(4,\{5\}\big)\Big\}$

2. Let $A = \{1,3\}$ and $B = \{2,4\}$.

   (a) Draw a picture of the set $A \times B$ as a subset of $\mathbb{R}^2$.

   (b) Compute $\mathcal{P}(A \times B)$.

   (c) What is the cardinality of $\mathcal{P}(A) \times \mathcal{P}(B)$?

3. Determine whether the following statements are true or false. Justify your answers.

   (a) If $\{7\} \in \mathcal{P}(A)$, then $\{7\} \notin A$.

   (b) Suppose that $A \subsetneq \mathcal{P}(B) \subsetneq C$ where $|A| = 2$. Then $|C|$ can be 5, but $|C|$ cannot be 4.

   (c) If $|B| = |A| + 1$, then $\mathcal{P}(B)$ has at least two more elements than $\mathcal{P}(A)$.

   (d) If A, B, C, D are cardinality-two subsets of $\{1,2,3\}$. Then at least two of them are equal.

4. Here are three incorrect proofs of Theorem 6.7: $A \subseteq B \implies \mathcal{P}(A) \subseteq \mathcal{P}(B)$. Why does each fail?

   (a) Let $x \in \mathcal{P}(A)$. 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

   $$\mathcal{P}(A) = \big\{\emptyset, \{1\}, \{2\}, \{1,2\}\big\} \subseteq \big\{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, B\big\} = \mathcal{P}(B)$$

   (c) Let $\{x\} \in \mathcal{P}(A)$. Then $x \in A$. Since $A \subseteq B$, we have $x \in B$. But then $\{x\} \in \mathcal{P}(B)$.

5. (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. (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 $A \cap \mathcal{P}(A) \neq \emptyset$?

7. If you've studied combinatorics, you'll know that the binomial coefficient $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ denotes the number of distinct ways one can choose r objects from a set of n objects.

   (a) Prove directly: $\binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}$ whenever $1 \leq r \leq n$.

   (b) (Hard) Prove by induction that $\forall n \in \mathbb{N}_0$, $\sum_{r=0}^{n} \binom{n}{r} = 2^n$, by using part (a) in the induction step.

   (c) Explain why part (b) provides an alternative proof of Theorem 6.8.

   If you found this easy, try proving the binomial theorem: $\forall n \in \mathbb{N}_0$, $(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}$.

6.3 Indexed Collections of Sets: Union and Intersection Revisited

In Definition 4.11 we defined the union of *two* sets $A \cup B = \{x : x \in A \text{ or } x \in B\}$, which inductively extends to any *finite union* of sets

$$A_1 \cup \cdots \cup A_n = \{x : x \in A_k \text{ for some } k\}$$

In this section we consider a stronger definition and compute unions and intersections of (potentially) infinite collections of sets.

Definition 6.11. Given a set of sets $\{A_n\}$ (each A_n is a set!), we form its *union* and *intersection*:

$$\bigcup A_n = \{x : x \in A_n \textit{ for some } n\} \qquad x \in \bigcup A_n \iff \exists n \text{ such that } x \in A_n$$
$$\bigcap A_n = \{x : x \in A_n \textit{ for all } n\} \qquad x \in \bigcap A_n \iff \forall n \text{ we have } x \in A_n$$

We say that $\{A_n\}$ is *pairwise disjoint* if $A_m \cap A_n = \emptyset$ whenever $m \neq n$.

For all examples in this section, the sets A_n are *indexed*: each n lies in some indexing set, typically $\mathbb{N}, \mathbb{Z}$ or $\mathbb{R}$. It is typical to decorate the union/intersection symbols to indicate this: e.g., if $n \in \mathbb{N}$ we might use the notation $\bigcup_{n\in\mathbb{N}} A_n$ or $\bigcup_{n=1}^{\infty} A_n$.

Example 6.12. Here is a simple (finite) example to get us used to the notation. Let

$$A_1 = \{1,3,5\}, \quad A_2 = \{2,3,4,6\}, \quad A_3 = \{1,2,3,6\}$$

The indexed collection $\{A_n : n \in I\}$ is indexed by $I = \{1,2,3\}$. Then

$$\bigcup_{n=1}^{3} A_n = A_1 \cup A_2 \cup A_3 = \{1,2,3,4,5,6\} \qquad \bigcap_{n=1}^{3} A_n = A_1 \cap A_2 \cap A_3 = \{3\}$$

Lemma 6.13. *Let $\{A_n\}$ be a set of sets. For any $A_m \in \{A_n\}$,*

$$A_m \subseteq \bigcup A_n \quad \textit{and} \quad \bigcap A_n \subseteq A_m$$

A proof is almost immediate from the definition: can you supply it?

Nested Collections

When a collection of sets is indexed by the natural numbers $\mathbb{N}$ in such a way that successive sets satisfy a subset relation, we describe the collection as *nested,* for instance

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$$

Since $A_n \subseteq A_1$ for all n, we see that $\bigcup_{n=1}^{\infty} A_n = A_1$:

$$x \in \bigcup_{n=1}^{\infty} A_n \iff \exists n \in \mathbb{N}, x \in A_n \iff x \in A_1$$

Computing the *intersection* in such a situation typically requires much more care...

Example 6.14. Consider the nested collection $\{A_n : n \in \mathbb{N}\}$ of half-open intervals $A_n = \left[0, \frac{1}{n}\right)$:

$$m \leq n \implies \tfrac{1}{n} \leq \tfrac{1}{m} \implies A_n \subseteq A_m \implies A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$$

The union is therefore $\bigcup_{n=1}^{\infty} [0, \frac{1}{n}) = A_1 = [0, 1)$.

Before considering the full intersection, we first compute all *finite* intersections. The nesting condition says that a finite intersection is simply the smallest of the listed sets: for any constant $m \in \mathbb{N}$,

$$\bigcap_{n=1}^{m} \left[0, \tfrac{1}{n}\right) = A_1 \cap \cdots \cap A_m = A_m = \left[0, \tfrac{1}{m}\right)$$

To find the *infinite* intersection, it is very tempting to take limits:

$$\bigcap_{n=1}^{\infty} \left[0, \tfrac{1}{m}\right) \stackrel{?}{=} \lim_{m\to\infty} \bigcap_{n=1}^{m} \left[0, \tfrac{1}{m}\right) \stackrel{?}{=} \left[0, \lim_{m\to\infty} \tfrac{1}{m}\right) = [0, 0)$$

This is mathematical garbage! Nothing you know about limits justifies either questionable equality. Moreover, the 'answer' $[0, 0)$ could only mean the *empty set*, which is *incorrect*: we claim that

$$\bigcap_{n=1}^{\infty} \left[0, \tfrac{1}{n}\right) = \{0\}$$

Proof. First observe that

$$x \in \bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \left[0, \tfrac{1}{n}\right) \iff \forall n \in \mathbb{N},\ 0 \leq x < \tfrac{1}{n}$$

Certainly $x = 0$ satisfies these inequalities. It remains to eliminate the other possibilities.

- If $x < 0$, then $x \notin A_1 = [0, 1)$ and so $x \notin \bigcap A_n$.
- Suppose that $x > 0$. There exists[38] N large enough so that $\frac{1}{N} \leq x$. Otherwise said, $x \notin A_N$, whence $x \notin \bigcap A_n$.

■

If the last part of the argument seems difficult, try an example! If $x = 0.13$, observe that

$$0.1 < 0.13 \implies x \notin A_{10} \implies x \notin \bigcap_{n=1}^{\infty} A_n$$

By modifying the endpoints of the sets A_n we obtain slightly different results:

$$\bigcap_{n=1}^{\infty} \left(0, \tfrac{1}{n}\right) = \varnothing \qquad \bigcap_{n=1}^{\infty} \left(0, \tfrac{1}{n}\right] = \varnothing \qquad \bigcap_{n=1}^{\infty} \left[0, \tfrac{1}{n}\right] = \{0\}$$

How would the arguments differ from what we did above?

The moral is that you cannot naïvely apply limits to sequences of sets. If thinking about limits helps your intuition, great, but you can't trust it blindly!

[38] The existence of $N \geq \frac{1}{x}$ should be intuitive; it is in fact guaranteed by the Archimedean property (Exercise 2.4.12).

An indexed collection can also be nested the other way round, in which case the intersection is straightforward (though unions need more work)

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \implies \bigcap_{n=1}^{\infty} A_n = A_1$$

Examples 6.15. Here are a few more simple examples of computing unions and intersections of indexed collections; some are nested, some not.

1. Let $A_n = [n, n+1) \subseteq \mathbb{R}$, for each $n \in \mathbb{Z}$. For example $A_3 = [3,4)$ and $A_{-17} = [-17,-16)$. Every real number lies in precisely one such set (the sets A_n are pairwise disjoint), whence

   $$\bigcup_{n\in\mathbb{Z}} A_n = \mathbb{R} \quad \text{and} \quad \bigcap_{n\in\mathbb{Z}} A_n = \emptyset$$

   To prove the former, note that $x \in [n, n+1)$ where $n = \lfloor x \rfloor$ is the greatest integer less than or equal to x: i.e., $\forall x \in \mathbb{R}$, we have $x \in A_{\lfloor x \rfloor}$, whence $\mathbb{R} \subseteq \bigcup_{n\in\mathbb{Z}} A_n$ (the reversed subset inclusion is trivial since each $A_n \subseteq \mathbb{R}$).

2. For each $n \in \mathbb{N}$, let $A_n = [-n, n]$ be a closed interval. This is a nested collection

   $$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \implies \bigcap_{n=1}^{\infty} A_n = A_1 = [-1,1]$$

   Similarly to the previous example, $\forall x \in \mathbb{R}$ we have $x \in A_n$ where n is any integer $\geq |x|$: we conclude that $\bigcup_{n=1}^{\infty} A_n = \mathbb{R}$.

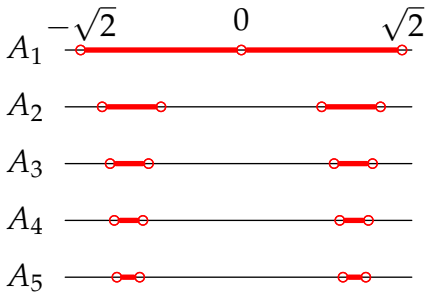
3. For each $n \in \mathbb{N}$, let $A_n = \{x \in \mathbb{R} : \left|x^2 - 1\right| < \frac{1}{n}\}$. Before computing the union and intersection of these sets, it is helpful to write each set as a pair of intervals. Note that

   $$\left|x^2-1\right| < \frac{1}{n} \iff -\frac{1}{n} < x^2 - 1 < \frac{1}{n} \iff \sqrt{1-\frac{1}{n}} < |x| < \sqrt{1+\frac{1}{n}}$$

   The sets are nested: $A_1 \supseteq A_2 \supseteq A_3 \supseteq A_4 \supseteq \cdots$, where

   $$A_n = \left(-\sqrt{1+\tfrac{1}{n}}, -\sqrt{1-\tfrac{1}{n}}\right) \cup \left(\sqrt{1-\tfrac{1}{n}}, \sqrt{1+\tfrac{1}{n}}\right)$$

   $$\implies \bigcup_{n=1}^{\infty} A_n = A_1 = (-\sqrt{2}, 0) \cup (0, \sqrt{2})$$

   For the intersection,

   $$\forall n \in \mathbb{N},\ x \in A_n \iff \forall n \in \mathbb{N},\ \left|x^2-1\right| < \frac{1}{n} \iff x^2 - 1 = 0$$

   It follows that $\bigcap_{n=1}^{\infty} A_n = \{1, -1\}$.

4. Let $A_0 = \{0,1\}$, $A_n = \{1\}$ is $n \geq 1$ is odd, and $A_n = \{2\}$ if $n \geq 2$ is even. Then,

   $$\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \left(\bigcup_{k=1}^{\infty} A_k\right) \cap \left(\bigcup_{k=2}^{\infty} A_k\right) \cap \cdots = \{0,1,2\} \cap \{0,1\} \cap \{0,1\} \cap \cdots = \{0,1\}$$

   Think about why $x \in \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \iff x$ lies in infinitely many of the sets A_n.

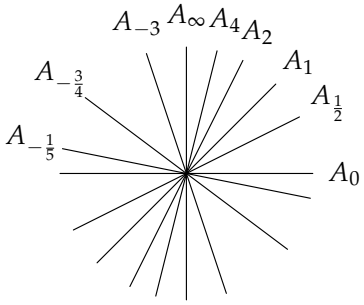
Unions: Don't Confuse Sets and Elements

When working with large unions, it is easy to confuse two sets:

- $\{A_n\}$ is a set of sets, each element of which is some set A_n.
- $\bigcup A_n$ is the set whose elements lie in at least one A_n.

It is important to understand the difference! Sometimes the indexed collection itself is the object of interest, other times we may want to use its union or intersection to define something new.

Example 6.16 (Projective space). Let A_m be the line[39] through the origin in $\mathbb{R}^2$ with gradient $m \in \mathbb{R} \cup \{\infty\}$. Since every point in $\mathbb{R}^2$ lies on such a line, and distinct lines intersect at the origin, we see that



$$\bigcup A_m = \mathbb{R}^2 \quad \text{and} \quad \bigcap A_m = \{(0,0)\}$$

The indexed collection is known as *projective space*

$$\mathbb{P}(\mathbb{R}^2) = \{A_m : m \in \mathbb{R} \cup \{\infty\}\}$$

and is interesting in its own right. Each *element* of projective space is a *line*, making $\mathbb{P}(\mathbb{R}^2)$ a very different set to $\mathbb{R}^2$. This example also shows that indexing sets don't have to be simple sets of integers!

In the next example we use an infinite union to define an interesting set.

Example 6.17 (Finite Decimals). For each $n \in \mathbb{N}$, let A_n be the set of decimals of length n,

$$A_n = \{0.a_1a_2\ldots a_n : \text{where each } a_i \in \{0,1,\ldots,9\}\}$$

For example $0.134 \in A_3$. Since $0.134 = 0.1340$, we also have $0.134 \in A_4$. This is a nested collection

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq A_4 \subseteq \cdots \implies \bigcap_{n\in\mathbb{N}} A_n = A_1 = \{0, 0.1, \ldots, 0.9\}$$

Now consider unions. If $m \in \mathbb{N}$, then

$$\bigcup_{n=1}^{m} A_n = A_m = \{x \in [0,1) : x \text{ has a decimal representation of length} \leq m\}$$

You might guess that the infinite union would be the entire[40] interval $[0,1]$, but this is incorrect:

$$\begin{aligned} x \in \bigcup_{n\in\mathbb{N}} A_n &\iff \exists n \in \mathbb{N} \text{ such that } x \in A_n \\ &\iff x \text{ is a decimal with some } \textit{finite} \text{ length } n \end{aligned}$$

The infinite union $\bigcup A_n$ is precisely the set of $x \in [0,1)$ which have a *finite* decimal representation! This is far from the entire interval: many rational numbers are excluded (e.g., $\frac{1}{3} = 0.3333\cdots \notin \bigcup A_n$), and the union contains *no irrational numbers*.

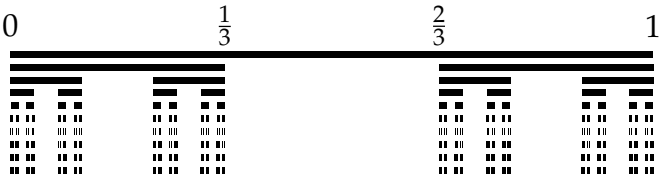
[39] The symbol ∞ is used to indicate the vertical line A_∞ with 'infinite gradient.'

[40] We would include $1 = 0.9999\cdots$

Optional! We finish this section with a bit of fun, using an infinite intersection to create a *fractal* set.

Example 6.18 (Cantor's middle-third set). Starting with the interval $C_0 = [0,1]$, construct a sequence of sets C_n by repeatedly removing the middle third of all intervals in C_n.

$$C_0 = [0,1]$$
$$C_1 = [0,\tfrac{1}{3}] \cup [\tfrac{2}{3},1]$$
$$C_2 = [0,\tfrac{1}{9}] \cup [\tfrac{2}{9},\tfrac{1}{3}] \cup [\tfrac{2}{3},\tfrac{7}{9}] \cup [\tfrac{8}{9},1]$$
$$\vdots$$



The sequence is drawn up to C_9, though you'll have to zoom in a long way to see the detail!

Cantor's middle-third set is defined to be the infinite intersection $\mathcal{C} := \bigcap_{n=0}^{\infty} C_n$.

Cantor's set has several interesting properties which we state non-rigorously.

Zero Measure (length) It should seem reasonable to write

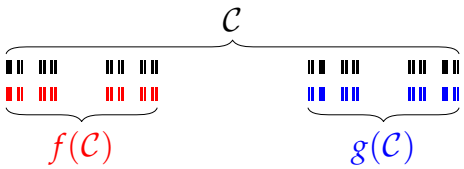
$$\text{len}(C_0) = 1, \quad \text{len}(C_1) = \tfrac{2}{3}, \quad \text{len}(C_2) = \left(\tfrac{2}{3}\right)^2, \quad \ldots \quad , \quad \text{len}(C_n) = \left(\tfrac{2}{3}\right)^n, \quad \ldots$$

where $\text{len}(C_n)$ is the sum of the lengths of all intervals in C_n. Since $\mathcal{C} \subseteq C_n$ for all n, we conclude that $\text{len}(\mathcal{C}) = 0$: Cantor's set contains *no intervals* of positive length.

Infinite Cardinality Cantor's set contains the endpoints of every interval removed at any stage of its construction. In particular, $\frac{1}{3^n} \in \mathcal{C}$ for all $n \in \mathbb{N}_0$, whence $\mathcal{C}$ is an *infinite set*.[41]

Self-similarity Let $f, g : \mathbb{R} \to \mathbb{R}$ be functions $f(x) = \frac{1}{3}x$ and $g(x) = \frac{1}{3}x + \frac{2}{3}$. Since $C_{n+1} = f(C_n) \cup g(C_n)$, we conclude

$$\mathcal{C} = f(\mathcal{C}) \cup g(\mathcal{C})$$



Cantor's set consists of *two shrunken copies of itself*, a classic property of fractals.

We'll analyze Cantor's set a little and consider a related construction in Exercises 14 & 15. Other fractal sets with a similar construction include the *Sierpiński carpet* and the *von Koch snowflake*.

Exercises 6.3. A reading quiz and several questions with linked video solutions can be found online.

1. Let $I = \{1,3,4\}$. Determine $\bigcup_{r\in I} S_r$ and $\bigcap_{r\in I} S_r$, where each $S_r = [r-1, r+3]$ is an interval.
2. For each integer n, consider the set $B_n = \{n\} \times \mathbb{R}$.
   (a) Draw a picture (in the Cartesian plane) of $\bigcup_{n=2}^{4} B_n = B_2 \cup B_3 \cup B_4$.
   (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 $\left(\bigcup_{n=2}^{4} B_n\right) \cap C$.
   (d) Compute $\bigcup_{n=2}^{4} (B_n \cap C)$. Compare with your answer to part (c).

[41] In fact it is more than merely infinite, it is *uncountably* so, as we'll discuss in Chapter 8. The bizarre contrast between this and the zero measure property was part of the reason Cantor introduced his set.

3. Give an example of four subsets A, B, C, D of $\{1,2,3,4\}$ such that all intersections of two subsets are different.

4. For each of collection, define an interval A_n such that the given collection is $\{A_n : n \in \mathbb{N}\}$. Then find both the union and intersection of the collection.

   (a) $\{[1, 2+1), [1, 2+\frac{1}{2}), [1, 2+\frac{1}{3}), \ldots\}$

   (b) $\{(-1, 2), (-\frac{3}{2}, 4), (-\frac{5}{3}, 6), (-\frac{7}{4}, 8), \ldots\}$

   (c) $\{(\frac{1}{4}, 1), (\frac{1}{8}, \frac{1}{2}), (\frac{1}{16}, \frac{1}{4}), (\frac{1}{32}, \frac{1}{8}), (\frac{1}{64}, \frac{1}{16}), \ldots\}$

5. 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. Use Definition 6.11 to prove that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \implies \bigcap_{n\in\mathbb{N}} A_n = A_1$

7. Let A_n be the set of decimals of length n, as described in Example 6.17.

   (a) Prove *directly* that the cardinality of A_n is 10^n.

   (b) Prove *by induction* that $|A_n| = 10^n$.

   (c) Prove that $\bigcup_{n=1}^{\infty} A_n \subseteq \mathbb{Q}$.

   (d) Explain why $\frac{1}{9} \notin \bigcup_{n=1}^{\infty} A_n$.

8. Suppose that $\forall n \in \mathbb{N},\ A_n \neq \emptyset$ and $m \geq n \implies A_m \subseteq A_n$. Prove or disprove the following:

   (a) $\bigcup_{n=1}^{293} A_n \neq \emptyset$ (b) $\bigcap_{n=1}^{293} A_n \neq \emptyset$ (c) $\bigcup_{n\in\mathbb{N}} A_n \neq \emptyset$ (d) $\bigcap_{n\in\mathbb{N}} A_n \neq \emptyset$

9. Compute the set $\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [\frac{1}{k}, 2+\frac{1}{k}]$. For a general family of sets $\{A_n : n \in \mathbb{N}\}$ Explain why it is is reasonable to write

$$x \in \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \iff x \text{ lies in all except finitely many } A_n$$

10. 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) $(\bigcup_{n\in I} A_n) \cap (\bigcup_{n\in I} B_n) \neq \bigcup_{n\in I}(A_n \cap B_n)$

    (b) $(\bigcap_{n\in I} A_n) \cup (\bigcap_{n\in I} B_n) \neq \bigcap_{n\in I}(A_n \cup B_n)$

11. (De Morgan's laws) Let $\{A_n : n \in I\}$ be an indexed family of sets and B a set. Prove

    (a) $B \setminus (\bigcup_{n\in I} A_n) = \bigcap_{n\in I}(B \setminus A_n)$

    (b) $B \setminus (\bigcap_{n\in I} A_n) = \bigcup_{n\in I}(B \setminus A_n)$

12. 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 $I = \emptyset$.

13. (Hard) Let $A_n = \{\frac{m}{n} \in \mathbb{Q} : 0 < m < n, m \in \mathbb{N}\}$, for each $n \in \mathbb{N}$.

    (a) State 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 $\bigcup_{n\in\mathbb{N}} A_n = \mathbb{Q} \cap (0,1)$.

    (d) Argue that further $\bigcup_{n\in\mathbb{N}} A_{2n} = \mathbb{Q} \cap (0,1)$.

    (e) Extend your proof to show that, for any fixed $p \in \mathbb{N}$, $\bigcup_{n\in\mathbb{N}} A_{pn} = \mathbb{Q} \cap (0,1)$.

14. (Hard) A *ternary representation*[42] of a number $x \in [0,1]$ is an expression

$$x = \sum_{n=1}^{\infty} \frac{a_n}{3^n} = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \cdots \text{ where each } a_n \in \{0,1,2\}$$

    We write $x = [0.a_1a_2a_3\cdots]_3$. For example $[0.12]_3 = \frac{1}{3} + \frac{2}{3^2} = \frac{5}{9}$.

    (a) Verify that $[0.02101]_3 = \frac{64}{243}$, $[0.22222\cdots]_3 = 1$ and $[0.020202020\cdots]_3 = \frac{1}{4}$.
    (*Hint: You'll need the geometric series formula* $\sum_{n=1}^{\infty} r^n = \frac{r}{1-r}$ *for the latter two*)

    (b) Let C_n be the n^{th} set in the construction of Cantor's middle-third set $\mathcal{C}$ (Example 6.18). Prove by induction that C_n is the set of all $x \in [0,1]$ with a ternary representation whose first n digits are only 0 or 2.
    (*Hints: Use* $C_{n+1} = f(C_n) \cup g(C_n)$*; What does division by 3 do to a ternary representation?*)

    (c) Argue that $\frac{1}{4} \in \mathcal{C}$, but that it is not an endpoint of any of the deleted middle-thirds removed during the construction of $\mathcal{C}$.

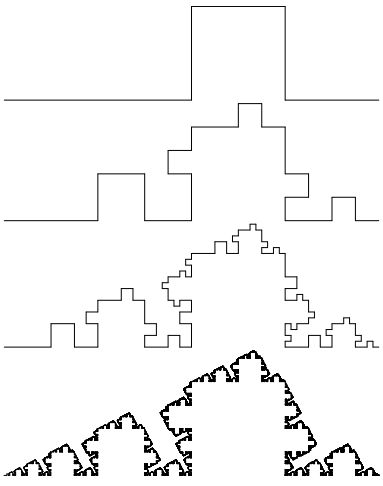
15. (Hard) We construct a modified Cantor set and fractal curve. Starting with $F_0 = [0,1]$, repeatedly delete the *third quarter* segment of each interval to obtain a sequence of sets $F_0, F_1, F_2, \ldots$:

$$F_1 = [0, \tfrac{1}{2}] \cup [\tfrac{3}{4}, 1], \quad F_2 = [0, \tfrac{1}{4}] \cup [\tfrac{3}{8}, \tfrac{1}{2}] \cup [\tfrac{3}{4}, \tfrac{7}{8}] \cup [\tfrac{15}{16}, 1], \quad \ldots$$

    (a) Prove that the sum of the lengths of all of all line segments in F_n is $\left(\frac{3}{4}\right)^n$.

    (b) Prove that $\bigcap_{n=1}^{\infty} F_n$ contains no intervals of positive length.

    (c) Instead of deleting the third quarter of each line segment, replace it with the other three sides of a square. The first three steps and the result of applying the process infinitely many times are shown in the pictures.

    After step 1 the curve has length $\ell_1 = \frac{3}{2}$ and the area under the curve is $A_1 = \frac{1}{4^2}$. After step 2 the length is $\ell_2 = \frac{9}{4}$ and the area $A_2 = A_1 + \frac{1}{4}A_1 + \frac{1}{16}A_1 = \frac{21}{4^4}$.

        i. Find the length ℓ_n of the curve after n steps. What is the 'length' of the resulting fractal curve?

        ii. Repeat for the *area* under each curve F_n. Prove that the area between the fractal and the x-axis is $\frac{1}{8}$.

[42] Analogous to a decimal representation $x = \sum_{n=1}^{\infty} \frac{a_n}{10^n} = \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \cdots$ where $a_n \in \{0,1,2,\ldots,9\}$.

7 Relations and Partitions

The mathematics of sets is rather basic until one has a notion of how to relate elements of sets to each other. We are already familiar with several examples of this, for instance:

1. The usual *order* of numbers (e.g., $3 < 7$) is a way of relating/comparing two elements of $\mathbb{R}$.
2. A *function* $f : A \to B$ relates elements in a set A with those in B.

In this chapter we discuss a general framework based on the Cartesian product (Section 6.1).

7.1 Binary Relations on Sets

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. The *domain* and *range* of $\mathcal{R}$ are the sets

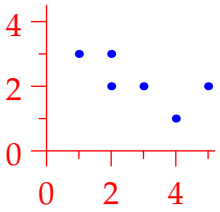
$$\text{dom}(\mathcal{R}) = \{a \in A : \exists b \in B, (a,b) \in \mathcal{R}\} \qquad \text{range}(\mathcal{R}) = \{b \in B : \exists a \in A, (a,b) \in \mathcal{R}\}$$

If $(a,b) \in \mathcal{R}$ we can also write $a \mathcal{R} b$, and say 'a is related to b.' Similarly $a \not\mathcal{R} b$ means $(a,b) \notin \mathcal{R}$.

Examples 7.2. 1. $\mathcal{R} = \{(1,3),(2,2),(2,3),(3,2),(4,1),(5,2)\}$ defines a relation on $\mathbb{N}$. Various true statements about this relation include

$$(2,2) \in \mathcal{R}, \qquad (4,2) \notin \mathcal{R}, \qquad 2 \not\mathcal{R} 5, \qquad 3 \mathcal{R} 2$$

In this case $\text{dom}(\mathcal{R}) = \{1,2,3,4,5\}$ and $\text{range}(\mathcal{R}) = \{1,2,3\}$.

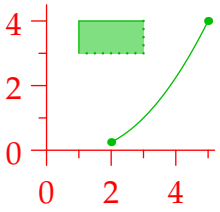


Example 1

2. $\mathcal{R} = \Big([1,3) \times (3,4]\Big) \cup \{(2t+1, t^2) : t \in [\frac{1}{2}, 2]\}$ is a relation on $\mathbb{R}$. This time we have $\text{dom}(\mathcal{R}) = [1,5]$ and $\text{range}(\mathcal{R}) = [\frac{1}{4}, 4]$.

3. For any set A, the set $\mathcal{R} = \{(a,a) : a \in A\}$ is a relation on A. The relation $\mathcal{R}$ is simply 'equals:'

$$(a,b) \in \mathcal{R} \iff a = b$$



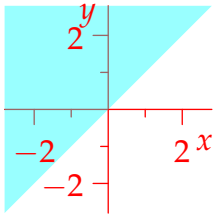
Example 2

4. On $\mathbb{R}$, the relation $\leq$ can be graphed: we plot all points $(x,y) \in \mathbb{R}^2$ such that $x \leq y$.

5. If A is the set of all humans, we may define $\mathcal{R} \subseteq A \times A$ by

$$(a,b) \in \mathcal{R} \iff a, b \text{ have a parent/child or sibling relationship}$$

In this example, the mathematical use of *relation* is similar to that in English: I am related to my sister, and my mother is related to me.



Example 4

6. If A is a set, then $\subseteq$ is a relation on the power set $\mathcal{P}(A)$ (the set of subsets). For instance, if $A = \{1,2,3\}$ then $\{1\} \subseteq \{1,3\}$ but $\{1,3\} \not\subseteq \{1\}$.

7. If $f : A \to B$ is a function then $\{(a, f(a)) : a \in A\}$ is a relation! We'll revisit this in Section 7.2.

With abstract relations, there are only a small number of things we can do.

Definition 7.3. Given a relation $\mathcal{R} \subseteq A \times B$, its *inverse* $\mathcal{R}^{-1} \subseteq B \times A$ is the set

$$\mathcal{R}^{-1} = \{(b,a) \in B \times A : (a,b) \in \mathcal{R}\}$$

To find the elements of $\mathcal{R}^{-1}$, you simply switch the components of each ordered pair in $\mathcal{R}$.
We say that $\mathcal{R}$ is *symmetric* if $\mathcal{R} = \mathcal{R}^{-1}$ (requires $A = B$): i.e., $(x,y) \in \mathcal{R} \iff (y,x) \in \mathcal{R}$.

Theorem 7.4. *For any relations $\mathcal{R}, \mathcal{S} \subseteq A \times B$:*

1. $\text{dom}(\mathcal{R}^{-1}) = \text{range}(\mathcal{R})$
2. $\text{range}(\mathcal{R}^{-1}) = \text{dom}(\mathcal{R})$
3. $(\mathcal{R}^{-1})^{-1} = \mathcal{R}$
2. $\mathcal{R} \subseteq \mathcal{S} \iff \mathcal{R}^{-1} \subseteq \mathcal{S}^{-1}$
5. $(\mathcal{R} \cup \mathcal{S})^{-1} = \mathcal{R}^{-1} \cup \mathcal{S}^{-1}$
4. $(\mathcal{R} \cap \mathcal{S})^{-1} = \mathcal{R}^{-1} \cap \mathcal{S}^{-1}$
7. *If $A = B$, then $\mathcal{R} \cup \mathcal{R}^{-1}$ and $\mathcal{R} \cap \mathcal{R}^{-1}$ are both symmetric*

Proof. Here are two of the arguments. Try the others yourself.

4. Suppose $\mathcal{R} \subseteq \mathcal{S}$. Then,

$$\begin{aligned} (b,a) \in \mathcal{R}^{-1} &\implies (a,b) \in \mathcal{R} && \text{(definition of inverse)} \\ &\implies (a,b) \in \mathcal{S} && \text{(since } \mathcal{R} \subseteq \mathcal{S}\text{)} \\ &\implies (b,a) \in \mathcal{S}^{-1} && \text{(definition of inverse)} \end{aligned}$$

Therefore $\mathcal{R}^{-1} \subseteq \mathcal{S}^{-1}$. Replacing $\mathcal{R}, \mathcal{S}$ with $\mathcal{R}^{-1}, \mathcal{S}^{-1}$ and applying part 1, we also see that

$$\mathcal{R}^{-1} \subseteq \mathcal{S}^{-1} \implies (\mathcal{R}^{-1})^{-1} \subseteq (\mathcal{S}^{-1})^{-1} \implies \mathcal{R} \subseteq \mathcal{S}$$

7. We prove the first half. By part 3,

$$(\mathcal{R} \cup \mathcal{R}^{-1})^{-1} = \mathcal{R}^{-1} \cup (\mathcal{R}^{-1})^{-1} = \mathcal{R}^{-1} \cup \mathcal{R} = \mathcal{R} \cup \mathcal{R}^{-1}$$

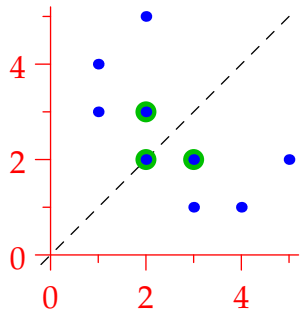
■

Example (7.2.1, cont.). $\mathcal{R} = \{(1,3),(2,2),(2,3),(3,2),(4,1),(5,2)\}$ is not symmetric. Indeed

$$\mathcal{R}^{-1} = \{(3,1),(2,2),(3,2),(2,3),(1,4),(2,5)\} \neq \mathcal{R}$$

as should be plain: e.g., $(1,3) \in \mathcal{R}$ but $(3,1) \notin \mathcal{R}$. However,

$$\mathcal{R} \cap \mathcal{R}^{-1} = \{(2,2),(2,3),(3,2)\} \quad \text{and}$$
$$\mathcal{R} \cup \mathcal{R}^{-1} = \{(1,3),(3,1),(2,2),(2,3),(3,2),(4,1),(1,4),(5,2),(2,5)\}$$



are both symmetric. Observe the symmetry in the picture. One obtains $\mathcal{R}^{-1}$ by *reflecting*[43] $\mathcal{R}$ in the line $y = x$: a *symmetric* relation on $\mathbb{R}$ has *reflection symmetry* in this line.

[43] Just as one does for inverse functions in calculus...

Exercises 7.1. A reading quiz and practice question can be found online.

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. What are its domain and range?
   (b) What is the inverse of $\mathcal{R}$?

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 $\mathcal{S}$ be the relation on $\mathbb{R}$ defined by

   $$a\,\mathcal{S}\,b \iff \exists x \in \mathbb{Q}\setminus\{0\} \text{ such that } a = x^2 b$$

   Is this relation symmetric?

3. Draw pictures of the following relations on the set of real numbers $\mathbb{R}$.

   (a) $\mathcal{R} = \{(x,y) : y \le 2 \text{ and } y \ge x \text{ and } y \ge 2-x\}$
   (b) $\mathcal{S} = \{(x,y) : (x-4)^2 + (y-1)^2 \le 9\}$

   State the domain and range and draw the inverse of each relation.

4. A relation is defined on $\mathbb{N}$ by $a\,\mathcal{R}\,b \Leftrightarrow \frac{a}{b} \in \mathbb{N}$. Let $c, d \in \mathbb{N}$. Under what conditions can we write $c\,\mathcal{R}^{-1}\,d$?

5. 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) Find $\text{dom}(\mathcal{R})$, $\text{range}(\mathcal{R})$ and $\mathcal{R}^{-1}$.
   (b) Compute the relations $\mathcal{R} \cup \mathcal{R}^{-1}$ and $\mathcal{R} \cap \mathcal{R}^{-1}$, and check that they are symmetric.

6. For the relation $\mathcal{R} = \{(x,y) : x \le y\}$ on $\mathbb{N}$, what is $\mathcal{R}^{-1}$, and what is the intersection $\mathcal{R} \cap \mathcal{R}^{-1}$?

7. Let $\mathcal{R} \subseteq \mathbb{Z} \times \mathbb{Z}$ be the relation $m\,\mathcal{R}\,n$ iff $m \mid n$. Compute $\mathcal{R} \cap \mathcal{R}^{-1}$.

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 $\mathcal{R} \cap \mathcal{R}^{-1} = \emptyset$?

9. Give formal proofs of the remaining parts of Theorem 7.4.

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.

11. Let $\mathcal{R}$ be a relation on a set A and define $\mathcal{S} = \mathcal{R} \cup \mathcal{R}^{-1}$. Prove that $\mathcal{S}$ is the *smallest symmetric relation on A containing $\mathcal{R}$* in the following sense: if

    $$\mathrm{T} = \left\{\mathcal{T} \subseteq A \times A : \mathcal{T} \text{ symmetric and } \mathcal{R} \subseteq \mathcal{T}\right\}$$

    then

    $$\mathcal{S} = \bigcap_{\mathcal{T} \in \mathrm{T}} \mathcal{T}$$

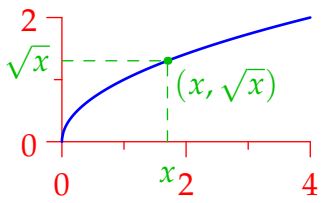
    ($\mathcal{S}$ *is known as the* symmetric closure *of* $\mathcal{R}$)

7.2 Functions revisited

In Section 4.3, we naïvely defined a function $f : A \to B$ as a *rule* associating to each element $a \in A$ an element $f(a) \in B$. But what do we mean by a *rule*? We address this issue by turning Example 7.2.7 on its head: a function $f : A \to B$ is precisely its *graph*.

Example 7.5. The function $f : [0,4] \to [0,2] : x \mapsto \sqrt{x}$ corresponds to the relation

$$\{(x, \sqrt{x}) : x \in [0,4]\} \subseteq [0,4] \times [0,2]$$



The difficulty is that we cannot use the notation $f(a)$ until we know that we have a function...

Definition 7.6. A *function* from A to B is a relation $f \subseteq A \times B$ satisfying two conditions:

1. $\mathrm{dom}(f) = A$ (each $a \in A$ is related to *at least one* $b \in B$)
2. $(a, b_1), (a, b_2) \in f \implies b_1 = b_2$ (each $a \in A$ is related to *at most one* $b \in B$)

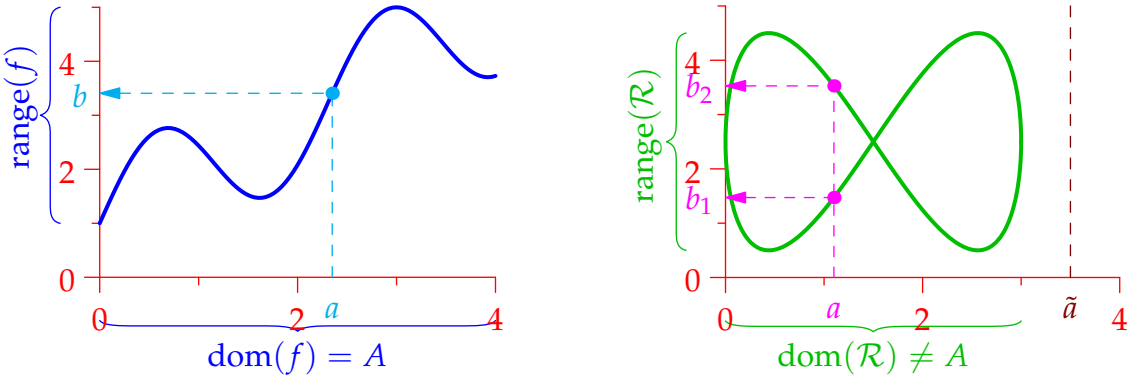
A relation $f \subseteq A \times B$ is a function if and only if *every* $a \in A$ is the first entry of *exactly one* ordered pair $(a, b) \in f$. This is simply an abstraction of the *vertical line test* from calculus.

Example 7.7. Let $A = [0,4]$ and $B = [0,5]$. Two relations $f, \mathcal{R} \subseteq A \times B$ are drawn below.

The first relation defines a function $f : A \to B$ since every $a \in A$ corresponds to exactly one $b \in B$: the vertical line through any $a \in A$ intersects the graph in exactly one point (a, b).

The second relation $\mathcal{R}$ does not define a function. In fact it fails *both* parts of the definition:

1. The vertical line through $\tilde{a} \in A$ does not intersect the graph: $\tilde{a} \notin \mathrm{dom}(\mathcal{R})$.
2. The vertical line through $a \in A$ intersects the graph in *two* points $(a, b_1) \neq (a, b_2)$.



Injectivity (Definition 4.18) can be rephrased in this new context: $f : A \to B$ injective means

$$(\forall a \in A)\ \ (a_1, b), (a_2, b) \in f \implies a_1 = a_2$$

Surjectivity has the same meaning as before: $\mathrm{range}(f) = B$.

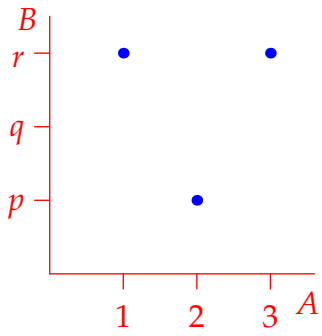
Example 7.8. Let $A = \{1,2,3\}$ and $B = \{p,q,r\}$. The relation

$$f = \{(1,r),(2,p),(3,r)\}$$

satisfies both conditions necessary to be a function. In elementary language, $f(1) = r$, $f(2) = p$ and $f(3) = r$. Moreover f is:

Not injective since $(1,r),(3,r) \in f$ and $1 \neq 3$

Not surjective since $\text{range}(f) = \{p,r\} \neq B$



The function $f : A \to B$

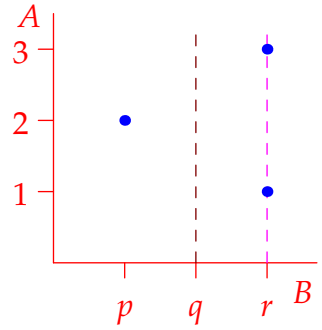
The inverse relation

$$f^{-1} = \{(r,1),(p,2),(r,3)\} \subseteq B \times A$$

is *not* a function due to failing *both* conditions of Definition 7.6.

1. $\text{dom}(f^{-1}) = \{p,r\} \neq B$. The vertical line through $b = q$ does not intersect (the graph of) f^{-1}.
2. $(r,1) \in f^{-1}$ and $(r,3) \in f^{-1}$, but $1 \neq 3$. The vertical line through $b = r$ intersects f^{-1} twice.

Note how the manner of these failures are *identical* to the failures of injectivity and surjectivity…



$f^{-1} \subseteq B \times A$: not a function

The example highlights one advantage of the relational approach: the inverse of a function $f : A \to B$ *always exists*; it is the *inverse relation* $f^{-1} \subseteq B \times A$. The question is whether f^{-1} is also a function. From the example and our discussions in Section 4.3, you should strongly suspect the answer.

Theorem 7.9. *Let $f : A \to B$ be a function and consider its inverse relation $f^{-1} \subseteq B \times A$. Then*

$$f^{-1} \text{ is a function} \iff f \text{ is bijective (both injective and surjective)}$$

Proof. Recalling Definition 7.6, we see that f^{-1} is a function if and only if the two conditions hold:

1. $\text{dom}(f^{-1}) = B$
2. $(b,a_1),(b,a_2) \in f^{-1} \implies a_1 = a_2$

By Theorem 7.4, condition 1 is equivalent to $\text{range}(f) = B$; that is, f is surjective.
Condition 2 is equivalent to $(a_1,b),(a_2,b) \in f \implies a_1 = a_2$: i.e., f is injective. ∎

Example (7.5 cont.). $f = \{(x,\sqrt{x}) : x \in [0,4]\} \subseteq [0,4] \times [0,2]$ is

Injective since $(x_1,y),(x_2,y) \in f \implies x_1 = y^2$ and $x_2 = y^2$, whence $x_1 = x_2$.

Surjective since $\text{range}(f) = \{\sqrt{x} : x \in [0,4]\} = [0,2]$.

Since f is bijective, its inverse relation is a *function*: $f^{-1} = \{(y,y^2) : y \in [0,2]\}$ or, as we'd normally write, $f^{-1} : [0,2] \to [0,4] : y \mapsto y^2$.

Exercises 7.2. A reading quiz and practice questions can be found online.

1. Suppose f is the relation

$$f = \{(1,1),(2,3),(3,5),(4,7)\} \subseteq \{1,2,3,4\} \times \{1,2,3,4,5,6,7\}$$

   (a) Show that we have 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 this function?

   (b) Is f injective? Justify your answer.

   (c) Suppose $g \subseteq \{1,2,3,4\} \times B$ is another relation so that the *graphs* of f and g are identical: that is, *as sets*,

$$\{(a, f(a)) : a \in \{1,2,3,4\}\} = \{(a, g(a)) : a \in \{1,2,3,4\}\}$$

   If g is a bijective function, what is B?

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) $\mathcal{U} = \{(x,y) \in [0,1] \times [0,1] : x^2 + y^2 = 1\}$

3. (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 7.6 to prove that the relation f^{-1} is *not* a function.

   (d) Prove directly from Definition 4.18 that f is not injective and not surjective. Compare your arguments with your answer to part (c).

4. Repeat the previous question for $f : \mathbb{R} \to \mathbb{R} : x \mapsto \sqrt{x^2 - 4x + 5}$.

5. (a) Express the function $f : \mathbb{R} \to \mathbb{R} : x \mapsto x^3$ as a relation on $\mathbb{R}$.

   (b) What is the inverse relation f^{-1}? Is it a function? Justify your answer.

6. Let A be any set and define the relation $\mathrm{id}_A = \{(a,a) : a \in A\}$ on A. Show that id_A is a bijective function (it is called the *identity function* on A). What is its inverse?

7. Consider the relation $f = \{(x,x) : x \in \mathbb{Q}\} \cup \{(x,0) : x \notin \mathbb{Q}\} \subseteq \mathbb{R}^2$

   (a) Prove that f is a function $\mathbb{R} \to \mathbb{R}$.

   (b) Is f injective? Surjective? Explain.

8. Suppose $f \subseteq A \times B$ is a function.

   (a) If $U \subseteq A$, explain why it would be reasonable to write the *image of* U as

$$f(U) = \{b \in B : \exists u \in U \text{ with } (u,b) \in f\}$$

   (*Recall Definition 4.16 if you're unsure what this means*)

   (b) If $V \subseteq B$, how would you write the inverse image $f^{-1}(V)$ in this language?

7.3 Equivalence Relations & Partitions

What do we mean when we say that two objects are *equal*? Mathematicians use the word flexibly, often in reference to non-identical objects which share some common property.[44] You've been doing this for years, indeed our very first result (Theorem 1.1: even + even = even) has exactly this flavor. To help develop a more flexible notion of equality, consider three important concepts.

Example 7.10. Let X be a set. The following are immediate:

Reflexivity: $\forall x \in X,\ x = x$

Symmetry: $(\forall x, y \in X)\ x = y \implies y = x$

Transitivity: $(\forall x, y, z \in X)\ x = y$ and $y = z \implies x = z$

We turn this simple example on its head to create an abstract, generalized notion of equality.

Definition 7.11. We define the following properties for a relation[45] $\sim$ on a set X:

Reflexivity: $\forall x \in X,\ x \sim x$ (every element of X is related to itself)

Symmetry: $x \sim y \implies y \sim x$ (if x is related to y, then y is related to x)

Transitivity: $x \sim y$ and $y \sim z \implies x \sim z$ (if x is related to y and y is related to z, then x is related to z)

An *equivalence relation* on X is a relation $\sim$ satisfying all three properties.

Example 7.10 says that 'equals' is indeed an equivalence relation on any set X. Things would be very boring if this were the only example...

Example 7.12. Define a relation $\sim$ on $\mathbb{Z}$ by

$$x \sim y \iff x - y \text{ is even} \qquad \text{(otherwise said, } x \equiv y \pmod 2\text{)}$$

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: Given $x, y \in \mathbb{Z}$, $x \sim y \implies x - y$ even $\implies y - x$ even $\implies y \sim x$.

Transitivity: Given $x, y, z \in \mathbb{Z}$,

$$\begin{aligned} x \sim y \text{ and } y \sim z &\implies x - y \text{ even and } y - z \text{ even} \\ &\implies x - z = (x - y) + (y - z) \text{ is even} \qquad \text{(Theorem 1.1!)} \\ &\implies x \sim z \end{aligned}$$

As we'll see shortly, an equivalence relation algebraically characterizes what it means for elements to share a common property: here $x \sim y$ if and only if they have the same *parity*.

[44] This goes back at least to Euclid, who used *equal* to refer to congruent triangles. To Euclid, congruent triangles were *essentially identical* and therefore not worth distinguishing.

[45] *Symmetry* is precisely as in Definition 7.3. As we've done, the universal quantifiers for symmetry and transitivity are typically hidden. The symbol $\sim$ ('tilde,' or 'twiddles') is commonly used for an abstract equivalence relation. It is the same symbol used to denote *similar triangles*: congruence and similarity are both equivalence relations on the set of triangles!

It would also be somewhat dull if every relation were an equivalence relation. In fact most are not.

Examples 7.13. 1. Consider the relation $\leq$ on the natural numbers $\mathbb{N}$. We check each condition:

Reflexivity: True. $\forall x \in \mathbb{R},\ x \leq x$.

Symmetry: False. For example, $2 \leq 3$ but $3 \not\leq 2$.

Transitivity: True. $\forall x, y, z \in \mathbb{R}$, if $x \leq y$ and $y \leq z$, then $x \leq z$.

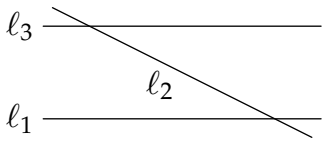
Plainly $\leq$ is *not* an equivalence relation on $\mathbb{N}$.

2. Let X be the set of lines in the plane and define $\ell_1 \mathcal{R} \ell_2 \iff \ell_1, \ell_2$ intersect.

Reflexivity: True. Every line intersects itself, so $\ell \mathcal{R} \ell$ for all $\ell \in X$.

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, if ℓ_1, ℓ_3 are parallel and ℓ_2 crosses both, then $\ell_1 \mathcal{R} \ell_2$, $\ell_2 \mathcal{R} \ell_3$ and $\ell_1 \not\mathcal{R} \ell_3$.



Again $\mathcal{R}$ is not an equivalence relation on X.

3. The relation $\mathcal{R} = \{(1,1), (1,3), (3,1), (3,3)\}$ is not an equivalence relation on $X = \{1,2,3\}$.

Reflexivity: False. For instance $(2,2) \notin \mathcal{R}$ (otherwise said, $2 \not\mathcal{R} 2$).

Symmetry: True. In all four cases, $(x,y) \in \mathcal{S} \implies (y,x) \in \mathcal{R}$ is clear.

Transitivity: True. Check this yourself; there are six valid combinations!

The usefulness of equivalence relations comes when we group together all related elements.

Definition 7.14. Given an equivalence relation $\sim$ on X, the *equivalence class* of $x \in X$ is the set

$$[x] = \{y \in X : y \sim x\} \qquad (y \in [x] \iff y \sim x)$$

If $y \in [x]$, we call y a *representative* of the equivalence class $[x]$. The *quotient* of X by $\sim$ is the set of equivalence classes:

$$X/_\sim = \{[x] : x \in X\} \qquad \text{(read '}X \text{ modulo } \sim\text{')}$$

Examples 7.15. We start by revisiting our first two examples.

1. (Example 7.10) For the equivalence relation $x \sim y \iff x = y$ on a set X, each equivalence class has precisely one element $[x] = \{x\}$; the quotient is the set of singleton subsets,

$$X/_\sim = \{\{x\} : x \in X\}$$

2. (Example 7.12) For the relation $x \sim y \iff x - y$ is even, there are two equivalence classes:

$$[0] = \{\ldots, -4, -2, 0, 2, 4, \ldots\} = 2\mathbb{Z} = \{\text{even integers}\}$$
$$[1] = \{\ldots, -3, -1, 1, 3, 5, \ldots\} = 1 + 2\mathbb{Z} = \{\text{odd integers}\}$$

and the quotient has two elements: $\mathbb{Z}/_\sim = \{[0], [1]\}$.

Note that any even number is a representative of the first class in the even/odd example: for instance

$$4 \in [0] \text{ since } 4 - 0 = 4 \text{ is even.}$$

Indeed $[4] = [0]$, which leads us a simple piece of bookkeeping...

Lemma 7.16. *Suppose $\sim$ is an equivalence relation. Then $x \sim y \iff [x] = [y]$.*

Observe how the proof uses all three defining properties of an equivalence relation.

Proof. ($\Leftarrow$) By reflexivity, $x \in [x]$. Thus $[x] = [y] \implies x \in [y] \implies x \sim y$.

($\Rightarrow$) Suppose $x \sim y$. We prove that $[x] = [y]$ by showing that each side is a subset of the other.

($\subseteq$) Let $z \in [x]$. By definition, $z \sim x$. By transitivity,

$$z \sim x \text{ and } x \sim y \implies z \sim y \implies z \in [y]$$

($\supseteq$) By symmetry, we also have $y \sim x$. Repeating the previous argument yields $[y] \subseteq [x]$. ∎

Example 7.17. Let $X = \{\text{students taking this course}\}$ and defined, $\forall x, y \in X$,

$$x \sim y \iff x \text{ achieves the same letter-grade as } y$$

It should be clear that this is an equivalence relation (try saying out loud what reflexivity, symmetry & transitivity mean here!). There is one equivalence class for each letter-grade awarded, each class being the set of all students who obtain a given grade. If we label the equivalence classes $A^+, A, A^-, B^+, \ldots, F$, where, say, $B = \{\text{students obtaining a B-grade}\}$, then the quotient

$$X/_\sim = \{A^+, A, A^-, B^+, \ldots, F\}$$

is essentially the set of possible letter-grades! Suppose Laura & Jorge both achieve a B^+; both are representatives of this equivalence class and the following propositions are all true:

$$\text{Laura} \in B^+, \quad \text{Jorge} \in B^+, \quad \text{Laura} \sim \text{Jorge}, \quad [\text{Laura}] = [\text{Jorge}] = B^+$$

The example is a special case of a general construction.

Theorem 7.18. *Suppose f is a function with domain X. Then*

$$x \sim y \iff f(x) = f(y)$$

defines an equivalence relation on X. There is one equivalence class $[x] = \{y \in X : f(x) = f(y)\}$ for each value in range(f).

The proof is an exercise. A suitable function in the previous example would be

$$f : X \to \{A^+, A, A^-, B^+, \ldots, F\} \quad \text{where} \quad f(x) = \text{“}x\text{'s letter grade”}$$

This converse to the theorem always holds: if $\sim$ is an equivalence relation then the so-called *canonical map* $f : X \to X/_\sim : x \mapsto [x]$ satisfies $x \sim y \iff f(x) = f(y)$!

Examples 7.19. 1. The function

$$f : \mathbb{Z} \to \{0,1,2,3,4\} : x \mapsto x^2 \pmod 5 \qquad \text{(take the remainder on division by 5)}$$

defines an equivalence relation on $\mathbb{Z}$:

$$x \sim y \iff x^2 \equiv y^2 \pmod 5$$

2. The function $f : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto x^2 + y^2$ defines an equivalence relation[46] on $\mathbb{R}^2$:

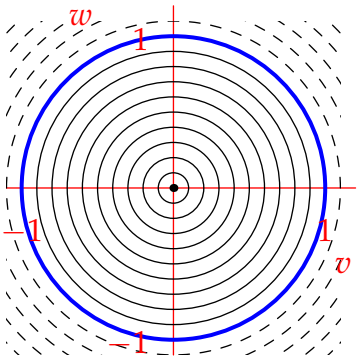
$$(x,y) \sim (v,w) \iff x^2 + y^2 = v^2 + w^2$$

As stated in the Theorem, there is one equivalence classes for each element $r^2 \in \text{range}(f) = \mathbb{R}_0^+$: if $x^2 + y^2 = r^2$, then

$$[(x,y)] = \left\{(v,w) \in \mathbb{R}^2 : v^2 + w^2 = r^2\right\}$$

is the *circle of radius r centered at the origin*. The equivalence class $[(1,0)]$ highlighted in the picture. Moreover, the quotient set is

$$\mathbb{R}^2/_\sim = \{\text{circles centered at the origin}\}$$



Partitions When you cut a cake, each crumb ends up in exactly one slice. To *partition* a set is to do precisely this: split into disjoint subsets so that each element lies in exactly one subset.

Definition 7.20. Let X be a set. A collection $\{A_n\}$ of non-empty subsets $A_n \subseteq X$ *partitions* X if

1. $X = \bigcup A_n$ (each $x \in X$ lies in *at least one* subset A_n)
2. If $A_m \neq A_n$, then $A_m \cap A_n = \emptyset$ (each $x \in X$ lies in *at most one*[47] subset A_n)

Partitions are intimately related to equivalence relations, as the next example illustrates.

Example 7.21. Partition $X = \{1,2,3,4,5\}$ into subsets

$$A_1 = \{1,3\}, \quad A_2 = \{2,4\}, \quad A_3 = \{5\}$$

and consider the relation $\sim$ on X defined by[48]

$$x \sim y \iff x, y \text{ are in the } \textit{same} \text{ subset } A_n$$

Run through the checklist: $\sim$ is reflexive, symmetric & transitive; it is an equivalence relation! Moreover, its equivalence classes are precisely the partitioning subsets A_1, A_2 and A_3:

$$[1] = [3] = \{1,3\} = A_1, \qquad [2] = [4] = \{2,4\} = A_2, \qquad [5] = \{5\} = A_5$$

[46] Be careful: $\sim$ is a subset of $\mathbb{R}^2 \times \mathbb{R}^2$, whereas each *equivalence class* is a subset of $\mathbb{R}^2$!
[47] *Distinct* sets A_n are *pairwise disjoint* (Definition 4.11). It is sometimes useful in examples to permit $A_m = A_n$.
[48] Otherwise said, $\sim$ is the set $\{(1,1),(1,3),(3,1),(3,3),(2,2),(2,4),(4,2),(4,4),(5,5)\} \subseteq X \times X$.

This tight relationship between partitions and equivalence relations is completely general: equivalence relations provide a straightforward *algebraic* method of working with partitions.

Theorem 7.22. 1. *If $\sim$ is an equivalence relation on X, then its equivalence classes partition X.*

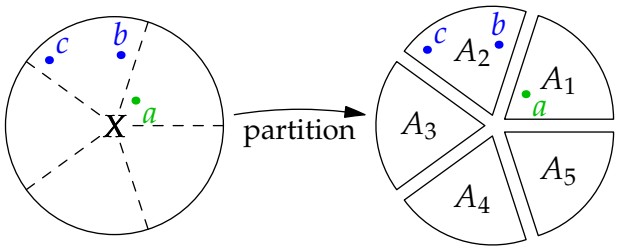
2. *A partition $\{A_n\}$ of X defines an equivalence relation $\sim$ on X by*

$$x \sim y \iff \exists A_n \text{ such that } x, y \in A_n$$

Its equivalence classes are the distinct subsets A_n: otherwise said, ${}^{X}/_{\sim} = \{A_n\}$.

In short, *equivalence relations and partitions are essentially the same thing!*[49] The picture illustrates part 2 and the cake-cutting metaphor:

$$A_1 = [a], \quad A_2 = [b] = [c],$$
$$b \sim c, \quad a \nsim b, \quad a \nsim c$$



Before reading the proof, look back at *every* example of an equivalence relation in this section and convince yourself that the equivalence classes really do partition the 'big set.' In both parts of the proof look for where the assumptions are used—reflexive, symmetric, transitive in part 1; the two partition conditions in part 2—and think about why they are needed.

Proof. 1. Given an equivalence relation $\sim$ on X, we must establish two claims:

(a) Every $x \in X$ lies in some equivalence class. This follows by reflexivity: for each $x \in X$,

$$x \sim x \implies x \in [x]$$

Every element of X lies in the equivalence class defined by itself!

(b) Distinct equivalence classes are pairwise disjoint: $[x] \neq [y] \implies [x] \cap [y] = \emptyset$. We establish the contrapositive. If $[x] \cap [y] \neq \emptyset$, then $\exists z \in [x] \cap [y]$. But then

$$\begin{aligned} z \sim x \text{ and } z \sim y &\implies x \sim z \text{ and } z \sim y && \text{(symmetry)} \\ &\implies x \sim y && \text{(transitivity)} \\ &\implies [x] = [y] && \text{(Lemma 7.16)} \end{aligned}$$

2. We need only demonstrate the reflexivity, symmetry and transitivity of $\sim$.

Reflexivity: $X = \bigcup A_n \implies$ every $x \in X$ lies in some A_n. Thus $x \sim x$ for all $x \in X$.
Symmetry: $x \sim y \implies \exists A_n$ such that $x, y \in A_n \implies y, x \in A_n \implies y \sim x$.
Transitivity: If $x \sim y$ and $y \sim z$, then $\exists A_m, A_n$ such that $x, y \in A_m$ and $y, z \in A_n$. Then

$$\begin{aligned} y \in A_m \cap A_n &\implies A_m \cap A_n \neq \emptyset \implies A_m = A_n \\ &\implies x, z \in A_n \\ &\implies x \sim z \end{aligned}$$

■

[49] Even more is true. If you think carefully, conditions 1 & 2 in the definition of partition (7.20) now form a disguised version of the vertical line test for the canonical map (Theorem 7.18)!

Exercises 7.3. A reading quiz and practice questions can be found online.

1. Show that $x \sim y \iff x \equiv y \pmod 3$ defines an equivalence relation on $\mathbb{Z}$. What are the equivalence classes?

2. (a) Let $\sim$ be the relation on $\mathbb{Z}$ defined by $a \sim b \iff a+b$ is even. Show that $\sim$ is an equivalence relation and determine its equivalence classes.

   (b) If 'even' is replaced by 'odd' in part (a), which of the properties reflexive, symmetric, transitive does $\sim$ now possess?

3. For each 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) $(v,w) \sim (x,y) \iff v^2w = x^2y$

4. Let $X = \{1,2,3,4,5,6\}$. The distinct equivalence classes resulting from an equivalence relation $\sim$ on X are $\{1,4,5\}$, $\{2,6\}$, and $\{3\}$. What is $\sim$? Give your answer as a subset of $X \times X$.

5. $\subseteq$ is a relation on any set of sets. Is $\subseteq$ reflexive, symmetric, transitive? Prove your assertions.

6. Suppose X is a set with at least two elements. Which of the properties reflexive, symmetric, transitive are satisfied by the relation $\neq$?

7. Let $X = \{ax^3 + bx^2 + cx + d : a,b,c,d \in \mathbb{R}\}$ be the set of polynomials of degree ≤ 3. Define a relation $\mathcal{R}$ on X by

   $$p \mathcal{R} q \iff p \text{ and } q \text{ have a common real root } \big(\exists x \in \mathbb{R} : p(x) = q(x)\big)$$

   For instance $p(x) = (x-1)^2$ and $q(x) = x^2 - 1$ have the root 1 in common, so $p \mathcal{R} q$. Determine which of the properties reflexive, symmetric and transitive are possessed by $\mathcal{R}$.

8. Let $A = \{2^m : m \in \mathbb{Z}\}$. A relation $\sim$ is defined on the set $\mathbb{Q}^+$ of positive rational numbers by

   $$a \sim b \iff ab^{-1} \in A$$

   Show that $\sim$ is an equivalence relation and describe the elements in the equivalence class $[3]$.

9. A relation is defined on the set $X = \{a + b\sqrt{2} : a,b \in \mathbb{Q},\ a + b\sqrt{2} \neq 0\}$ by $x \sim y \iff \frac{x}{y} \in \mathbb{Q}$. Show that $\sim$ is an equivalence relation and determine the distinct equivalence classes.

10. For the purposes of this question, a real number is *x small* if $|x| \leq 1$. Let $\mathcal{R}$ be the relation on the set of real numbers defined by $x \mathcal{R} y \iff x - y$ is small.

    Prove or disprove: $\mathcal{R}$ is an equivalence relation on $\mathbb{R}$.

11. Find the equivalence classes for the relation $\sim$ in Example 7.19.1.

12. For each relation $\mathcal{R}$ on $\mathbb{Z}$, decide whether it is reflexive, symmetric, or transitive, and whether it 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 $a \equiv b \pmod 4$

13. For Example 7.13.3, compute the 'classes' $[x] = \{y \in X : x \mathcal{R} y\}$. What do you observe?

14. Let $X = \{1,2,3\}$. Define the relation $\mathcal{S} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,3)\}$ on X.

    (a) Which of the properties reflexive, symmetric, transitive are satisfied by $\mathcal{S}$?
    (b) Let $A_n = \{x \in X : x \,\mathcal{S}\, n\}$. Show that $\{A_1, A_2, A_3\}$ do not partition X.
    (c) Repeat parts (a) and (b) for the relation $\mathcal{T}$ on X, where

    $$\mathcal{T} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,3)\}$$

    (*Warning! Some of the sets A_1, A_2, A_3 might be the same in these examples*)

15. (Example 7.19.2) Prove directly that the circles $A_r = \{(x,y) : x^2 + y^2 = r^2\}$ partition $\mathbb{R}^2$: i.e.,

    $$\mathbb{R}^2 = \bigcup_{r \geq 0} A_r \quad \text{and} \quad A_{r_1} \cap A_{r_2} \neq \emptyset \implies A_{r_1} = A_{r_2}$$

16. Determine whether each collection $\{A_n : n \in \mathbb{R}\}$ partitions $\mathbb{R}^2$. 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) $A_n = \{(x,y) \in \mathbb{R}^2 : y^4 - y^2 = x - n\}$

17. Let X be the set of all humans. If $x \in X$, 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*?

18. A relation $\mathcal{R}$ is *antisymmetric* if $\big((x,y) \in \mathcal{R}\big) \wedge \big((y,x) \in \mathcal{R}\big) \implies x = y$. Give examples of relations $\mathcal{R}$ on $X = \{1,2,3\}$ 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.

19. Let $\mathcal{R}$ be a relation on X and define its *reflexive closure* $\mathcal{S} = \mathcal{R} \cup \{(x,x) : x \in X\}$. Prove that $\mathcal{S}$ is reflexive and that, if $\mathcal{T}$ is any reflexive relation for which $\mathcal{R} \subseteq \mathcal{T}$, then $\mathcal{S} \subseteq \mathcal{T}$.

20. (a) Prove Theorem 7.18.
    (b) If f has domain X, explain why $\{f^{-1}(\{b\}) : b \in \text{range}(f)\}$ forms a partition of X.

21. Define a relation $\sim$ on $\mathbb{R}^2 \setminus \{(0,0)\}$ by

    $$(x,y) \sim (v,w) \iff \exists \lambda \neq 0 \text{ such that } (\lambda x, \lambda y) = (v,w)$$

    (a) Prove that $\sim$ is an equivalence relation.
    (b) What does this relation have to do with projective space $\mathbb{P}(\mathbb{R}^2)$ (Example 6.16)?

22. (*If you've studied linear algebra*) We say that square matrices A, B are *similar* if there exists a matrix M such that $B = MAM^{-1}$.

    (a) Prove that similarity is an equivalence relation on the set of $n \times n$ matrices.
    (b) What is the equivalence class of the identity matrix?
    (c) Show that $\left(\begin{smallmatrix} -11 & 15 \\ -5 & 9 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 4 & 10 \\ 0 & -6 \end{smallmatrix}\right)$ are similar.
    (*Hint: diagonalize the matrices!*)

7.4 Well-definition, Rings and Congruence

We return to our discussion of congruence (Section 3.1) 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 7.23. *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 merely a generalization of Example 7.12 and Exercise 7.3.1. You should prove this yourself. The equivalence classes are precisely the integers which are congruent modulo n:

$$\begin{aligned}[a] &= \{x \in \mathbb{Z} : x \equiv a \pmod{n}\} \\ &= \{x \in \mathbb{Z} : x \text{ and } a \text{ have the same remainder as when divided by } n\} \\ &= \{x \in \mathbb{Z} : x - a \text{ is divisible by } n\}\end{aligned}$$

In this language, we can restate what it means for two equivalence classes to be equal.

Lemma 7.24. $[a] = [b] \iff a \equiv b \pmod{n} \iff \exists k \in \mathbb{Z}$ *such that* $b = a + kn$.

If the meaning of *any* of the above is unclear, re-read the previous section! The equivalence classes of $\sim_n$ partition the integers into exactly n subsets, one for each remainder. The quotient set is therefore

$$\mathbb{Z}/_{\sim_n} = \{[0], [1], \ldots, [n-1]\}$$

We use this set to define an extremely important object.

Definition 7.25. Define operations $+_n$ and $\cdot_n$ on the quotient 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 $\cdot_n$.

It is important to appreciate that $+_n$ and $\cdot_n$ are *new operations*, defining addition/multiplication of *equivalance classes* in terms of standard addition/multiplication of integers.

Example 7.26. We work in the ring $\mathbb{Z}_8$. According to the definition,

$$[3] +_8 [6] = [3 + 6] = [9] = [1]$$

There is a potential difficulty: each equivalence class is *large* (e.g., $[3] = \{\ldots, -5, 3, 11, 19, \ldots\}$), so we have lots of *choice* regarding representatives. For instance, since $[3] = [11]$ and $[6] = [22]$ we should be able to observe that

$$[11] +_8 [22] = [3] +_8 [6]$$

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 $+_8$, things turn out exactly as we'd like,

$$[11] +_8 [22] = [11 + 22] = [33] = [1] = [3] +_8 [6]$$

Consider things more abstractly. Given equivalence classes X and Y, here is the process for computing the equivalence class $X +_n Y$:

1. *Choose representatives* $x \in X$ and $y \in Y$ so that $X = [x]$ and $Y = [y]$.
2. *Sum the integers* x and y to get a *new integer* $x + y \in \mathbb{Z}$.
3. *Take the equivalence class* $X +_n Y = [x + y]$.

The concern is that there are *infinitely many possible choices* for elements $x \in X$ and $y \in Y$ in step 1. If $+_n$ is to make sense, we must obtain the *same* equivalence class $[x + y]$ **regardless of our choices of** x **and** y**.** This indeed happens, as we indicate with a special piece of terminology.

Theorem 7.27. *The operations* $+_n$ *and* $\cdot_n$ *are well-defined.*

This is nothing more than a rehash of Theorem 3.9: compare it with what follows until you are comfortable we are doing the same thing! Observe that

$$[x] = \{z \in \mathbb{Z} : z \equiv x \pmod{n}\} = \{x + kn : k \in \mathbb{Z}\}$$

Otherwise said, all representatives (all possible choices in step 1) of the equivalence class $[x]$ have the form $x + kn$ for some $k \in \mathbb{Z}$. Thinking similarly for y, the Theorem requires only that we prove

$$\forall k, l \in \mathbb{Z}, \quad [x + kn] +_n [y + ln] = [x] +_n [y] \quad \text{and} \quad [x + kn] \cdot_n [y + ln] = [x] \cdot_n [y]$$

Proof. We prove that $+_n$ is well-defined.

$$\begin{aligned} [x + kn] +_n [y + ln] &= [(x + kn) + (y + ln)] && \text{(definition of } +_n) \\ &= [x + y + (k + l)n] = [x + y] && \text{(Lemma 7.24)} \\ &= [x] +_n [y] && \text{(definition of } +_n) \end{aligned}$$

The argument for $\cdot_n$ is similar. ∎

Compare our equivalence class notation with that from Section 3.1. For instance

$$[-3] +_8 [12] = [1] \quad \text{means the same thing as} \quad -3 + 12 \equiv 1 \pmod{8} \qquad (*)$$

Aside: Advanced Notation Given the applicability of $\mathbb{Z}_n$, it is customary in higher-level mathematics to drop the square brackets and all subscripts, simply writing

$$\mathbb{Z}_n = \{0, 1, 2, \ldots, n - 1\}$$

In this language, $(*)$ would simply be written $-3 + 12 = 1$ in $\mathbb{Z}_8$. It is critical to appreciate here that $-3, 12$ and 1 denote *sets* (equivalence classes), not *integers.* Until you are 100% certain of this, you should stick to either of the notations in $(*)$.

It is something of a mathematical joke to write $1 + 1 = 0$ (in $\mathbb{Z}_2$). Even in this context, $1 + 1$ still equals 2; the issue is that $2 = 0$ in $\mathbb{Z}_2$, hence the seemingly paradoxical statement. This is really a very simple claim about *sets*: that the sum of any two odd numbers is even!

Functions and Partitions Our construction of $\mathbb{Z}_n$ is a special case of a more general situation.[50]

Definition 7.28. Let $\sim$ be an equivalence relation on a set X and let A be any set. Suppose we construct $f : {}^{X}/_{\sim} \to A$ using a rule of the form

$$f([x]) := \text{“do something to a representative } x\text{”}$$

We say that f is *well-defined* if this construction defines a legitimate function. More formally:

$$[x] = [y] \implies f([x]) = f([y])$$

Examples 7.29. 1. Consider $\mathbb{Z}_3 = {}^{\mathbb{Z}}/_{\sim}$, that is $X = \mathbb{Z}$ where $x \sim y \iff x \equiv y \pmod 3$. The rule $f([x]) = 2x + 1$ does *not* produce a well-defined function $f : \mathbb{Z}_3 \to \mathbb{Z}$ since, for instance

$$[0] = [9] \quad \text{but} \quad f([0]) = 1 \neq 19 = f([9])$$

2. The rule $f([x]) = [2x + 1]$ does produce a well-defined function $f : \mathbb{Z}_3 \to \mathbb{Z}_3$ however:

$$\begin{aligned} [x] = [y] &\implies y = x + 3k \text{ for some integer } k \\ &\implies f([y]) = f([x + 3k]) = [2x + 6k + 1] = [2x + 1] = f([x]) \end{aligned}$$

since $6x \equiv 0 \pmod 3$. If this seems abstract, note that f is easily stated in tabular form since there are only three possible inputs!

$[x]$	$[0]$	$[1]$	$[2]$
$f([x])$	$[1]$	$[0]$	$[2]$

3. This last example is something of an advert for the advanced notation in the previous aside. We ask for which integers k the rule $f_k(x) = kx$ produces a well-defined function $f_k : \mathbb{Z}_4 \to \mathbb{Z}_6$. If this is confusing, re-rewrite using either of the notations

$$f_k(x) = kx \pmod 6 \quad \text{or} \quad f_k([x]_4) = [kx]_6$$

Start with a special case to get a feel for things. A table of values for $f_1(x)$ shows an immediate problem!

x	0	1	2	3	4	5	6	$\cdots$
$f_1(x)$	0	1	2	3	4	5	0	$\cdots$

In $\mathbb{Z}_4$, we have $0 = 4$, however $f_1(0) = 0$ and $f_1(4) = 4$ *which are not equal in* $\mathbb{Z}_6$! It follows that f_1 is *not well-defined*: it isn't a function (and never was!).

Now proceed systematically.

$$x \equiv y \pmod 4 \implies x - y = 4n \implies kx - ky = 4kn \qquad \text{(for some } n \in \mathbb{Z}\text{)}$$

It follows that f_k is well-defined ($f_k(x) = f_k(y) \in \mathbb{Z}_6$) 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,

$$f_k \text{ is well-defined} \iff 6 \mid 4k \iff 3 \mid 2k \iff 3 \mid k$$

Since $kx \in \mathbb{Z}_6$, equivalent values of k modulo 6 won't change f_k: there are only *two* well-defined functions $f_0(x) = 0$ and $f_3(x) = 3x$.

x	0	1	2	3	4	5	6	$\cdots$
$f_0(x)$	0	0	0	0	0	0	0	$\cdots$
$f_3(x)$	0	3	0	3	0	3	0	$\cdots$

You'll have much more practice with problems like these when you study group theory.

[50] Theorem 7.27 verifies the well-definition of two functions $+_n, \cdot_n : {}^{\mathbb{Z}}/_{\sim_n} \times {}^{\mathbb{Z}}/_{\sim_n} \to {}^{\mathbb{Z}}/_{\sim_n}$

Just for Fun: Geometric Applications (optional!)

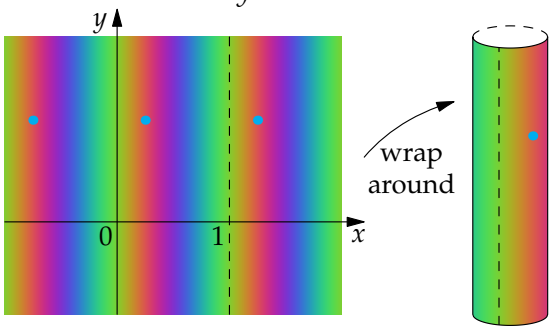
We consider how equivalence relations permit the easy construction of certain geometric objects and of functions on such.

Examples 7.30. 1. Define an equivalence relation on $\mathbb{R}^2$ by

$$(a,b) \sim (c,d) \iff a-c \in \mathbb{Z} \text{ and } b = d$$

The equivalence classes are horizontal strings of points with the same y co-ordinate:

$$[(a,b)] = \{(a+n,b) : n \in \mathbb{Z}\}$$



The set of equivalence classes $\mathbb{R}^2/_\sim$ may be visualized as a *cylinder* by imagining rolling the plane into a tube of circumference 1 so that all points in a given equivalence class coincide.[51]

Now consider the rule $f([(x,y])) = y\sin(2\pi x)$.

$$\begin{aligned}[(x,y)] = [(v,w)] &\implies y = w \text{ and } x = v+n, \text{ for some } n \in \mathbb{Z} \\ &\implies f([(x,y)]) = y\sin(2\pi x) = w\sin(2\pi v) = f([(v,w)])\end{aligned}$$

Otherwise said, $f : \mathbb{R}^2/_\sim \to \mathbb{R}$ is a well-defined function whose *domain is the cylinder.* More generally, and $F : \mathbb{R}^2 \to \mathbb{R}$ for which $(x,y) \sim (v,w) \implies F(x,y) = F(v,w)$ may be viewed as a function on the cylinder.

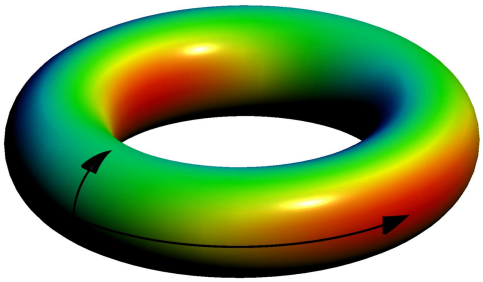
2. As a natural extension, see if you can visualize why the equivalence classes of the relation

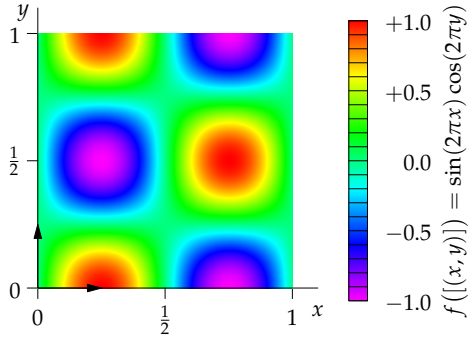
$$(a,b) \sim (c,d) \iff a-c \in \mathbb{Z} \text{ and } b-d \in \mathbb{Z}$$

on $\mathbb{R}^2$ may be identified with a torus (the surface of a ring-doughnut). The function

$$f([x,y]) = \sin(2\pi x)\cos(2\pi y)$$

is well-defined and may thought of as having domain the torus. The pictures below illustrate f, where the colors correspond.





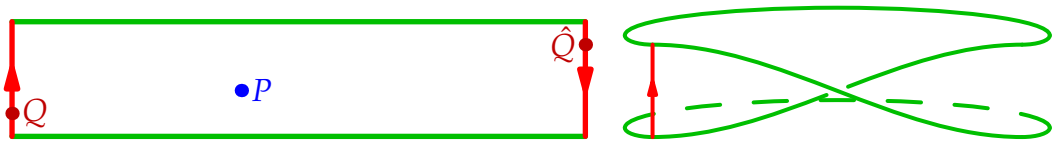
[51] Alternatively, imagine piercing a roll of toilet paper and unrolling it: the single puncture becomes a row of (almost!) equally spaced holes. Unfortunately for the analogy, toilet paper has purposeful thickness!

Equivalence relations and partitions are regularly used in this manner to describe objects in geometry and topology. Here is a final famous example written using the language of partitions.

Example 7.31 (Möbius Strip). Partition the a rectangle X as follows:

- If $P \in X$ does not lie on the left or right edges, place it in a subset $\{P\}$ by itself.
- Otherwise, pair Q with a point on the other edge identified in the opposite direction $\{Q, \hat{Q}\}$.

These subsets clearly partition the rectangle X and thus describe an equivalence relation $\sim$ on X. The quotient set $X/\sim$ may be visualized via the classic construction of a Möbius strip: give the rectangle a half-twist and glue the two edges so that both points in each equivalance class coincide. This construction allows us to analyse the strip while working on the rectangle: if you walk off the right side of the rectangle (at $\hat{Q}$) you simply end up at the corresponding point (Q) on the left side!



Rectangle Half-twist and glue

Exercises 7.4. A reading quiz and practice questions can be found online.

1. Let X be the set of students in a (large) math class and define an equivalence relation on X via

   $$x \sim y \iff x, y \text{ either both wear glasses, or both do not wear glasses}$$

   Does the rule $f([x]) =$ "x's name" describe a well-defined function? Explain.

2. (a) Explicitly check that $[7] + [21] = [98] + [-5]$ in $\mathbb{Z}_{13}$.
   (b) Suppose that $[5] \cdot [7] = [8] \cdot [9]$ makes sense in the ring $\mathbb{Z}_n$. Find n.

3. Prove the second half of Theorem 7.27, that $\cdot_n$ is well-defined.

4. Suppose that p is prime and that in $\mathbb{Z}_p$, we have $[a] \neq [0]$. Show $[a]^2 \neq [0]$.
   (*Hint: Recall Exercise 3.2.13*)

5. Give an explicit proof of Theorem 7.23.

6. Determine whether the following are well-defined.
   (a) $f : \mathbb{Z}_3 \to \mathbb{Z}_5 : [x]_3 \mapsto [x^3]_5$
   (b) $g : \mathbb{Z}_{10} \to \mathbb{Z}_{20} : [x]_{10} \mapsto [x^2]_{20}$
   (c) $h : \mathbb{Z}_8 \to \mathbb{Z}_{16}$ where $h(x) = 2x$ (equivalently $h(x) = 2x \pmod{16}$, or $h([x]_8) = [2x]_{16}$).
   (d) $j : \mathbb{Z}_8 \to \mathbb{Z}_{12}$ where $j(x) = 2x$.

7. (a) Suppose $\forall n \in \mathbb{Z}$, $(x + 4n)^2 \equiv x^2 \pmod{m}$. Find all integers $m \geq 2$ for which this is true.
   (b) 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.

8. In the sense of Example 7.30, can we view $F(x, y) = (y^2 - 1)\sin^2(\pi x)$ as a function whose domain is the cylinder? Explain.

9. (a) Suppose $f : {}^{X}\!/_{\sim} \to A$ is well-defined. State what it means for f to be *injective.* What do you notice?
   (b) Prove that $f : \mathbb{Z}_7 \to \mathbb{Z}_{35} : [x]_7 \mapsto [15x]_{35}$ is a well-defined, injective function.
   (c) Repeat part (b) for $g : \mathbb{Z}_{100} \to \mathbb{Z}_{300} : [x]_{100} \mapsto [9x]_{300}$.
   (*Hint: You may find it useful that* $9 \cdot (-11) \equiv 1 \pmod{100}$)

10. Define a partition of the sphere $S^2 = \{(x,y,z) : x^2 + y^2 + z^2 = 1\}$ into subsets consisting of pairs of antipodal points: $\{(x,y,z),(-x,-y,-z)\}$
    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 Roman Surface; look it up!

11. Consider the relation $(a,b) \sim (c,d) \iff ad = bc$ on $\mathbb{Z} \times \mathbb{N}$.
    (a) Prove that $\sim$ is an equivalence relation.
    (b) List several elements of the equivalence class $[(2,3)]$. Repeat for $[(-3,7)]$. What does $\sim$ have to do with the set of rational numbers $\mathbb{Q}$?
    (c) Define operations $+$ and $\times$ on ${}^{\mathbb{Z}\times\mathbb{N}}\!/_{\sim}$ by
    $$[(a,b)] + [(c,d)] = [(ad+bc, bd)] \qquad [(a,b)] \times [(c,d)] = [(ac,bd)]$$
    Prove that $+$ and $\times$ are well-defined (*do this without using division!*).
    (d) (Hard) Prove that $f([(x,y)]) = \frac{x}{y}$ is a *well-defined bijection* $f : {}^{\mathbb{Z}\times\mathbb{N}}\!/_{\sim} \to \mathbb{Q}$, and that f transforms $+$ and $\times$ into the usual addition and multiplication on $\mathbb{Q}$: that is,
    $$f([(a,b)] + [(c,d)]) = f([(a,b)]) + f([(c,d)]), \text{ etc.}$$
    (*This is essentially a formal definition of the ring of rational numbers!*)

12. 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 an example of a surjective 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 $f : {}^{\mathbb{R}}\!/_{\sim} \to [0,1)$.

13. Suppose $\sim$ is an equivalence relation on X and let $\gamma(x) = [x]$. Prove the following:
    (a) If $f : {}^{X}\!/_{\sim} \to A$ is well-defined, then $F = f \circ \gamma : X \to A$ satisfies $x \sim y \implies F(x) = F(y)$.
    (b) 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 $F = f \circ \gamma$.

14. (Just for fun!) Prove that you can cut a Möbius strip round the middle yet still end up with a single loop.
    Look up the description of a Klein bottle: how is the relationship between it and the Möbius strip similar to the torus/cylinder relationship?

8 Cardinality and Infinite Sets

During the late 1800's, Georg Cantor assaulted the foundations of mathematics by introducing certain infinite sets, in particular his middle-third set (Example 6.18). Cantor ideas about infinity met significant resistance: some mathematicians and philosophers felt his approach to be unnatural; he even inflamed religious scholars who considered his investigations an affront to the divine!

Despite initial antipathy, Cantor's notion of cardinality is now universally accepted. By unearthing several contradictions inherent in contemporary (naïve) set theory, mathematicians became convinced that a rigorous *axiomatic* approach was necessary. Developing an axiomatic foundation to mathematics became a core goal of early 20$^{\text{th}}$ century mathematics.

8.1 Cantor's Notion of Cardinality

Recall that the *cardinality* $|A|$ of a finite set A is simply the number of elements of A (Definition 4.8). While "number of elements" is meaningless for infinite sets, cardinality also gives rise to a *relation* $\leq$ which can be used to *compare* sets; this interpretation does extend to infinite sets...

Example 8.1. Consider sets

$$A = \{\text{fish}, \text{dog}\} \qquad B = \{\alpha, \beta, \gamma\}$$

While the elements of A and B are completely different, the sets themselves may be compared using cardinality: we write $|A| \leq |B|$ to indicate that A has no more elements than B. In Theorem 4.25, we saw that this mandates the existence of an *injective function* $f : A \to B$. It should be clear that

$$f(\text{fish}) = \alpha, \qquad f(\text{dog}) = \beta$$

defines a suitable function.

Theorem 4.25 tells us how to compare cardinalities of finite sets *using functions* and *without counting elements*. Cantor's seemingly innocuous idea was to turn this *theorem* for finite sets into a general *definition* of cardinality.

Definition 8.2. The *cardinality* of a set A is denoted $|A|$. We compare cardinalities as follows:

- $|A| \leq |B| \iff \exists f : A \to B$ injective
- $|A| = |B| \iff \exists g : A \to B$ bijective

$|A| < |B|$ means $|A| \leq |B|$ and $|A| \neq |B|$, that is $\exists f : A \to B$ injective but $\nexists g : A \to B$ bijective.

Lemma 8.3. *On any collection of sets, $A \sim B \iff |A| = |B|$ defines an equivalence relation.*

Cardinality is an abstract property whereby sets can be *compared* rather than a *value* attaching to a given set.[52] Regardless, it should be clear that cardinality partitions any collection of sets: every set has a cardinality, and no set has more than one cardinality. It is moreover natural to identify the cardinalities of finite sets with the *cardinal numbers* $0, 1, 2, 3, 4, \ldots$

[52] We *could* use the Lemma to define $|A| := [A]$ as the equivalence class of A, though this likely feels confusing!

Countably Infinite Sets

To make progress, it is helpful to introduce a symbol for the cardinality of the simplest infinite set.

Definition 8.4. We say that A is a *countably infinite*[53] set and write $|A| = \aleph_0$ (read *aleph-nought* or *aleph-null*), if its cardinality equals that of the set of *natural numbers* $\mathbb{N}$:

$$|A| = \aleph_0 \iff \exists g : \mathbb{N} \to A \text{ bijective}$$

In the next section we'll see why a new symbol is needed; why ∞ doesn't suffice.

Examples 8.5.

1. The function $g : \mathbb{N} \to \mathbb{N}_{\geq 2}$ defined by $g(n) = n + 1$ is a bijection, whence $\mathbb{N}_{\geq 2} = \{2, 3, 4, 5, \ldots\}$ is countably infinite.[54]
2. Let $2\mathbb{N} = \{2, 4, 6, 8, 10, \ldots\}$ be the set of positive even integers. The function

   $$h : \mathbb{N} \to 2\mathbb{N} : n \mapsto 2n$$

   is a bijection. It follows that $|2\mathbb{N}| = |\mathbb{N}| = \aleph_0$ and that $2\mathbb{N}$ is countably infinite.

Theses examples demonstrate a perplexing property of infinite sets. For instance, $2\mathbb{N}$ is a *proper subset* in bijective correspondence with $\mathbb{N}$! It feels like we want to say two contradictory things:

- $\mathbb{N}$ has the 'same number of elements' as $2\mathbb{N}$ (bijectivity of h).
- $\mathbb{N}$ has 'twice the number of elements' as $2\mathbb{N}$ ('half' of $\mathbb{N}$ is even, and 'half' odd).

The source of our discomfort is that *number of elements* is meaningless for infinite sets. However, if we replace this phrase with *cardinality*, then both statements are true![55] The existence of a proper subset with the same cardinality can indeed be used as a *definition* of infinite set (Exercise 13).

Rather than hunting for an explicit bijection, the simplest approach to these problems is often to list the elements of a set A so that it 'looks like' the natural numbers, with an initial element a_1 followed by all others in sequence:

$$A = \{a_1, a_2, a_3, a_4, \ldots\} \quad \text{indicates a bijection } g : \mathbb{N} \to A : n \mapsto a_n, \text{ whence } |A| = \aleph_0$$

For instance, we could have approached our examples by listing elements in tabular form:

$\mathbb{N}$	n	1	2	3	4	5	6	7	8	9	10	$\cdots$
$\mathbb{N}_{\geq 2}$	$g(n)$	2	3	4	5	6	7	8	9	10	11	$\cdots$
$2\mathbb{N}$	$h(n)$	2	4	6	8	10	12	14	16	18	20	$\cdots$

The required bijections are immediately visible with little need to state them explicitly. We apply this technique to two important examples of countably infinite sets.

[53] Some authors say *denumerable* and use *countable* for sets with $|A| \leq \aleph_0$. *Aleph* is the first letter of the Hebrew alphabet.

[54] This is a version of *Hilbert's Grand Hotel Problem.* Imagine a hotel with an infinite number of rooms: Room 1, Room 2, Room 3, Room 4, etc. Even if the hotel is full, by moving everyone one room down the hall, space can always be found for an additional guest. The second example is another version, where infinitely many new guests must be accommodated.

[55] We won't pursue cardinal arithmetic in any detail, but it is completely legitimate to write $2\aleph_0 = \aleph_0$ for the second example. The first example would be $1 + \aleph_0 = \aleph_0$.

List the set of *integers* in a non-standard order, alternating positive and negative terms:

$$\mathbb{Z} = \{z_1, z_2, z_3, z_4, \ldots\} = \{0, 1, -1, 2, -2, 3, -3, 4, -4, \ldots\}$$

The function $g : \mathbb{N} \to \mathbb{Z} : n \mapsto z_n$ is bijective, so we conclude:

Theorem 8.6. *The integers $\mathbb{Z}$ are a countably infinite set.*

You might feel that our argument was too quick! Is it really obvious what g is? Bijectivity is the observation that every integer appears *exactly once* in our non-standard ordering. If you *really* want, you can construct an explicit formula for g from a tabular representation

$$\begin{array}{c|cccccccccc} n & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \cdots \\ \hline g(n) & 0 & 1 & -1 & 2 & -2 & 3 & -3 & 4 & -4 & \cdots \end{array} \implies g(n) = \begin{cases} \frac{1}{2}n & \text{if } n \text{ even} \\ -\frac{1}{2}(n-1) & \text{if } n \text{ odd} \end{cases}$$

and prove bijectivity using the formula. In practice, it is not worth the cost to be this explicit.

As you build up examples, you no longer have to compare directly to the natural numbers. Since composition of bijective functions is bijective (Theorem 4.22/transitivity in Lemma 8.3),

$$B \text{ is countably infinite} \iff \exists A \text{ countably infinite and } \exists g : A \to B \text{ bijective}$$

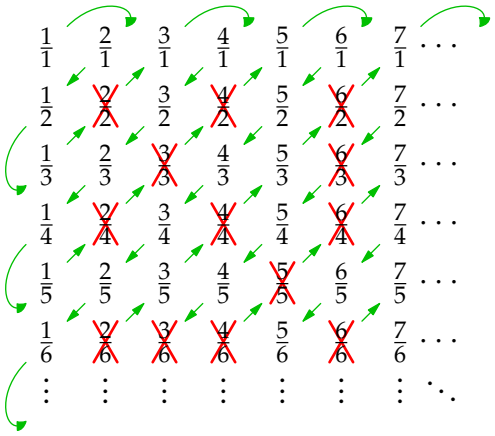
For instance, the set of even integers $2\mathbb{Z}$ is denumerable via the bijection $g : \mathbb{Z} \to 2\mathbb{Z} : z \mapsto 2z$.

We use this approach to help prove the first of Cantor's truly counter-intuitive revelations.

Theorem 8.7. *The rational numbers $\mathbb{Q}$ are countably infinite.*

Any sensible person should feel that there are far, far more rational numbers than integers, yet the sets have the same cardinality. Bizarre!

Proof. For each pair of natural numbers a, b, place the fraction $\frac{a}{b}$ in the b^{th} row, a^{th} column of an infinite square. List the positive rational numbers by tracing diagonals in a snake-like manner and deleting any number that has already been traced ($\frac{2}{2} = \frac{1}{1}$, $\frac{6}{4} = \frac{3}{2}$, etc.).



Since $\frac{a}{b}$ each appears in diagonal $a + b - 1$ and repeats are deleted, *every positive rational number appears exactly once.* We therefore obtain an enumeration

$$\mathbb{Q}^+ = \left\{ \frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{1}{3}, \frac{3}{1}, \frac{4}{1}, \frac{3}{2}, \frac{2}{3}, \frac{1}{4}, \frac{1}{5}, \cdots \right\}$$

and conclude that $\mathbb{Q}^+$ is countably infinite. Plainly this corresponds to a bijection $g : \mathbb{N} \to \mathbb{Q}^+$. To finish the proof, extend g by defining the bijective function

$$h : \mathbb{Z} \to \mathbb{Q} : n \mapsto \begin{cases} g(n) & \text{if } n > 0 \\ 0 & \text{if } n = 0 \\ -g(-n) & \text{if } n < 0 \end{cases}$$

which identifies the negative rationals with $\mathbb{Z}^-$. By Theorem 8.6, we deduce that $|\mathbb{Q}| = |\mathbb{Z}| = \aleph_0$. ∎

Other countably infinite sets appear to be even larger than $\mathbb{Q}$! For example:

- Cartesian products such as $\mathbb{N} \times \mathbb{N}$.
- The *algebraic numbers* $\mathbb{A} = \{x \in \mathbb{C} : p(x) = 0 \text{ for some polynomial } p \text{ with integer coefficients}\}$. Every rational number is algebraic ($\frac{a}{b}$ is a root of $p(x) = bx - a$), but so are many irrationals ($\sqrt{2}$ is a root of $p(x) = x^2 - 2$). Non-algebraic numbers (e.g., π and e) are termed *transcendental.*

Arguments for these examples are left to the exercises.

The Least Infinite Cardinal?

We introduced $\aleph_0$ to represent the cardinality of the 'simplest' infinite set $\mathbb{N}$. Here is a compelling reason why the natural numbers should indeed be considered the *most simple* infinite set.

Theorem 8.8. *A is finite 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 therefore be considered the least infinite cardinal.

Proof. ($\Rightarrow$) Express the set in roster notation $A = \{a_1, \ldots, a_n\}$. We must prove two things:[56]

($|A| \leq \aleph_0$) Define $f : A \to \mathbb{N}$ 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.

($|A| \neq \aleph_0$) We show there are no bijections $\mathbb{N} \to A$. Suppose $g : \mathbb{N} \to A$ and consider the set

$$g(\{1, \ldots, n+1\}) = \{g(1), \ldots, g(n+1)\} \subseteq A$$

Since A has n elements, at least two of the values $g(1), \ldots, g(n+1)$ must be equal. Therefore g is not injective and consequently not bijective.

($\Leftarrow$) See Exercise 11. ∎

We'll address the existence of infinite sets with cardinality *larger* than $\aleph_0$ in the next section.

Exercises 8.1. A reading quiz and practice question can be found online.

1. Refresh your proof skills by proving explicitly that the following functions are bijections:
   (a) $g : \mathbb{N} \to \mathbb{N}_{\geq 2} : n \mapsto n + 1$ (b) $h : \mathbb{N} \to 2\mathbb{N} : n \mapsto 2n$
2. Prove that $\mathbb{Z}_{\geq -3}$ is countably infinite by constructing a bijective function $g : \mathbb{N} \to \mathbb{Z}_{\geq -3}$.
3. Prove that the set $3\mathbb{Z} + 2 = \{3n + 2 : n \in \mathbb{Z}\}$ is countably infinite.
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 providing an explicit bijection with a known countably infinite set.
5. State a bijection $g : (0, 1) \to (4, 6)$ which shows that these intervals have the same cardinality.

[56] The $n = 0$ case ($A = \emptyset$) works, though it feels strange: $f = \emptyset$ is a suitable injective (!) function $f : \emptyset \to \mathbb{N}$, and there are no functions $g \subseteq \mathbb{N} \to \emptyset$. It will help to think about the formal definition of *function* (Section 7.2)!

6. Prove Lemma 8.3 (*revisit Theorem 4.22 on the composition of bijective functions.*)

7. Prove that $A \subseteq B \implies |A| \leq |B|$ (*show there exists an injective function $f : A \to B$*).

8. (a) Prove that $\mathbb{N} \times \mathbb{N}$ is countably infinite by modifying the proof of Theorem 8.7.
   (b) Combine part (a) with Theorem 8.7 to prove that $\mathbb{Q} \times \mathbb{Q}$ is countably infinite.
   (c) Suppose A_n is countably infinite for each $n \in \mathbb{N}$ and list elements as follows:

   $$A_1 = \{a_{11}, a_{12}, a_{13}, a_{14}, \ldots\}$$
   $$A_2 = \{a_{21}, a_{22}, a_{23}, a_{24}, \ldots\}$$
   $$A_3 = \{a_{31}, a_{32}, a_{33}, a_{34}, \ldots\}, \text{ etc.}$$

   Prove that $\bigcup A_n$ is countably infinite (*a countable union of countable sets is countable*).

Warning! The remaining exercises are significantly trickier.

9. Suppose $A \neq \emptyset$. Prove that $|A| \leq |B|$ if and only if there exists a surjective function $g : B \to A$.
   (*Hint: Use g to construct an injective $f : A \to B$, and vice versa*)

10. Let $\mathbb{A} = \{x \in \mathbb{C} : p(x) = 0 \text{ for some polynomial } p \text{ with integer coefficients}\}$ be the set of *algebraic* numbers. We prove that $\mathbb{A}$ is countably infinite.
    (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 $P_M = \{a_d x^d + \cdots + a_1 x + a_0 : d + |a_0| + \cdots + |a_d| = M\}$. Explain why P_M is finite.
    (c) Show that $R_M = \{x \in \mathbb{C} : p(x) = 0 \text{ for some } p \in P_M\}$ is finite by using the fact that a degree d polynomial has at most d roots.
    (d) Prove that $\mathbb{A} = \bigcup_{M \in \mathbb{N}} R_M$ and conclude that $\mathbb{A}$ is countably infinite.

11. We complete the proof of Theorem 8.8: if $|A| < \aleph_0$, then A is a finite set.
    We prove by contradiction. Suppose A is infinite and that $|A| < \aleph_0$. Then there exists an injective function $f : A \to \mathbb{N}$. List the range of f in *increasing* order:

    $$\text{range}(f) = \{n_1, n_2, n_3, \ldots\} \qquad n_1 < n_2 < n_3 < \cdots$$

    (a) Show that for all $k \in \mathbb{N}$, there exists a unique $a_k \in A$ satisfying $f(a_k) = n_k$.
    (b) Define $g : \mathbb{N} \to A$ by $g(k) = a_k$. Prove that g is a bijection and hence obtain a contradiction.

12. Suppose C is countably infinite and let $c \in C$.
    (a) Show that there exists a bijection $h : \mathbb{N} \to C$ with $h(1) = c$.
    (b) Prove that $D = C \setminus \{c\}$ is countably infinite.
    (*Hint: h be as in part (a) and use g from Example 8.5 to construct $k : \mathbb{N} \to D$*)

13. (a) Suppose $|A| \geq \aleph_0$. Show that there exists $C \subseteq A$ for which $|C| = \aleph_0$.
    (b) Prove that a set A is infinite if and only if it has a proper subset B with $|B| = |A|$.
    (*Hint: use part (a) and Exercise 12*)

14. Describe an *explicit* bijection $g : [0, 1] \to (0, 1)$, and thus demonstrate that the intervals have the same cardinality.
    (*Hint: $\{\frac{1}{n} : n \in \mathbb{N}_{\geq 2}\}$ is a countably infinite subset of $(0, 1)$*)

8.2 Uncountable Sets

Since $\mathbb{Q}$ seems so large, you might think that there couldn't be sets with strictly larger cardinality. But we haven't yet thought about the real numbers...

Definition 8.9. A set A is *uncountable* if $\aleph_0 < |A|$. Otherwise said, there exists an injection $f : \mathbb{N} \to A$ but no bijection $g : \mathbb{N} \to A$.

Theorem 8.10. *The interval $(0,1)$ of real numbers is uncountable.*

We denote the cardinality of the interval $(0,1)$ by $\mathfrak{c}$ for *continuum*. The theorem may therefore be written $\aleph_0 < \mathfrak{c}$. We prove by showing that $\aleph_0 \leq \mathfrak{c}$ and $\aleph_0 \neq \mathfrak{c}$.

Proof. ($\aleph_0 \leq \mathfrak{c}$) The function $f : \mathbb{N} \to (0,1) : n \mapsto \frac{1}{n+1}$ is plainly injective:

$$f(n) = f(m) \implies \frac{1}{n+1} = \frac{1}{m+1} \implies n = m$$

($\aleph_0 \neq \mathfrak{c}$) Suppose, for contradiction, that $g : \mathbb{N} \to (0,1)$ is a bijection. Express the sequence of values $g(1), g(2), g(3), \ldots$ as decimals:[57]

$$\begin{aligned}
g(1) &= 0.b_{11}b_{12}b_{13}b_{14}b_{15}b_{16}\cdots \\
g(2) &= 0.b_{21}b_{22}b_{23}b_{24}b_{25}b_{26}\cdots \\
g(3) &= 0.b_{31}b_{32}b_{33}b_{34}b_{35}b_{36}\cdots \\
g(4) &= 0.b_{41}b_{42}b_{43}b_{44}b_{45}b_{46}\cdots \\
g(5) &= 0.b_{51}b_{52}b_{53}b_{54}b_{55}b_{56}\cdots \\
&\vdots
\end{aligned} \qquad \text{where each } b_{ij} \in \{0,\ldots,9\}$$

Define a new decimal

$$x := 0.x_1x_2x_3x_4x_5\cdots \in (0,1) \qquad \text{where} \quad x_n = \begin{cases} 1 & \text{if } b_{nn} \neq 1 \\ 2 & \text{if } b_{nn} = 1 \end{cases}$$

Since x disagrees with $g(n)$ at the n^{th} decimal place, we see that $x \neq g(n)$: that is, x is *not* in the above list. However $x \in (0,1)$ and g is surjective, so x *must* be in the list: contradiction. ∎

The second part of the proof is known as *Cantor's diagonal argument*, since we compare the constructed decimal x with the diagonal of an infinite square of integers. Since the interval $(0,1)$ is uncountable, and $(0,1) \subseteq \mathbb{R}$, it is immediate that the real numbers are also uncountable. Using only the ideas developed so far (a combination of Exercises 8.1.5 and 14), we could prove directly that every interval of finite length has cardinality $\mathfrak{c}$. It is easier, however, to delay this momentarily...

More amazingly, Cantor's middle-third set (Example 6.18) also has cardinality $\mathfrak{c}$, despite seeming vanishingly small! The details, and more, are in Exercise 12.

[57] A number $x \in (0,1)$ has two decimal representations if and only if one of them terminates and the other ultimately becomes an infinite sequence of 9's: e.g., $0.135 = 0.1349999\cdots$. For the purposes of this proof, choose the terminating decimal when it exists. We restrict to $x_n = 1, 2$ later in the proof to keep away from these double representations.

Non-explict Comparison of Cardinalities

The following result is very useful for comparing cardinalities.

Theorem 8.11 (Cantor–Schröder–Bernstein). *If $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.*

This seems like it should be obvious, but pause for a moment: it is *not* a result about *numbers!* The theorem should be understood in the context of Definition 8.2, in which language it becomes:

> If there exist *injections* $f : A \to B$ and $g : B \to A$, then there exists a *bijection* $h : A \to B$

The proof is beautiful, though a little long to reproduce here; check out any elementary text on set theory if you're interested. Its usefulness is that it allows us to equate cardinalities without the explicit construction of *bijective* functions; *injective* functions are typically much easier to conjure!

Example 8.12. Observe that the following functions are both injective (only—not bijective!):

$$f : (0,1) \to [0,1] : x \mapsto x \qquad (\text{range}(f) = (0,1) \subsetneq [0,1])$$
$$g : [0,1] \to (0,1) : x \mapsto \frac{1}{2}x + \frac{1}{4} \qquad (\text{range}(g) = [\tfrac{1}{4}, \tfrac{3}{4}] \subsetneq (0,1))$$

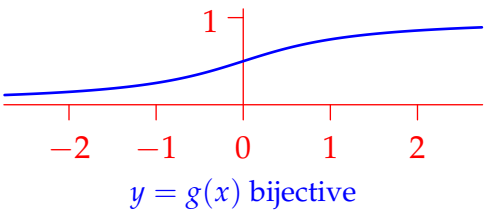
By Cantor–Schröder–Bernstein, the sets $(0,1)$ and $[0,1]$ have the same cardinality ($\mathfrak{c}$). Note how fast this was; we didn't have to construct an explicit bijection $h : (0,1) \to [0,1]$, a much harder business (see Exercise 8.1.14)

As advertised a few moments ago, the Example generalizes to cover all intervals.

Corollary 8.13. *All intervals (of positive length) have cardinality $\mathfrak{c}$.*

Proof. Let A be an interval (could be infinite length) and choose a subinterval $(a,b) \subseteq A$. The following functions are *injective*:

- $f : (0,1) \to A$ where $f(x) = a + (b-a)x$; this maps $(0,1)$ onto $\text{range}(f) = (a,b) \subseteq A$.
- $\iota : A \to \mathbb{R}$ where $\iota(x) = x$.
- $g : \mathbb{R} \to (0,1)$ where $g(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1} x$; this is in fact bijective with inverse $g^{-1}(y) = \tan\left(\pi y - \frac{\pi}{2}\right)$.



Putting everything together:

$$|(0,1)| \overset{f}{\leq} |A| \overset{\iota}{\leq} |\mathbb{R}| \overset{g}{=} |(0,1)|$$

By Cantor–Schröder–Bernstein, all these sets have the same cardinality $\mathfrak{c}$. ∎

Further examples can be found in the exercises.

Cantor's Paradoxical Theorem

For a final punchline, we generalize Theorem 6.8 which, for finite sets A, asserted that $|\mathcal{P}(A)| = 2^{|A|}$ is *strictly larger* than A itself. We now have the technology to attack this for *infinite* sets.

Theorem 8.14 (Cantor). *If A is any set, then $|A| \lneq |\mathcal{P}(A)|$.*

The main implication is that *there is no largest cardinality!* We can always construct a larger cardinality set by taking the power set. For example, $|\mathcal{P}(\mathbb{R})| > |\mathbb{R}| = \mathfrak{c}$. Want a set with even larger cardinality? Try $\mathcal{P}\big(\mathcal{P}(\mathbb{R})\big)$, or $\mathcal{P}\big(\mathcal{P}(\mathcal{P}(\mathbb{R}))\big)$! We can continue this process indefinitely.

Proof. If $A = \emptyset$, the result is trivial. Otherwise, first observe that $f : a \mapsto \{a\}$ defines an injective function $f : A \to \mathcal{P}(A)$, whence $|A| \leq |\mathcal{P}(A)|$.

To complete the argument we must show that no *bijective* function $g : A \to \mathcal{P}(A)$ can exist. Suppose, for a contradiction, that $g : A \to \mathcal{P}(A)$ is bijective, and consider the set

$$X = \big\{a \in A : a \notin g(a)\big\}$$

Tack stock for a moment, since X is hard to think about. Since $g(a)$ is a *subset* of A, the condition $a \notin g(a)$ is legitimate, whence X is a genuine *subset* of A. A simple example will hopefully help.

Example 8.15. Let $A = \{1, 2, 3\}$ and define a function $g : A \to \mathcal{P}(A)$ by

$$g(1) = \{1, 2\}, \qquad g(2) = \{1, 3\}, \qquad g(3) = \emptyset$$

Since $1 \in g(1)$, $2 \notin g(2)$, and $3 \notin g(3)$, we see that $X = \{2, 3\}$. Since our goal is to prove that no *bijection* $A \to \mathcal{P}(A)$ can exist, it is important to note that this function g is *not bijective*; indeed $X \notin \text{range}(g) = \big\{\{1, 2\}, \{1, 3\}, \emptyset\big\}$ shows that this g is not surjective.

Proof Continued. By assumption, g is surjective. Since $\text{range}(g) = \mathcal{P}(A)$, the set X *lies in* $\text{range}(g)$. Otherwise said, $X = g(a)$ for some $a \in A$. We ask whether a is an element of X:

$$\begin{aligned} a \in X &\iff a \notin g(a) && \text{(definition of } X\text{)} \\ &\iff a \notin X && \text{(since } X = g(a)\text{)} \end{aligned}$$

Look at what we have concluded: $a \in X \iff a \notin X$ is plainly a contradiction! We conclude that no bijection $g : A \to \mathcal{P}(A)$ exists, and so $|A| \lneq |\mathcal{P}(A)|$. ∎

Cantor's theorem played a role in pushing set theory towards axiomatization, in part because of the following paradox. If a 'set' is merely a collection of objects, we may consider the 'set of all sets' S. Its power set $\mathcal{P}(S)$ is a set of sets, so must be a subset of S, whence $|\mathcal{P}(S)| \leq |S|$. However, by Cantor's theorem, $|S| \lneq |\mathcal{P}(S))|$, resulting in the manifest absurdity $|S| \lneq |S|$!

The remedy is a rigorous definition of 'set.' *Axiomatic set theory* describes a small number of legitimate ways to build sets, of which we've seen several in these notes: e.g., union, power set, set-builder notation. In particular, the 'set of all sets' *cannot* be legitimately constructed.[58]

[58] The critical condition for preventing Cantor's paradox is that set-builder notation $\{x \in A : P(x)\}$ can only produce a *subset* of an already existing set A. The 'set of all sets' would have the form $\{x : P(x)\}$ where x is *unrestricted*.

Some Final Thoughts on the Limits of Proof

Throughout this course we've learned about some of the basic methods and concepts used by mathematicians. In particular, we've learned how to use proofs to demonstrate the truth of statements about mathematical objects. As we finish, it makes sense to reflect on the limits of our methods.

By the early 20th century, the discovery of various paradoxes and contradictions (like Cantor's) led to a foundational crisis in mathematics. If a concept as basic as *set* is contradictory, how can we have faith in any mathematical conclusion?! The result of this crisis was an effort to place 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. Such an axiomatic foundation ideally would satisfy two conditions:

- *Consistency*: No contradiction could be derived from the axioms.
- *Completeness*: All true mathematical statements could be derived from the axioms.

All hope for such a foundation was crushed in 1931, when Kurt Gödel published his famous *Incompleteness Theorems* which showed that no such axiomatic system could exist. Essentially, Gödel showed that any consistent axiomatic system strong enough to produce some basic arithmetic, there are *undecideable* statements; neither derivable nor refutable from the axioms. Perhaps even worse, no such system can prove its own consistency.

While the strongest aims of early 20th axiomatics cannot be accomplished, contemporary research 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*—as your studies progress, you'll see that many of the objects you study can be formalized as sets together with functions and relations between sets. We've started this work already: Chapter 7 says that functions and relations are themselves sets! Numbers like 0, 1, 2, $\frac{12}{19}$ or even $\pi = 3.14\ldots$ 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 ZFC is subject to the limitations imposed by Gödel's theorems,[59] they have proven able to formalize most of the mathematics actually used by current mathematicians, and have not (thus far!) produced any inconsistencies. While there is plenty of fun to be had exploring set theory,most modern mathematicians feel little need to dwell on the foundational issues of last century!

Exercises 8.2. A reading quiz and practice question can be found online.

1. Decide the cardinality of each set. No working is necessary.

   (a) $\mathbb{N}_{\leq 12}$ (b) $\mathbb{Z}_{\leq 12}$ (c) $(0,5]$ (d) $[2,\pi] \cap \mathbb{Q}$ (e) $\mathcal{P}(\{\mathbb{R}\})$ (f) $\bigcap_{x \in \mathbb{R}^+} [3 - \frac{1}{x}, 3 + \frac{1}{x})$

2. Find *explicit* bijections (thus showing that the given intervals have the same cardinality):

   (a) $f : [2,3) \to [1,5)$ (b) $g : [2,3) \to (1,5]$ (c) $h : (-3,2) \to \mathbb{R}$ (d) $j : \mathbb{R} \to (1,\infty)$

   (*Hint: The proof of Corollary 8.13 should provide some inspiration—be creative*)

3. Let $B = [3,5) \cup (6,10)$. Use the Cantor–Schröder–Bernstein Theorem to prove that $|B| = \mathfrak{c}$.

   (*Hint: State injective functions $f : (0,1) \to B$ and $g : B \to (0,1)$*)

[59] Perhaps the most famous undecidable statement in ZFC is relevant to our recent discussion: the *continuum hypothesis* is the claim that no set has cardinality strictly between $\aleph_0$ and $\mathfrak{c}$; that intervals are the simplest ('smallest') uncountable sets.

4. (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 prove that, for any natural number k, $|\underbrace{\mathbb{N} \times \cdots \times \mathbb{N}}_{k \text{ times}}| = \aleph_0$
   (d) Use part (b) to give an alternative proof that $|\mathbb{Q}^+| = \aleph_0$.

5. Revisit the proof of Cantor's Theorem, and Example 8.15.

   (a) Suppose $g : \{1,2,3,4\} \to \mathcal{P}(\{1,2,3,4\})$ is defined by

   $$g(1) = \{2,3\}, \qquad g(2) = \{1,2\}, \qquad g(3) = \emptyset, \qquad g(4) = \{1,2,4\}$$

   Compute the set $X = \{a \in \{1,2,3,4\} : a \notin g(a)\}$.
   (b) Repeat part (a) for $g : \mathbb{N} \to \mathcal{P}(\mathbb{N}) : n \mapsto \{x \in 2\mathbb{N} : x \leq n\}$.
   (*Hint: Try some examples first! What is* $g(1)$*? Is* $1 \in g(1)$*? What about* $2 \in g(2)$*?*)

6. 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 $x \in \mathbb{N} \implies x + \sqrt{2} \in \mathbb{I}$. Hence conclude that $\aleph_0 \leq |\mathbb{I}|$.
   (c) Argue that the irrational numbers are uncountable ($|\mathbb{I}| = \mathfrak{c}$ is harder to show).
   (d) Show that the *transcendental* (non-algebraic) numbers are uncountable (see Exercise 8.1.10).

7. Give an example of an uncountable set I and an indexed collection $\{A_n : n \in I\}$ for which all the following conditions hold:

   - Each A_n is countably infinite.
   - If $m \neq n$, then $A_m \neq A_n$.
   - For all m, n, either $A_m \subseteq A_n$ or $A_m \supseteq A_n$.
   - $\bigcup_{n \in I} A_n$ is countably infinite.

8. 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:

   $$X = \{A : A \notin A\}$$

   (a) Suppose X is a set. By asking yourself if X is a member of itself, deduce a contradiction.
   (*The point of Russell's paradox is that objects like* X *cannot be considered sets!*)
   (b) Russell's paradox is one version of an ancient logical conundrum with many guises. E.g.,

   > A town has one barber, who cuts the hair of all townspeople, and only those people, who do not cut their own hair: Who cuts the barber's hair?

   Can you explain the connection between this and Russell's paradox?

9. In Exercise 6.3.14, we saw that Cantor's middle-third set $\mathcal{C}$ is the set of all numbers in $[0,1]$ possessing a ternary expansion consisting only of 0's and 2's. By modifying the proof of Theorem 8.10, argue that $\mathcal{C}$ is uncountable.

   (*Exercise 12 establishes the stronger result* $|\mathcal{C}| = \mathfrak{c}$)

The final questions are more of a challenge.

10. Express a real number $x \in (0,1)$ as a decimal $x = 0.x_1x_2x_3x_4\ldots$ where we choose the terminating decimal whenever there is a choice (footnote 57). Prove that

$$f : (0,1) \times (0,1) \to (0,1) \quad \text{defined by} \quad f(x,y) = 0.x_1y_1x_2y_2x_3y_3\ldots$$

is *injective*. Hence conclude that $\left|\mathbb{R}^2\right| = \mathfrak{c}$.

11. If A and B are non-empty sets we let A^B denote the set of all *functions* $f : B \to A$.

    (a) If $A = \{0,1\}$ and $B = \{a,b,c\}$, list all elements of A^B. What is $\left|A^B\right|$?

    (b) If A and B are finite sets, show that $\left|A^B\right| = |A|^{|B|}$.

    (c) Let B be a set and $Y \subseteq B$. The *characteristic function* of Y is $\chi_Y : B \to \{0,1\}$ where

$$\chi_Y(x) = \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{if } x \notin Y \end{cases}$$

    Prove that $\Phi : \mathcal{P}(B) \to \{0,1\}^B$ defined by $\Phi(Y) = \chi_Y$ is *bijective*.
    (*If B is finite, this provides another proof that $|\mathcal{P}(B)| = 2^{|B|}$*)

12. Let $x \in [0,1]$. A *binary expansion* of x is a sequence (b_n) of zeros and ones such that

$$x = \sum_{n=1}^{\infty} \frac{b_n}{2^n}$$

    The binary expansion of $x \in [0,1]$ is (almost) unique;[60] if there is a choice, take the terminating expansion. Define a function $f : [0,1] \to \mathcal{P}(\mathbb{N})$ (the set of subsets of $\mathbb{N}$) by

$$f(x) = \{n \in \mathbb{N} : b_n = 1 \text{ in the binary expansion of } x\}$$

    (a) Prove that f is injective, and that, consequently, $\mathfrak{c} \leq |\mathcal{P}(\mathbb{N})|$.

    (b) Is f surjective?
    (*Hint: consider the set $X = \{2,3,4,\ldots\} \in \mathcal{P}(\mathbb{N})$*)

    (c) Let $\mathcal{C}$ be Cantor's middle-third set. Prove that $g : \mathcal{P}(\mathbb{N}) \to \mathcal{C}$ is a bijection, where

$$g(X) = \sum_{n \in X}^{\infty} \frac{2}{3^n}$$

    (d) Use the Cantor–Schröder–Bernstein Theorem to conclude that $|\mathcal{P}(\mathbb{N})| = |\mathcal{C}| = \mathfrak{c}$.
    (*Together with Exercise 11, this shows why we could write $\mathfrak{c} = 2^{\aleph_0}$*)

[60] Binary expansions are unique unless x has a terminating expansion, in which case the there is a second involving an infinite sequence of 1's: e.g., $[0.011111\cdots]_2 = [0.1]_2$. This example is beloved of computer scientists who, following Exercise 11, might view $\mathcal{P}(\mathbb{N}) \sim \{0,1\}^{\mathbb{N}}$ as the set of *binary sequences/strings* $(x_1, x_2, x_3, \ldots)$, where each $x_j \in \{0,1\}$.