Math 13 — An Introduction to Abstract Mathematics

Neil Donaldson & Alessandra Pantano

December 2, 2015

Contents

Useful Texts

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

Learning Outcomes

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

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

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

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

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

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

1 Introduction

What is Mathematics?

For many students this course is a game-changer. A crucial part of the course is the acceptance that upper-division mathematics is very different from what is presented at grade-school and in the calculus sequence. Some students will resist this fact and spend much of the term progressing through the various stages of grief (denial, anger, bargaining, depression, acceptance) as they discover that what they thought they excelled at isn't really what the subject is about. Thus we should start at the beginning, with an attempt to place the mathematics you've learned within the greater context of the subject.

The original Greek meaning of the word *mathemata* is the supremely unhelpful, "That which is to be known/learned." There is no perfect answer to our question, but a simplistic starting point might be to think of your mathematics education as a progression.

Arithmetic College Calculus Abstract Mathematics

In elementary school you largely learn *arithmetic* and the basic notions of shape. This is the mathematics all of us need in order to function in the real world. If you don't know the difference between 15,000 and 150,000, you probably shouldn't try to buy a new car! For the vast majority of people, arithmetic is the only mathematics they'll ever need. Learn to count, add, and work with percentages and you are thoroughly equipped for most things life will throw at you.

Calculus discusses the relationship between a quantity and its rate of change, the applications of which are manifold: distance/velocity, charge/current, population/birth-rate, etc. Elementary calculus is all about solving problems: What is the area under the curve? How far has the projectile traveled? How much charge is in the capacitor? By now you will likely have computed many integrals and derivatives, but perhaps you have not looked beyond such computations. A mathematician explores the theory behind the calculations. From an abstract standpoint, calculus is the beautiful structure of the Riemann integral and the Fundamental Theorem, understanding *why* we can use anti-derivatives to compute area. To an engineer, the fact that integrals can be used to model the bending of steel beams is crucial, while this might be of only incidental interest to a mathematician. Perhaps the essential difference between college calculus and abstract mathematics is that the former is primarily interested in the *utility* of a technique, while the latter focuses on structure, veracity and the underlying beauty. In this sense, abstract mathematics is much more of an art than a science. No-one measures the quality of a painting or sculpture by how useful it is, instead it is the structure, the artist's technique and the quality of execution that are praised. Research mathematicians, both pure and applied, view mathematics the same way.

In areas of mathematics other than Calculus, the link to applications is often more tenuous. The structure and distribution of prime numbers were studied for over 2000 years before, arguably, any serious applications were discovered. Sometimes a real-world problem motivates generalizations that have no obvious application, and may never do so. For example, the geometry of projecting 3D objects onto a 2D screen has obvious applications (TV, computer graphics/design). Why would anyone want to consider projections from 4D? Or from 17 dimensions? Sometimes an application will appear later, sometimes not.[1] The reason the mathematician studies such things is because the structure appears beautiful to them and they want to appreciate it more deeply. Just like a painting.

[1] There are very useful applications of high-dimensional projections, not least to economics and the understanding of sound and light waves.

The mathematics you have learned so far has consisted almost entirely of computations, with the theoretical aspects swept under the rug. At upper-division level, the majority of mathematics is presented in an abstract way. This course will train you in understanding and creating abstract mathematics, and it is our hope that you will develop an appreciation for it.

Proof

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

1. An argument that establishes the truth of a fact.
2. A test or trial of an assertion.[2]

In mathematics we always mean the former, while in much of science and wider culture the second meaning predominates. Indeed mathematics is one of the very few disciplines in which one can categorically say that something is *true* or *false.* In reality we can rarely be so certain. A greasy salesman in a TV advert may claim that to have *proved* that a certain cream makes you look younger; a defendant may be *proved* guilty in court; the gravitational constant *is* 9.81ms^{-2}. Ask yourself what these statements mean. The advert is just trying to sell you something, but push harder and they might provide some justification: e.g. 100 people used the product for a month and 75 of them claim to look younger. This is a *test,* a proof in the second sense of the definition. Is a defendant really guilty of a crime just because a court has found them so; have there never been any miscarriages of justice? Is the gravitational constant precisely 9.81ms^{-2}, or is this merely a good approximation? This kind of pedantry may seem over the top in everyday life: indeed most of us would agree that if 75% of people think a cream helps, then it probably is doing something beneficial. In mathematics and philosophy, we think very differently: the concepts of true and false and of proof are very precise.

So how do mathematicians reach this blissful state where everything is either right or wrong and, once proved, is forever and unalterably certain? The answer, rather disappointingly, is by cheating. *Nothing* in mathematics is true except with reference to some assumption. For example, consider the following theorem:

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

We all believe that this is true, but can we *prove* it? In the sense of the second definition of proof, it might seem like all we need to do is to test the assertion: for example $4 + 6 = 10$ is even. In the first sense, the *mathematical* sense, of proof, this is nowhere near enough. What we need is a *definition* of even.[3]

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

The proof of the theorem now flows straight from the definition.

[2] It is this notion that makes sense of the seemingly oxymoronic phrase *The exception proves the rule.* It is the exception that *tests* the validity of the rule.

[3] And more fundamentally of *sum* and *integer.*

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

$$x + y = 2m + 2n = 2(m + n). \qquad (*)$$

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

There are several important observations:

- 'Any' in the statement of the theorem means the proof must work *regardless* of what even integers you choose. It is not good enough to simply select, for example, 4 and 16, then write $4 + 16 = 20$. This is an *example,* or test, of the theorem, not a mathematical proof.
- According to the definition, $2m$ and $2n$ together represent *all possible pairs* of even numbers.
- The proof makes direct reference to the definition. The vast majority of the proofs in this course are of this type. If you know the definition of every word in the statement of a theorem, you will often discover a proof simply by writing down the definitions.
- The theorem itself did not mention any *variables.* The proof required a calculation for which these were essential. In this case the variables m and n come for free *once you write the definition of evenness!* A great mistake is to think that the proof is nothing more than the calculation $(*)$. This is the easy bit, and it means nothing without the surrounding sentences.

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

Conjectures

In this course, you will discover that one of the most creative and fun aspects of mathematics is the art of formulating, proving and disproving conjectures. To get a taste, consider the following:

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

Conjecture 1.4. *For every positive integer n, the integer $n^2 + n + 41$ is prime.*

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

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

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

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

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

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

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

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

Let us work with Conjecture 1.3. If n is an odd integer, then, by definition, we can write it as $n = 2k+1$ for some integer k. Then

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

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

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

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

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

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

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

We have therefore expressed $k^2 + k$ as a product of two consecutive integers. This is great, because for any two consecutive integers, one is even and the other is odd, and so their product must be even. We have now proved that the conjecture is true. Conjecture 1.3 is indeed a *theorem!* Everything we've done so far has been investigative, and is laid out in an untidy way. We don't want the reader to have to wade through all of our scratch work, so we formalize the above argument. This is the final result of our deliberations; investigate, spot a pattern, conjecture, prove, and finally present your work in as clean and convincing a manner as you can.

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

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

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

We distinguish two cases. If k is even, then $k(k+1)$ is even and so $4k(k+1)$ is divisible by 8.
If k is odd, then $k+1$ is even. Therefore $k(k+1)$ is again even and $4k(k+1)$ divisible by 8.
In both cases $n^2 - 1 = 4k(k+1)$ is divisible by 8. This concludes the proof. ■

It is now time to explore Conjecture 1.4. The question here is whether or not $n^2 + n + 41$ is a prime integer for *every* positive integer n. We know that when $n = 1$, 2, 3, 4, 5, 6 or 7 the answer is yes, but examples do not make a proof. At this point, we do not know whether the conjecture is true or false. Let us investigate the question further. Suppose that n is any positive integer; we must ask whether it is possible to factor $n^2 + n + 41$ as a product of two positive integers, neither of which is one.[4] When $n = 41$ such a factorization certainly exists, since we can write

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

Our *counterexample* shows that there exists at least one value of n for which $n^2 + n + 41$ is *not* prime. We have therefore disproved the conjecture that 'for all positive integers n, $n^2 + n + 41$ is prime,' and so Conjecture 1.4 is false!

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

Conjectures: *True* or *False*?

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

1. The sum of any three consecutive integers is even.
2. There exist integers m and n such that $7m + 5n = 4$.
3. Every common multiple of 6 and 10 is divisible by 60.
4. There exist integers x and y such that $6x + 9y = 10$.
5. For every positive real number x, $x + \frac{1}{x}$ is greater than or equal to 2.
6. If x is any real number, then $x^2 \geq x$.

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

7. If n is any integer, $n^2 + 5n$ must be even.
8. If x is any real number, then $|x| \geq -x$.
9. Consider the set $\mathbb{R}$ of all real numbers. For all x in $\mathbb{R}$, there exists y in $\mathbb{R}$ such that $x < y$.
10. Consider the set $\mathbb{R}$ of all real numbers. There exists x in $\mathbb{R}$ such that, for all y in $\mathbb{R}$, $x < y$.
11. The sets $A = \{n \in \mathbb{N} : n^2 < 25\}$ and $B = \{n^2 : n \in \mathbb{N} \text{ and } n < 5\}$ are equal. Here $\mathbb{N}$ denotes the set of natural numbers.

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

2 Logic and the Language of Proofs

In order to read and construct proofs, we need to start with the langauge in which they are written: *logic*. Logic is to mathematics what grammar is to English. Section 2.1 will not look particularly mathematical, but we'll quickly get to work in Section 2.2 using logic in a mathematical context.

2.1 Propositions

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

Examples. 1. $17 - 24 = 7$.

2. 39^2 is an odd integer.
3. The moon is made of cheese.
4. Every cloud has a silver lining.
5. God exists.

In order to make sense, these propositions require a clear *definition* of every concept they contain. There are many concepts of God in many cultures, but once it is decided *which* we are talking about, it is clear that They either exist or do not. This example illustrates that a question need not be indisputably answerable (by us) in order to qualify as a proposition. Indeed mostly when people argue over propositions and statements, what they are really arguing over are the defintions!

Anything that is not true or false is not a proposition. *January* 1^{st} is not a proposition, neither is *Green*.

Truth Tables

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

P	Q
T	T
T	F
F	T
F	F

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

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

Connecting Propositions: Conjunction, Disjunction and Negation

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

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

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

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

P	$\neg P$
T	F
F	T

It is best to use *and, or* and *not* when speaking about these concepts: conjunction, disjunction and negation may make you sound educated, but at the serious risk of not being understood!

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

P. Irvine is a city in California.

Q. Irvine is a town in Ayrshire, Scotland.

R. Irvine has seven letters.

Clearly P is true while R is false. If you happen to know someone from Scotland, you might know that Q is true.[5] We can now compute the following (increasingly grotesque) combinations...

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

How did we establish these facts? Some are quick, and can be done in your head. Consider, for instance, the statement $(\neg R) \wedge P$. Because R is false, $\neg R$ is true. Thus $(\neg R) \wedge P$ is the conjunction of two true statements, hence it is true. Similarly, we can argue that $R \vee P$ is true (because R is false and P is true), so the negation $\neg(R \vee P)$ is false.

Establishing the truth value of the final proposition $(\neg P) \vee [((\neg R) \vee P) \wedge Q]$ requires more work. You may want to set up a truth table with several auxiliary columns to help you compute:

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

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

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

Conditional and Biconditional Connectives

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

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

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

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

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

Observe that the expressions $P \implies Q$ and $P \iff Q$ are themselves *propositions*. They are, after all, sentences which are either true or false!

Synonyms

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

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

For instance, the following propositions all mean exactly the same thing:

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

Are you comfortable with what P and Q are here?

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

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

This is the most immediately confusing part of the truth table for the conditional connective. Here is a mathematical example, written with an English translation at the side.

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

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

If you want to understand connectives more deeply than this, then take a logic or philosophy course! For example, although neither statement makes the least bit of sense in English;

$$\begin{gathered} 17 \text{ is odd} \implies \text{Mexico is in China is } \textit{false}, \\ \text{whilst} \\ 17 \text{ is even} \implies \text{Mexico is in China is } \textit{true}. \end{gathered}$$

Such bizarre constructions are happily beyond the consideration of this course!

Theorems and Proofs

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

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

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

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

We can write the theorem in terms of propositions and connectives:

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

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

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

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

The Converse and Contrapositive

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

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

In general, we can't say anything about the truth value of the converse of a true statement. The contrapositive of a true statement is, however, *always* true.

Theorem 2.7. *The contrapositive of an implication is logically equivalent the original implication.*

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

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

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

Example. Let P and Q be the following statements:

P. Claudia is holding a peach.

Q. Claudia is holding a piece of fruit.

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

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

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

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

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

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

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

This is clearly true.

The fact that $P \implies Q$ and $\neg Q \implies \neg P$ are logically equivalent allows us, when convenient, to prove $P \implies Q$ by instead proving its contrapositive...

Proof by Contrapositive

Here is another basic theorem.

Theorem 2.8. *Let x and y be two integers. If $x + y$ is odd, then exactly one of x or y is odd.*

The statement of the theorem is an implication of the form $P \implies Q$. Here we have

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

Q. Exactly one of x or y is odd.

A direct proof would require that we assume P is true and logically deduce the truth of Q. The problem is that it is hard to work with these propositions, especially Q. The negation of Q is, however, much easier:

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

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

Since $P \implies Q$ is logically equivalent to the simpler-seeming contrapositive $(\neg Q) \implies (\neg P)$, we choose to prove the latter. This is, after all, equivalent to proving the original implication.

Proof. There are two cases: x and y are both even, or both odd.
Case 1: Let $x = 2m$ and $y = 2n$ be even. Then $x + y = 2(m + n)$ is even.
Case 2: Let $x = 2m + 1$ and $y = 2n + 1$ be odd. Then $x + y = 2(m + n + 1)$ is even. ∎

The above is an example of a *proof by contrapositive*.

De Morgan's Laws

Two of the most famous results in logic are attributable to Augustus De Morgan, a very famous 19th century logician.

Theorem 2.9 (De Morgan's laws)**.** *Let P and Q be any propositions. Then:*

1. $\neg(P \wedge Q) \iff \neg P \vee \neg Q$.
2. $\neg(P \vee Q) \iff \neg P \wedge \neg Q$.

The first law says that the negation of $P \wedge Q$ is logically equivalent to $\neg P \vee \neg Q$: the two expressions have the *same truth table.* Here is a proof of the first law. Try the second on your own.

Proof.

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

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

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

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

I rode the subway *and* I had coffee.

What is its negation (opposite)? Clearly it is:

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

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

You will see these laws again when we think about sets.

Aside: Think about the meaning!

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

Mark and Mary have the same height.

is the proposition:

Mark and Mary do not have the same height.

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

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

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

Negating Conditionals

As our discussion of contrapositives makes clear, you will often want to understand the negation of a statement. In particular, it is important to understand the negation of a conditional $P \implies Q$. Is it enough to say 'P doesn't imply Q'? And what could this mean? To answer the question you can use truth tables, or just think.

Here is the truth table for $P \implies Q$ and its negation: recall that negation simply swaps T and F.

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

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

Theorem 2.10. *$\neg(P \implies Q)$ is logically equivalent to $P \wedge \neg Q$ (read 'P and not Q').*

Now *think* rather than calculate. What is the opposite of the following implication?

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

Hopefully it is clear that the negation is:

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

The implication *'therefore'* has disappeared and a conjuction *'and'* is in its place.

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

The converse, $Q \implies P$.

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

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

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

If x is even then x^2 is even.

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

P. x is even.

Q. x^2 is even.

Hence the negation is $P \wedge \neg Q$, which is:

x is even and x^2 is odd.

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

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

Tautologies and Contradictions

There are two final related concepts that are helpful for understanding proofs.

Definition 2.11. A *tautology* is a logical expression that is always true, regardless of what the component statments might be.
A *contradiction* is a logical expression that is always false.

The easiest way to detect these is simply to construct a truth table.

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

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

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

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

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

Aside: Algebraic Logic

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

$$P \wedge Q \iff Q \wedge P \qquad P \vee Q \iff Q \vee P$$
$$(P \wedge Q) \wedge R \iff P \wedge (Q \wedge R), \qquad (P \vee Q) \vee R \iff P \vee (Q \vee R),$$
$$(P \wedge Q) \vee R \iff (P \vee R) \wedge (Q \vee R), \qquad (P \vee Q) \wedge R \iff (P \wedge R) \vee (Q \wedge R).$$

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

Exercises

2.1.1 Express each of the following statements in the "If . . . , then . . . " form.

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

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

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

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

2.1.2 Suppose that "Girls smell of roses" and "Boys have dirty hands" are true statements and that "The Teacher is always right" is a false statment. Which of the following are true?
Hint: Label each of the given statements, and think about each of the following using connectives.

(a) If girls smell of roses, then the Teacher is always right.

(b) If the Teacher is always right, then boys have dirty hands.

(c) If the Teacher is always right or girls smell of roses, then boys have dirty hands.

(d) If boys have dirty hands and girls smell of roses, then the Teacher is always right.

2.1.3 Write the negation (in words) of the following claim:

If Jack and Jill climb up the hill, then they fall down and like pails of water.

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

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

2.1.6 Use De Morgan's laws to prove that $P \implies Q$ is logically equivalent to $\neg P \vee Q$.

2.1.7 Prove that the expressions $(P \implies Q) \wedge (Q \implies P)$ and $P \iff Q$ are logically equivalent (have the same truth table). Why does this make sense?

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

2.1.9 Prove that $(\neg P \wedge Q) \vee (P \wedge \neg Q) \iff \neg(P \iff Q)$ is a tautology:

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

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

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

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

(i) Someone was playing pool.

(ii) Colin was late.

2.1.11 Suppose that "Ford is tired and Zaphod has two heads" is a false statement. What can we conclude if we discover each of the following? *Treat the two scenarios separately.*

(a) Ford is tired.

(b) Ford is tired if and only if Zaphod has two heads.

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

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

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

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

(a) What can you say about S?

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

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

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

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

2.2 Methods of Proof

There are four standard methods for proving $P \implies Q$. In practice, long proofs will use several of these.

Direct Assume P and logically deduce Q.

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

Contradiction Assume that P and $\neg Q$ are true and deduce a *contradiction*. Since $P \wedge \neg Q$ implies a contradiction, this shows that $P \wedge \neg Q$ must be *false*. Because $P \wedge \neg Q$ is equivalent to $\neg(P \implies Q)$, this is enough to conclude that $P \implies Q$ is true (Theorem 2.10).

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

The direct method has the advantage of being easy to follow logically. The contrapositive method has its advantage when it is difficult to work directly with the propositions P, Q, especially if one or both involve the *non-existence* of something. Working with their negations might give you the existence of ingredients with which you can calculate. Proof by contradiction has a similar advantage: assuming both P and $\neg Q$ gives you two pieces of information with which you can calculate. Logically speaking there is no difference between the three methods, beyond how you visualize the argument.

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

Theorem 2.12. *Suppose that x is an integer. If $3x + 5$ is even, then $3x$ is odd.*

Direct Proof. We show that if $3x + 5$ is even then $3x$ is odd. Assume that $3x + 5$ is even, then $3x + 5 = 2n$ for some integer n. Hence

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

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

Contrapositive Proof. We show that if $3x$ is even then $3x + 5$ is odd. Assume that $3x$ is even, and write $3x = 2n$ for some integer n. Then

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

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

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

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

But this says that 5 is even: a contradiction. ∎

Some simple proofs

We now give several examples of simple proofs. The only notation needed to speed things along is that of some basic sets of numbers: $\mathbb{N}$ for the positive integers, $\mathbb{Z}$ for the integers, $\mathbb{R}$ for the real numbers, and $\in$ for 'is a member of the set'. Thus $2 \in \mathbb{Z}$ is read as '2 is a member of the set of integers', or more concisely, '2 is an integer'.

Theorem 2.13. *Let $m, n \in \mathbb{Z}$. Both m and n are odd if and only if the product mn is odd.*

There are two theorems here:

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

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

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

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

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

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

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

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

∎

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

In the second part of the proof, we did not need to consider whether m was even or odd: if n was even, the product mn would be even regardless. The second part would have been very difficult to prove directly: Assume mn is odd, then $mn = 2k + 1$, so... We are stuck!

Theorem 2.14. *If $3x + 5$ is even, then x is odd.*

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

Direct Proof. Simply quote the two previous theorems. Because $3x + 5$ is even, $3x$ must be odd by Theorem 2.12. Now, since $3x$ is odd, both 3 and x are odd by Theorem 2.13. ∎

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

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

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

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

$$5 = (3x + 5) - 3x = 2m - 6k = 2(m - 3k) \text{ is even. Contradiction.}$$

∎

Selecting a method of proof is often a matter of taste. You should be able to see the advantages and disadvantages of the various approaches. The direct proof is more logically straightforward, but it depends on two previous results. The contrapositive and the contradiction arguments are quicker and more self-contained, but they require a deeper familiarity with logic.[6]

Potential Mistakes: Generality and 'Without Loss of Generality'

There are many common mistakes that you should be careful to avoid. Here are two incorrect 'proofs' of the $\implies$ direction of Theorem 2.13.

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

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

Fake Proof 2. Let $m = 2k + 1$ and $n = 2k + 1$ be odd. Then, $mn = (2k + 1)(2k + 1) = 2(2k^2 + 2k) + 1$ is odd. ∎

The problem with this second 'proof' is that it is not sufficiently general. m and n are supposed to be *any* odd integers, but by setting both of them equal to $2k + 1$, we've chosen m and n to be the

[6] For even more variety, here is a direct proof of Theorem 2.14 that does not use any previous theorem. Suppose $3x + 5$ is even, so $3x + 5 = 2n$ for some integer n. Then

$$x = (3x + 5) - 2x - 5 = 2n - 2x - 5 = 2(x - n - 3) + 1 \text{ is odd.}$$

You will often have a variety of possible approaches: this just makes proving theorems even more fun!

same! Notice how the correct proof uses $m = 2k + 1$ and $n = 2l + 1$, where we place no restriction on the integers k and l.

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

Think back to how this was used in the the proof of Theorem 2.13. If at least one of integers m, n is even, then we lose nothing by assuming that it is the second integer n. The labels m, n are arbitrary: if n happened not to be even, we could simply relabel the integers, changing their order so that the second is now even.

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

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

Here is a palpably ludicrous 'theorem' which illustrates another potential mistake.

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

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

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

■

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

More simple proofs

Theorem 2.15. *Suppose $x \in \mathbb{R}$. Then $x^3 + 2x^2 - 3x - 10 = 0 \implies x = 2$.*

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

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

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

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

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

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

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

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

On balance the contrapositive proof is probably the nicest, but you may decide for yourself.

Aside: Being Excessively Logical

The statement of Theorem 2.15 is an implication $P \implies Q$ where P and Q are:

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

You can make life very hard for yourself by being overly logical. For instance, you may wish take a third proposition R. $x \in \mathbb{R}$, and state the theorem as $R \implies (P \implies Q)$. This is the way of pain! It's easier to assume that you're always dealing with real numbers as a universal constraint, and ignore it entirely in the logic.

One can always append a third proposition to the front of any theorem, namely, "all math I already know." Try to resist the temptation to be so logical that your arguments become unreadable!

Theorem 2.16. *If $n \in \mathbb{Z}$ is divisible by $p \in \mathbb{N}$, then n^2 is divisible by p^2.*

Before trying to prove this, recall what 'n is divisible by p' means: that $n = pk$ for some integer k. With the correct definition, the proof is immediate.

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

Remember: state the definition of everything important in the theorem and often the proof will be staring you in the face.

Proof by Cases

The next proof involves breaking things into cases. The relevant definition here is that of *remainder.* An integer n is said to have remainder $r = 0, 1$, or 2 upon division by 3 if we can write $n = 3k + r$ for some integer k. With a little thought, it should be clear that every integer is of the form $3k, 3k + 1$, or $3k + 2$. This is analogous to how all integers are either even $(2k)$ or odd $(2k + 1)$. We will consider remainders more carefully in Chapter 3.

Theorem 2.17. *If n is an integer, then n^2 has remainder 0 or 1 upon dividing by 3.*

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

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

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

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

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

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

Thus n^2 has remainder 0 or 1 and cannot have remainder 2. ∎

Non-existence Proofs

When a Theorem claims that something does not exist, it is generally a good time for a contrapositive or contradiction proof. This is since 'does not exist' is already a *negative* condition. A contradiction or contrapositive proof of a theorem $P \implies Q$ already involve the negated statement $\neg Q$. If Q states that something does not exist, then $\neg Q$ states that it does! To see this in action, consider the following result.

Theorem 2.18. *$x^{17} + 12x^3 + 13x + 3 = 0$ has no positive (real number) solutions.*

First we interpret the theorem as an implication: throughout we assume that x is a real number.

If x is a solution to the equation $x^{17} + 12x^3 + 13x + 3 = 0$, then $x \leq 0$.

The theorem is of the form $P \implies Q$, with:

P. $x^{17} + 12x^3 + 13x + 3 = 0$, $\qquad$ Q. $x \leq 0$.

The negation of Q is simply '$x > 0$.' To prove the theorem by contradiction, we assume P and not Q, and deduce a contradiction.

Proof. Assume that x satisfies $x^{17} + 12x^3 + 13x + 3 = 0$, and that $x > 0$. Because all terms on the left hand side are positive, we have $x^{17} + 12x^3 + 13x + 3 > 0$. A contradiction. ∎

Note how quickly the proof is written: it assumes that you, and any reader, are familiar with the underlying logic of a contradiction proof without it needing to be spelled out. The discussion we undertook before writing the proof would be considered scratch work: you shouldn't include it a final write-up.

If you recall the Intermediate and Mean Value Theorems from Calculus, you should be able to prove that there is exactly one (necessarily negative!) solution to the above polynomial equation.

The AM-GM inequality

Next we give several proofs of a famous inequality relating the arithmetic and geometric means of two or more numbers.

Theorem 2.19. *If x, y are positive real numbers, then $\frac{x+y}{2} \geq \sqrt{xy}$ with equality if and only if $x = y$.*

First a direct proof: note how the implication signs are stacked to make the argument easy to read.

Direct Proof. Clearly $(x-y)^2 \geq 0$ with equality $\iff x = y$. Now multiply out:

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

The square-root in $(*)$ is well-defined because $x + y$ is positive.[a] Moreover, it is clear that the final inequality is an equality if and only if all of them are, which is if and only if $x = y$. ∎

[a] We are using the fact that the function $f(t) = t^2$ is increasing for t positive.

The argument for 'with equality if and only if $x = y$' depended on all of the implications in the proof are *biconditionals.*

The following contradiction proof incorporates exactly the same calculation, but is laid out in a different order. This is not always possible, and you have to take great care when trying it. You will likely agree that the direct proof is easier to follow.

Contradiction Proof. Suppose that $\frac{x+y}{2} < \sqrt{xy}$. Since $x + y \geq 0$, this is if and only if $(x+y)^2 < 4xy$. Now multiply out and rearrange:

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

Since squares of real numbers are non-negative, this is a contradiction. Thus $\frac{x+y}{2} \geq \sqrt{xy}$.
Now suppose that $\frac{x+y}{2} = \sqrt{xy}$. Following the biconditionals through the proof, we see that this is if and only if $(x-y)^2 = 0$, from which we recover $x = y$. Hence result. ∎

Aside: The general AM-GM inequality

Both the statement and the proof of the general inequality are more difficult. You might be surprised that an argument involving 'raising to the nth power' doesn't work. Try it and see why… The general proof is harder and we present it at a higher level, leaving out some of the more obvious details. This helps us view the proof as a whole, and makes the logical flow clearer. The only prerequisite is a little calculus, namely the First Derivative Test at the end of the first paragraph.

Theorem 2.20. *If $x_1, \ldots, x_n > 0$ then $\frac{x_1+\cdots+x_n}{n} \geq \sqrt[n]{x_1 \cdots x_n}$, with equality if and only if $x_1 = \cdots = x_n$.*

Proof. Consider the function $f(x) = e^{x-1} - x$. Its derivative is $f'(x) = e^{x-1} - 1$, which is zero if and only if $x = 1$. The sign of the derivative changes from negative to positive at $x = 1$, whence this is a local minimum. f has no other critical points and its domain is the whole real line, whence $x = 1$ is the location of the *global* minimum of f. Since $f(1) = 0$, we have $e^{x-1} \geq x$ with equality if and only if $x = 1$.

Now consider the average $\mu = \frac{x_1+x_2+\cdots+x_n}{n}$. Applying our inequality to $x = \frac{x_i}{\mu}$, we have

$$\frac{x_i}{\mu} \leq \exp\left(\frac{x_i}{\mu} - 1\right), \qquad \text{for each } i = 1, 2, \ldots, n. \tag{$*$}$$

Since all x_i are positive, we may multiply the expressions $(*)$ while preserving the inequality:

$$\frac{x_1}{\mu} \cdots \frac{x_n}{\mu} \leq \exp\left(\frac{x_1}{\mu} - 1 + \cdots + \frac{x_n}{\mu} - 1\right) = \exp(n-n) = 1. \tag{$\dagger$}$$

Thus $\mu^n \geq x_1 \cdots x_n$ from which the result, $\mu \geq \sqrt[n]{x_1 \cdots x_n}$, follows.

Equality is if and only if all the inequalities $(*)$ are equalities, which is if and only if $x_i = \mu$ for all $i = 1, \ldots, n$. That is, all the x_i are equal. ∎

Given the theorem and proof are both more difficult, there are a few things you should do to help convince yourself of their legitimacy.

1. Write down some examples. E.g. if $x_1 = 20, x_2 = 27, x_3 = 50$, the inequality reads

$$\frac{97}{3} \geq \sqrt[3]{20 \cdot 27 \cdot 50} = 30.$$

2. Observe that Theorem 2.19 is a special case.
3. Work through the proof, inserting comments and extra calculations until you are convinced that the argument is correct. For example, the calculation $\frac{x_1+\cdots+x_n}{\mu} = \frac{\mu n}{\mu} = n$ was omitted from (†): anyone with the prerequisite knowledge to read the rest of the proof should easily be able to supply this.

It is perfectly reasonable to ask how you would know to try such a proof. The answer is that you wouldn't. You should appreciate that a proof like this is a distillation of thousands of attempts and improvements, perhaps over many years. No-one came up with this argument as a first attempt!

Combining and Subdividing Theorems

Sometimes it is useful to break a proof into pieces, akin to viewing a computer program as a collection of subroutines that you combine for the finale. Usually the purpose is to make the proof of a difficult result more readable, but it can be done to emphasize the importance of certain aspects of your work. Mathematics does this by using *lemmas* and *corollaries.*

> *Lemma:* a theorem whose importance you want to downplay. Often the result is individually unimportant, but becomes more useful when incorporated as part of a larger theorem.
>
> *Corollary:* a theorem which follows quickly from a larger result. Corollaries can be used to draw attention to a particular aspect or a special case of a theorem.

In many mathematical papers the word *theorem* is reserved only for the most important results, everything else being presented as a lemma or corollary. The choice of what to call a result is entirely one of presentation. If you want your paper to be more readable, or to highlight the what you think is important, then lemmas and corollaries are your friends!

Here is a famous example of a lemma at work.

Lemma 2.21. *Suppose that $n \in \mathbb{Z}$. Then n^2 is even $\iff$ n is even.*

Prove this yourself: the ($\Rightarrow$) direction is easiest using the contrapositive method, while the ($\Leftarrow$) direction works well directly.

Theorem 2.22. *$\sqrt{2}$ is irrational.*

This is tricky for a few reasons. The theorem does not appear to be of the form $P \implies Q$, but in fact it is. Consider:

Q. $\sqrt{2}$ is irrational.

P. Everything you already know in mathematics!

Of course P is entirely unhelpful; How would we start a direct proof when we don't know what to choose from the whole universe of mathematics? A contrapositive proof might also be difficult: $\neg Q$ straightforwardly states that $\sqrt{2}$ is rational, but $\neg P$ is the cryptic statement, 'something we know happens to be false.' But what is the *something*? Instead we use a proof by contradiction.

Proof. Suppose that $\sqrt{2} = \frac{m}{n}$ for some $m, n \in \mathbb{N}$, where m, n have no common factors.
Then $m^2 = 2n^2$ which says that m^2 is even.
By Lemma 2.21 we have that m is even.
Thus $m = 2k$ for some $k \in \mathbb{Z}$.
But now, $n^2 = 2k^2$, from which (Lemma 2.21 again) we see that n is even.
Thus m and n have a common factor of 2. This is a contradiction. ∎

First observe how Lemma 2.21 was used to make the proof easier to read. Now try to make sense of the proof. The main challenge comes in the first line. Once we assume that $\sqrt{2} = \frac{m}{n}$, we can immediately insist that m, n have no common factors. It is important to realize that this is *not* the assumption being contradicted. Indeed it is no real restriction *once we assume that* $\sqrt{2}$ *is rational.* If you find this approach difficult, you may prefer the alternative proof given in the exercises.

Here is another famous result involving prime numbers.

Definition 2.23. A positive integer $p \geq 2$ is *prime* if its only positive divisors are itself and 1.

The first few primes are $2, 3, 5, 7, 11, 13, 17, 19, \ldots$. It follows[7] from the definition that all positive integers ≥ 2 are either primes or *composites* (products of primes).

Theorem 2.24. *There are infinitely many prime numbers.*

To break down the proof we first prove a lemma: the symbol $:=$ is read 'defined to be equal to.'

Lemma 2.25. *Suppose that $p_1, \ldots, p_n$ are integers ≥ 2. Then $\Pi := p_1 p_2 \cdots p_n + 1$ is not divisible by p_i for any i.*

Proof. Suppose that Π is divisible by p_i. Observe that

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

Since $p_1 \cdots p_n$ is divisible by p_i, the left hand side of this equation is divisible by p_i. But then 1 must be disvisible by p_i. Since $p_i \geq 2$, this is a contradiction. ∎

[7] This is not obvious: we will prove it much later in Theorem 5.16.

Proof of theorem. Again we prove by contradiction. Assume that there are exactly n prime numbers $p_1, \ldots, p_n$ and consider $\Pi := p_1 \cdots p_n + 1$. By the lemma, Π is not divisible by any of the primes $p_1, \ldots, p_n$. There are two cases:

(a) Π is prime, in which case it is a *larger* prime than anything in our list $p_1, \ldots, p_n$.

(b) Π is composite, in which case it is divisible by a prime. But this prime cannot be in our list $p_1, \ldots, p_n$.

In either case we've shown that there is another prime not in the list $p_1, \ldots, p_n$, and we've contradicted our assumption that we had *all* the primes.

■

The lemma approach was almost essential for this example, since both the lemma and the theorem were proved by contradiction. Nesting one contradiction argument within another is a recipe for serious confusion!

Exercises

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

2.2.2 Prove or disprove the following conjectures.

(a) There is an even integer which can be expressed as the sum of three even integers.

(b) Every even integer can be expressed as the sum of three even integers.

(c) There is an odd integer which can be expressed as the sum of two odd integers.

(d) Every odd integer can be expressed as the sum of three odd integers.
To get a feel about whether a claim is true or false, try out some examples. If you believe a claim is false, provide a specific counterexample. If you believe a claim is true, give a (formal) proof.

2.2.3 Prove or disprove the following conjectures:

(a) The sum of any 3 consecutive integers is divisible by 3.

(b) The sum of any 4 consecutive integers is divisible by 4.

(c) The product of any 3 consecutive integers is divisible by 6.

2.2.4 Augustus De Morgan satisfied his own problem:

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

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

(b) Suppose you have an acquaintance who satisfies the same problem. How old will they turn in 2014?

Give a formal argument which justifies that you are correct.

2.2.5 Prove that if n is a natural number greater than 1, then $n! + 2$ is even.
Here $n!$ denotes the factorial *of the integer n. Look up the definition if you forgot about it.*

2.2.6 Let $x, y \in \mathbb{Z}$. Prove that if xy is odd, then x and y are odd.

2.2.7 (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?

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

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

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

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

2.2.9 Given below is the proof of a result. What is the result?

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

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

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

2.2.10 Suppose that x, and y are real numbers. Prove that if $3x + 5y$ is irrational, then at least one of x and y is irrational. *Recall that x is irrational if it cannot be written as a ratio of integers.*

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

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

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

2.2.14 You meet three old men, Alain, Boris, and César, each of whom is a Truthteller or a Liar. Truthtellers speak only the truth; Liars speak only lies. You ask Alain whether he is a Truthteller or a Liar. Alain answers with his back turned, so you cannot hear what he says.
"What did he say?" you ask Boris.
Boris says: "Alain says he is a Truthteller."
César says: "Boris is lying."
Is César a Truthteller or a Liar? Explain your answer.

2.2.15 *(Snake-like integers)* Let's say that an integer y is *Snake-like* if and only if there is some integer k such that $y = (6k)^2 + 9$.

(a) Give three examples and three non-examples of Snake-like integers.

(b) Given $y \in \mathbb{Z}$, compute the negation of the statement, 'y is Snake-like.'

(c) Show that every Snake-like integer is a multiple of 9.

(d) Show that the statements, 'n is Snake-like,' and, 'n is a multiple of nine,' are not equivalent.

2.2.16 Assume that Ben's father lives in Peru. Consider the following implication β:

> If Ben's father is an artist and does not have any friends in Asia, then Ben plays tennis or ping-pong, or he appeared in at least one picture of the May 1992 Time magazine.

(a) Find the contrapositive of β.

(b) Find the converse of β.

(c) Find the negation of β.

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

2.2.17 Here is an alternative argument that $\sqrt{2}$ is irrational. Suppose that $\sqrt{2} = \frac{m}{n}$ where $m, n \in \mathbb{N}$. This time we don't assume that m, n have no common factors.

(a) m, n satisfy the equation $m^2 = 2n^2$. Prove that there exist positive integers m_1, n_1 which satisfy the following three conditions:

$$m_1^2 = 2n_1^2, \qquad m_1 < m, \qquad n_1 < n.$$

(b) Show that there exist two sequences of decreasing positive integers $m > m_1 > m_2 > \cdots$ and $n > n_1 > n_2 > \cdots$ which satisfy $m_i^2 = 2n_i^2$ for all $i \in \mathbb{N}$.

(c) Is it possible to have an infinite sequence of decreasing *positive* integers? Why not? Show that we obtain a contradiction and thus conclude that $\sqrt{2} \notin \mathbb{Q}$.

This is an example of the *method of infinite descent*, which is very important in number theory.

2.2.18 You are given the following facts.

(a) All polynomials are continuous.

(b) (Intermediate Value Theorem) If f is continuous on $[a, b]$ and L lies between $f(a)$ and $f(b)$, then $f(x) = L$ for some $x \in (a, b)$.

(c) If $f'(x) > 0$ on an interval, then f is an increasing function.

Use these facts to give a formal proof that $x^{17} + 12x^3 + 13x + 3 = 0$ has *exactly one solution* x, and that x lies in the interval $(-1, 0)$.

2.3 Quantifiers

The proofs we've dealt with thusfar have been fairly straightforward. In higher mathematics, however, there are often definitions and theorems that involve many pieces, and it becomes unwieldy to write everything out in full sentences. Two space-saving devices called *quantifiers* are often used to contract sentences and make the larger structure of a statement clearer.[8] Their use in formal logic is more complex, but for most of mathematics (and certainly this text) all you need is to be able to recognize, understand, and negate them. This last is most important for attempting contrapositive or contradiction proofs.

Definition 2.26. The *universal quantifier* $\forall$ is read 'for all'. The *existential quantifier* $\exists$ is read 'there exists.'

Many sentences in English can be restated using quantifiers:

Examples.

1. Every cloud has a silver lining: $\forall$ clouds, $\exists$ a silver lining.
2. All humans have a brain: $\forall$ humans, $\exists$ a brain.
3. There is an integer smaller than π: $\exists n \in \mathbb{Z}$ such that $n < \pi$.
4. π cannot be written as a ratio of integers: $\forall$ integers m, n, we have $\frac{m}{n} \neq \pi$.

Propositional Functions and Quantified Propositions

Definition 2.27. A *propositional function* is an expression $P(x)$ which depends on a variable x. The collection of allowed variables x is the *domain* of P. For each x, the expression $P(x)$ is a proposition in the usual sense.
The *quantified proposition* $\forall x,\ P(x)$ is an assertion that $P(x)$ is true *for all* values of x. Similarly $\exists x,\ P(x)$ asserts that $P(x)$ is true for *at least one* value of x.

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

$$P(x). \quad x^2 > 4.$$

For this example, $P(5)$ is true, whilst $P(-1)$ is false. More generally, $P(x)$ is true for some values of x (namely $x > 2$ or $x < -2$) and false for others ($-2 \leq x \leq 2$).
In this case the quantified proposition $\forall x \in \mathbb{R},\ P(x)$ is *false*, while $\exists x \in \mathbb{R},\ P(x)$ is true.

Aside: Clarity versus Concision

As we've observed, mathematics is something of an art form and, like with all art, different practitioners have different tastes. Some mathematicians write very concisely, keeping words to a minimum. Some write almost entirely in English. Most use a hybrid of quantifiers and English, aiming for a balance between brevity and clarity. For example, consider the famous *sum of four squares* theorem:

[8] At least that's the idea: very often they are *over-used* and achieve the opposite effect!

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

You will probably agree that the English version is easiest to follow, and the Full Logic the most abstract. However, the English version is less precise: 'sum of four squares' has to be interpreted. The Full Logic expression avoids this by introducing variables and a formula. The Hybrid expression aims for a balance between these extremes. The insertion of a single comma and the phrase 'such that' increases readabilty, while retaining the benefit of precision. Remember that the purpose of writing mathematics is so that someone else can read and understand what you've written *without* you being there to explain it to them. Your presentation style has an enormous effect on whether you are successful!

Similarly, in our previous example, the sentence '$\exists x \in \mathbb{R}$ such that $x^2 > 4$' is more understandable than our original formulation '$\exists x \in \mathbb{R},\ x^2 > 4$.'

Counterexamples and Negating Quantified Propositions

Besides the concision afforded by quantifiers, one of their benefits is a rule that allows for easy negation.

Theorem 2.28. *For any propositional function $P(x)$, we have:*

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

Like with all theorems, to understand it you should unpack it, write it in English, and come up with an example:

1. The negation of 'For all x, $P(x)$ is true' is

   There exists an x such that $P(x)$ is false.'

2. The negation of 'There exists an x such that $P(x)$ is true' is

   For all x, $P(x)$ is false.

Definition 2.29. A *counterexample* to $\forall x,\ P(x)$ is a single element t in the domain of P such that $P(t)$ is false.

Clearly $x = 1$ is a suitable counterexample to $\forall x \in \mathbb{R},\ x^2 > 4$.

Examples. Here are two examples, numbered corresponding to the parts of Theorem 2.28.

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

Somebody does not own a bicycle.

It certainly looks pedantic, but symbolically we might write:

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

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

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

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

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

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

Advice when Negating: Hidden and Excess Quantifiers

Theorem 2.28 seems very simple, but in practice it can be very easy to misuse. Here are some points to consider when negating quantifiers.

1. Don't forget the *meaning* of the sentence. Use the logical rules in Theorem 2.28 but also think it out in words. You should get the same result. Think about the finished sentence and read it aloud: if it *sounds* like the opposite of what you started with then it probably is!

2. The symbol $\nexists$ for 'does not exist' is much abused. Very occasionally its use is appropriate, but it too often demonstrates laziness or a lack of understanding. Avoid using it unless absolutely necessary.

3. Only switch the symbols $\forall$ and $\exists$ if they preceed a *proposition* and are truly used as logical quantifiers. In the following example, 'silver lining' is not a proposition.

   $$\forall \text{ clouds}, \exists \text{ a silver lining.}$$

   When negating, we don't switch $\exists$ to $\forall$. Indeed its negation is

   $$\exists \text{ a cloud without a silver lining.}$$

4. Beware of hidden quantifiers! Sometimes a quantifier is implied but not explicitly stated. This is very common when a statement contains an implication. Consider the following very easy theorem.

   If n is an odd integer, then n^2 is odd. $(*)$

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

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

In this form we can negate by combining the rules in Theorems 2.10 and 2.28. The pattern is

$$\neg\,[\forall n,\ P(n) \implies Q(n)] \iff \exists n,\ P(n) \text{ and } \neg Q(n).$$

The negation of $(*)$ is therefore,

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

The negation of $(*)$ is, of course, false!

Here is a harder example of a hidden quantifier, this time from Linear Algebra.

Definition 2.30. Vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are *linearly independent* if $a\mathbf{x} + b\mathbf{y} + c\mathbf{z} = \mathbf{0} \implies a = b = c = 0$.

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

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

To negate the definition, we must also negate the hidden quantifier:

Vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are *linearly dependent* if $\exists a, b, c$ not all zero such that $a\mathbf{x} + b\mathbf{y} + c\mathbf{z} = \mathbf{0}$.

The final challenge is recalling how to negate an implication: recall Theorem 2.10, and note that the negation of $a = b = c = 0$ is that *at least one* of a, b, c is non-zero.

Multiple quantifiers

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

Example. Show that the following statement is false.

$\forall x \in \mathbb{R},\ \exists y \in \mathbb{R}$ such that $xy = 3$.

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

$\exists x \in \mathbb{R}$ such that $\forall y \in \mathbb{R}$ we have $xy \neq 3$.

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

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

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

Putting it all together: Continuity

The definition of continuity from calculus combines multiple quantifiers, a hidden quantifier and an implication. The purpose of this text isn't to teach you the subtleties of what the following definition means, that's for a later Analysis class. We simply want to be able to read and negate such expressions.

Definition 2.31. Suppose that f is a function whose domain and codomain are sets of real numbers. We say that f is *continuous* at $x = a$ if,

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

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

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

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

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

Warning! The negation of $\forall \varepsilon > 0$ is *not* $\exists \varepsilon \leq 0$. Only the ultimate proposition[9] is negated!
For an example of this definition in use, see the exercises.

The Order of Quantifiers Matters!

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

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

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

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

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

1. Everyone has a friend.
2. There exists somebody who is friend with everybody.

Quite different!

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

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

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

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

[9] In this case $|x - a| < \delta \implies |f(x) - f(a)| < \varepsilon$.

Exercises

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

(a) All mathematics exams are hard.

(b) No football players are from San Diego.

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

2.3.2 Let P be the proposition: 'Every positive integer is divisible by thirteen.'

(a) Write P using quantifiers.

(b) What is the negation of P?

(c) Is P true or false? Prove your assertion.

2.3.3 Prove or disprove: There exist integers m and n such that $2m - 3n = 15$.

2.3.4 Prove or disprove: There exist integers m and n such that $6m - 3n = 11$.
Hint: The left-hand side is always divisible by...

2.3.5 Prove that between any two distinct rational numbers there exists another rational number.

2.3.6 Let p be an odd integer. Prove that $x^2 - x - p = 0$ has no *integer* solutions.

2.3.7 Prove: For every positive integer n, $n^2 + n + 3$ is an odd integer greater than or equal to 5.
There are two claims here: $n^2 + n + 3$ is odd, and $n^2 + n + 3 \geq 5$.

2.3.8 Consider the propositional function

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

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

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

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

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

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

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

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

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

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

2.3.10 Which of the following statements are true? Explain.

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

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

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

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

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

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

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

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

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

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

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

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

2.3.13 Prove or disprove each of 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.

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

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

Consider the following proposition:

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

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

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

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

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

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

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

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

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

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

(e) Prove that $x_n = \frac{1}{n}$ converges to zero.

2.3.16 This question uses Definition 2.31. You will likely find this difficult.

(a) Prove, directly from the definition, that $f(x) = x^2$ is continuous at $x = 0$. *If you are given $\epsilon > 0$, what should δ be?*

(b) Prove that $g(x) = \begin{cases} 1+x & \text{if } x \geq 0, \\ x & \text{if } x < 0, \end{cases}$ is discontinuous at $x = 0$.

(c) (Very hard) Let $h(x) = \begin{cases} x & \text{if } x \text{ is rational,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$ Prove that f is continuous *only* at $x = 0$.

2.3.17 In this question we prove Rolle's Theorem from calculus:
If f is continuous on $[a, b]$, differentiable on (a, b), and $f(a) = f(b) = 0$, then $\exists c \in (a, b)$ such that $f'(c) = 0$.
As you work through the question, think about where the hypotheses are used and why we need them.

(a) Recall the Extreme Value Theorem. The function f is continuous on $[a, b]$, so f is bounded and attains its bounds. Otherwise said,

$$\exists m, M \in [a, b] \text{ such that } \forall x \in [a, b] \text{ we have } f(m) \leq f(x) \leq f(M).$$

Suppose that $f(m) = f(M)$. Why is the conclusion of Rolle's Theorem obvious in this case?

(b) Now suppose that $f(m) \neq f(M)$. Argue that *at least one* of the following cases holds:

$$f(M) > 0 \quad \text{or} \quad f(m) < 0.$$

(c) Without loss of generality, we may assume that $f(M) > 0$. By considering the function $-f$, explain why.

(d) Assume $f(M) > 0$. Then $M \neq a$ and $M \neq b$. Consider the difference quotient,

$$\frac{f(M+h) - f(M)}{h}.$$

Show that if $0 < |h| < \min\{M - a, b - M\}$ then the difference quotient is well-defined (exists and makes sense).

(e) Suppose that $0 < h < b - M$. Show that

$$\frac{f(M+h) - f(M)}{h} \leq 0.$$

How do we know that $L^+ := \lim_{h \to 0^+} \frac{f(M+h)-f(M)}{h}$ exists? What can you conclude about L^+?

(f) Repeat part (d) for $L^- := \lim_{h \to 0^-} \frac{f(M+h)-f(M)}{h}$.

(g) Conclude that $L^+ = L^- = 0$. Why have we completed the proof?

3 Divisibility and the Euclidean Algorithm

In this section we introduce the notion of *congruence*: a generalization of the idea of separating all integers into 'even' and 'odd.' At its most basic it involves going back to elementary school when you first learned division and would write something similar to

$33 \div 5 = 6\,\mathrm{r}\,3$ '6 remainder 3.'

The study of congruence is of fundamental importance to Number Theory, and provides some of the most straightforward examples of Groups and Rings. We will cover the basics in this section—enough to compute with—then return later for more formal observations.

3.1 Remainders and Congruence

Definition 3.1. Let m and n be integers. We say that n *divides* m and write $n \mid m$ if m is divisible by n: that is if there exists some integer k such that $m = kn$. Equivalently, we say that n is a *divisor* of m, or that m is a *multiple* of n.

For example: $4 \mid 20$ and $17 \mid 51$, but $12 \nmid 8$.

When one integer does not divide another, there is a remainder left over.

Theorem 3.2 (The Division Algorithm)**.** *Let m be an integer and n a positive integer. Then there exist unique integers q (the* quotient*) and r (the* remainder*) which satisfy the following conditions:*

1. $0 \leq r < n$.
2. $m = qn + r$.

For example: If $m = 23$ and $n = 7$, then $q = 3$ and $r = 2$ because '$23 \div 7 = 3$ remainder 2.' More formally, $23 = 3 \cdot 7 + 2$, with $0 \leq 2 < 7$. Similarly, if $m = -11$ and $n = 3$, then $q = -4$ and $r = 1$ because $-11 = (-4) \cdot 3 + 1$, with $0 \leq 1 < 3$.

For practice, find a formula for all the integers that have remainder 4 *after division by* 6.

The proof of the Division Algorithm relies on the development of induction, to which we will return in Chapter 5. The theorem should be read as saying that n goes q times into m with r left over. The fact that the remainder is nicely defined allows us to construct an alternative form of arithmetic.

Definition 3.3. Let a, b be integers, and n a positive integer. We say that *a is congruent to b modulo n* and write $a \equiv b \pmod{n}$ if a and b have the same remainder upon dividing by n.
When the *modulus* n is clear, it tends to be dropped, and we just write $a \equiv b$.

For example: $7 \equiv 10 \pmod{3}$, since both have the same remainder (1) on dividing by 3. Can you find a formula for *all* the integers that are congruent to 10 modulo 3?

Let a be an integer. Consider the following conjectures. Are they true or false?

Conjecture 3.4. $a \equiv 8 \pmod 6 \implies a \equiv 2 \pmod 3$.

Conjecture 3.5. $a \equiv 2 \pmod 3 \implies a \equiv 8 \pmod 6$.

The first conjecture is true. Indeed, if $a \equiv 8 \pmod 6$, we can write $a = 6k + 8$ for some integer k. Then $a = 6k + 8 = 6k + 6 + 2 = 3(2k + 2) + 2$ so a has remainder 2 upon division by 3, showing that a is congruent to 2 modulo 3.
On the other hand, the second conjecture is false. All we need is a counterexample. Consider $a = 5$: clearly a is congruent to 2 modulo 3, but a is not congruent to 8 modulo 6 (because it has remainder 5, not 2, upon division by 6).

The following theorem is crucial, and provides an equivalent definition of congruence.

Theorem 3.6. $a \equiv b \pmod n \iff n \mid (b - a)$.

Proof. There are two separate theorems here, although both rely on the Division Algorithm (Theorem 3.2) to divide both a and b by n. Given a, b, n, the Division Algorithm shows that there exist unique quotients q_1, q_2 and remainders r_1, r_2 which satisfy

$$a = q_1 n + r_1, \qquad b = q_2 n + r_2, \qquad 0 \le r_1, r_2 < n. \tag{$*$}$$

Now we perform both directions of the proof.

($\Rightarrow$) Suppose that $a \equiv b \pmod n$. By definition, this means that a and b have the same remainder when divided by n. That is, $r_1 = r_2$. Now subtracting a from b gives us

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

which is divisible by n. Therefore $n \mid (b - a)$.

($\Leftarrow$) This direction is a little more subtle. We assume that $b - a$ is divisible by n. Thus $b - a = kn$ for some integer k. According to $(*)$, this implies that

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

is also a multiple of n. Now consider the condition on the remainders in $(*)$: since $0 \le r_1, r_2 < n$, we quickly see that

$$\begin{cases} 0 \le r_2 < n \\ -n < -r_1 \le 0 \end{cases} \implies -n < r_2 - r_1 < n.$$

This says that $r_2 - r_1$ is a multiple of n lying strictly between $\pm n$. The only possibility is that $r_2 - r_1 = 0$. Otherwise said, $r_2 = r_1$, whence a and b have the same remainder, and so $a \equiv b \pmod n$. ∎

If you are having some trouble with the final step, think about an example. Suppose that $n = 26$ and that and that $r_2 - r_1$ is an *integer* satisfying the inequalities $-26 < r_2 - r_1 < 26$. It should be obvious that $r_2 - r_1 = 0$.

To gain some familiarity with congruence, use Theorem 3.6 to show that

$$a \equiv b \pmod{n} \iff b \equiv a \pmod{n}.$$

Note that both this expression and the theorem contain a hidden quantifier, as discussed in Section 2.3. Morover, combining the theorem with Definition 3.1 leads to the observation that

$$a \equiv b \pmod{n} \iff \exists k \in \mathbb{Z} \text{ such that } b - a = kn,$$

that is, $b = a + kn$.

Congruence and Divisibility

The previous two theorems may appear a little abstract, so it's a good idea to recap the relationship between congruence and divisibility. The following observations should be immediate to you!

Let a be any integer and let n be a positive integer. Then

- a is congruent to either $0, 1, 2, \ldots$ or $n - 1$ modulo n.
- a is divisible by n if and only if $a \equiv 0 \pmod{n}$.
- a is *not* divisible by n if and only if $a \equiv 1,\ 2,\ 3, \ldots,\ n - 1 \pmod{n}$.

To test your level of comfort with the definition of congruence, and review some proof techniques, prove the following theorem.

Theorem 3.7. *Suppose that n is an integer. Then*

$$n^2 \not\equiv n \pmod{3} \iff n \equiv 2 \pmod{3}.$$

If you don't know how to start, try completing the following table:

n	n^2	$n^2 \equiv n \pmod{3}$
0	0	T
1		
2		

Now try to write a formal proof.

That the congruence sign $\equiv$ appears similar to the equals sign $=$ is no accident. In many ways it behaves exactly the same. In Chapter 7.3 we shall see that congruence is an important example of an *equivalence relation.*

Modular Arithmetic

The arithmetic of remainders is almost exactly the same as the more familiar arithmetic of real numbers, but comes with all manner of fun additional applications, most importantly cryptography and data security: your cell-phone and computer perform millions of these calculations every day! Here we spell out the basic rules of congruence arithmetic.[10]

Theorem 3.8. *Suppose throughout that a, b, c, d are integers, and that all congruences are modulo the same integer n.*

1. $a \equiv b$ *and* $c \equiv d \implies ac \equiv bd$
2. $a \equiv b$ *and* $c \equiv d \implies a \pm c \equiv b \pm d$

What the theorem says is that the operations of 'take the remainder' and 'add' (or 'multiply') can be performed in either order; the result will be the same. For example, consider $a = 29$, $b = 14$ and $n = 6$. We can add a and b then take the remainder when dividing by n:

$$29 + 14 = 43 = 6 \cdot 7 + 1.$$

Instead we could first take the remainders of a and b modulo 6 and then add these:

$$5 + 2 = 7, \quad \text{which has the } \textit{same remainder } 1.$$

Either way, we may write the result as a congruence,

$$29 + 14 \equiv 1 \pmod 6.$$

Proof of Theorem 3.8. Suppose that $a \equiv b$ and $c \equiv d$. By Theorem 3.6 we have $a - b = kn$ and $c - d = ln$ for some integers k, l. Thus

$$ac = (b + kn)(d + ln) = bd + n(bl + kd + kln) \Rightarrow ac - bd = n(bl + kd + kln)$$

is divisible by n. Hence $ac \equiv bd$.
Try the second argument yourself. ∎

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

Examples. 1. What is the remainder when 39^{23} is divided by 10? At the outset this appears impossible. Ask your calculator and it will tell you that $39^{23} \approx 3.93 \times 10^{36}$, which is of no help at all! Instead think about the rules of arithmetic modulo 10. Since $39 \equiv 9 \equiv -1 \pmod{10}$, we quickly notice that

$$39 \cdot 39 \equiv (-1) \cdot (-1) \equiv 1 \pmod{10},$$

[10] The usual associative, commutative and distributive laws of arithmetic

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

all follow because $x = y \implies x \equiv y \pmod n$, regardless of n: equal numbers have the same remainder after all!

whence $39^2 \equiv 1 \pmod{10}$. Since positive integer exponents signify repeated multiplication, we can repeat the exercise to obtain

$$39^{23} \equiv (-1)^{23} \equiv -1 \equiv 9 \pmod{10}.$$

Therefore 39^{23} has remainder 9 when divided by 10. Otherwise said, the last digit of 39^{23} is a 9. If you ask a computer for all the digits you can check this yourself.

2. Now that we understand powers, more complex examples become easy. Here we compute modulo $n = 6$.

$$7^9 + 14^3 \equiv 1^9 + 2^3 \equiv 1 + 8 \equiv 9 \equiv 3 \pmod{6}.$$

Hence $7^9 + 14^3 = 40356351$ has remainder 3 when divided by 6.

3. Find the remainder when $124^{12} \cdot 65^{49}$ is divided by 11. This time we'll need to perform multiple calculations to keep reducing the base to something managable. Since $124 = 11^2 + 3$ and $65 = 11 \cdot 6 - 1$, we write

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

The remainder is therefore 2. There is no way to do this on a pocket calculator, since the original number $124^{12} \cdot 65^{49} \approx 9 \times 10^{113}$ is far too large to work with!

The primary difference between modular and normal arithmetic is, perhaps unsurprisingly, with regard to *division.*

Theorem 3.9. *If $ka \equiv kb \pmod{kn}$ then $a \equiv b \pmod{n}$.*

The modulus is divided by k as well as the terms, so the meaning of $\equiv$ changes. In Exercise 3.1.6 you will prove this theorem, and observe that, in general, we do not expect $a \equiv b \pmod{n}$.

Exercises

3.1.1 Find the remainder when $17^{251} \cdot 23^{12} - 19^{41}$ is divided by 5. *Hint:* $17 \equiv 2$ *and* $2^2 \equiv -1 \pmod{5}$.

3.1.2 Is the statement

$$n^2 \equiv n \pmod{3} \iff n \equiv 0 \pmod{3} \text{ or } n \equiv 1 \pmod{3},$$

identical to Theorem 3.7? Why/why not?

3.1.3 Prove that if $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$ then $3a - c^2 \equiv 3b - d^2 \pmod{n}$.

3.1.4 Find a natural number n and integers a, b such that $a^2 \equiv b^2 \pmod{n}$ but $a \not\equiv b \pmod{n}$.

3.1.5 Let p be a prime number greater than or equal to 3. Show that if $p \equiv 1 \pmod{3}$, then $p \equiv 1 \pmod{6}$. *Hint: p is odd.*

3.1.6 Suppose that $7x \equiv 28 \pmod{42}$. By Theorem 3.9, it follows that $x \equiv 4 \pmod{6}$.

(a) Check this explicitly using Theorem 3.6.

(b) If $7x \equiv 28 \pmod{42}$, is it possible that $x \equiv 4 \pmod{42}$?

(c) Is it always the case that $7x \equiv 28 \pmod{42} \implies x \equiv 4 \pmod{42}$? Why/why not?

(d) Prove Theorem 3.9.

3.1.7 If $a \mid b$ and $b \mid c$, prove that $a \mid c$.

3.1.8 Let a, b be positive integers. Prove that $a = b \iff a \mid b$ and $b \mid a$.

3.1.9 Here are two conjectures:

Conjecture 1 $a \mid b$ and $a \mid c \implies a \mid bc$.

Conjecture 2 $a \mid c$ and $b \mid c \implies ab \mid c$.

Decide whether each conjecture is true or false and prove/disprove your assertions.

3.1.10 Fermat's Little Theorem (to distinguish it from his 'Last') states that if p is prime and $a \not\equiv 0 \mod p$, then $a^{p-1} \equiv 1 \pmod{p}$.

(a) Use Fermat's Little Theorem to prove that $b^p \equiv b \pmod{p}$ for *any* integer b.

(b) Prove that if p is prime then $p \mid (2^p - 2)$.

(c) Prove that the converse is not true, that $2^n - 2$ being divisible by n does not imply that n is prime (take $n = 341$...).

3.2 Greatest Common Divisors and the Euclidean Algorithm

At its most basic, Number Theory involves finding *integer* solutions to equations. Here are two simple-sounding questions:

1. The equation $9x - 21y = 6$ represents a straight line. Are there any *integer points* on this line? That is, can you find integers x, y satisfying $9x - 21y = 6$?
2. What about on the line $4x + 6y = 1$?

Before you do anything else, try sketching both lines (lined graph paper will help) and try to decide if there are any integer points. If there are any, how many are there? Can you find them all?

In this section we will see how to answer these questions in general: for which lines $ax + by = c$ with $a, b, c \in \mathbb{Z}$, are there integer solutions, and how can we find them all? The method introduces the appropriately named *Euclidean algorithm,* a famous procedure dating at least as far back as Euclid's *Elements* (c. 300 BC.).

Definition 3.10. Let m, n be integers, not both zero. Their *greatest common divisor* $\gcd(m, n)$ is the largest (positive) divisor of both m and n. We say that m, n are *relatively prime* if $\gcd(m, n) = 1$.

Example. Let $m = 60$ and $n = 90$. The positive divisors of the two integers are listed in the table:

m	1	2	3	4	5	6	10	12	15	20	$\underline{30}$	60
n	1	2	3	5	6	9	10	15	18	$\underline{30}$	45	90

The greatest common divisor is the largest number common to both rows: clearly $\gcd(60, 90) = 30$.

Finding the greatest common divisor by listing all the positive divisors of a number is extremely tedious. This is where Euclid rides to the rescue.

Euclidean Algorithm. To find $\gcd(m, n)$ for two positive integers $m > n$:

(i) Use the division algorithm (Theorem 3.2) to write $m = q_1 n + r_1$ with $0 \leq r_1 < n$.

(ii) If $r_1 = 0$, set $\gcd(m, n) = n$. Otherwise,
If $r_1 > 0$, apply again: divide n by r_1 to obtain $n = q_2 r_1 + r_2$ with $0 \leq r_2 < r_1$.

(iii) If $r_2 = 0$, set $\gcd(m, n) = r_1$. Otherwise,
If $r_2 > 0$, apply again: divide r_1 by r_2 to obtain $r_1 = q_3 r_2 + r_3$ with $0 \leq r_3 < r_2$.

(iv) If $r_2 = 0$, set $\gcd(m, n) = r_1$. Otherwise,
Repeat the process: obtain a decreasing sequence of positive integers

$$r_1 > r_2 > r_3 > \ldots > 0$$

Theorem 3.11. *The Algorithm eventually produces a remainder of zero: $\exists r_{p+1} = 0$. The greatest common divisor of m, n is the last non-zero remainder:* $\gcd(m, n) = r_p$.

The proof is in the exercises. If m, n are not both positive, take absolute values first and apply the algorithm. For instance $\gcd(-6, 45) = 3$.

Example. Compute $\gcd(1260, 750)$ using the Euclidean algorithm: the steps are labeled as in the original algorithm. You might instead find it easier to create a table with and observe each remainder moving diagonally down and left at each successive step.

	m	n	q	r
(i) $1260 = 1 \times 750 + 510$	1260	750	1	510
(ii) $750 = 1 \times 510 + 240$	750	510	1	240
(iii) $510 = 2 \times 240 + 30$	510	240	2	30
(iv) $240 = 8 \times 30 + 0$	240	30	8	0

Theorem 3.11 says that $\gcd(1260, 750) = 30$, the last non-zero remainder.

As you can see, the Euclidean algorithm is very efficient.

Reversing the Algorithm: Integer Points on Lines

To apply the Euclidean algorithm to finding integer points on lines, we must turn it on its head. By starting with the second last line of the algorithm and substituing the previous lines one at a time, we can find integers x, y such that $\gcd(m, n) = mx + ny$. This is easiest to demonstrate by continuing our previous example:

Example (continued)**.** Find integers x, y such that $1260x + 750y = 30$.

Solve for 30 (the gcd of 1260 and 750) using step (iii), to get

$$30 = 510 - 2 \times 240.$$

Now use the equation in step (ii) to solve for 240 and substitute:

$$30 = 510 - 2 \times (750 - 510) = 3 \times 510 - 2 \times 750.$$

Finally, substitute for 510 using equation (i):

$$30 = 3 \times (1260 - 750) - 2 \times 750 = 3 \times 1260 - 5 \times 750.$$

We have expressed 30 as a linear combination of 1260 and 750, as desired. Reading off the coefficients of the combination, we get $x = 3$ and $y = -5$ therefore satisfy $1260x + 750y = 30$.
Note how the process to find x and y is twofold: first we find $\gcd(m, n)$ using the Euclidean Algorithm, then we do a series of back substitutions to recover x and y.

More generally, we have the following corollary.

Corollary 3.12. *Given any integers m, n there exist integers x, y such that* $\gcd(m, n) = mx + ny$.

We are now in a position to solve our motivating problem: finding all integer points on the line $ax + by = c$ where a, b, c are integers.

Theorem 3.13. *Let a, b, c be integers and $d = \gcd(a, b)$. Then the equation $ax + by = c$ has an integer solution (x, y) iff $d \mid c$. In such a case, all integer solutions are given by*

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

where (x_0, y_0) is any fixed integer solution, and t takes any integer value.

One uses the Euclidean Algorithm to find the initial solution (x_0, y_0), then applies $(*)$ to obatin all of them.[11] The proof is again in the exercises.

Examples. 1. Find all integer solutions to the equation $1260x + 750y = 90$.
We calculated earlier that $\gcd(1260, 750) = 30$. Thus $d = 30$. Since $d \mid c$ (that is, $30 \mid 90$), we know that there are integer solutions. We also calculated that

$$1260 \times 3 + 750 \times (-5) = 30.$$

Since we want 90, we simply multiply our pair $(3, -5)$ by three:

$$1260 \times 9 + 750 \times (-15) = 90.$$

whence $(x_0, y_0) = (9, -15)$ is an integer solution to the equation. The general solution is therefore

$$(x, y) = \left(9 + \frac{750}{30}t, -15 - \frac{1260}{30}t\right) = (9 + 25t, -15 - 42t), \text{ where } t \in \mathbb{Z}.$$

2. No consider the line $570x + 123y = 7$. We calculate the greatest common divisor using the Euclidean algorithm:

$$\left.\begin{aligned} 570 &= 4 \times 123 + 78 \\ 123 &= 1 \times 78 + 45 \\ 78 &= 1 \times 45 + 33 \\ 45 &= 1 \times 33 + 12 \\ 33 &= 2 \times 12 + 9 \\ 12 &= 1 \times 9 + 3 \\ 9 &= 4 \times 3 + 0 \end{aligned}\right\} \implies \gcd(570, 123) = 3.$$

Since $3 \nmid 7$, we conclude that the line $570x + 123y = 7$ has no integer points.

3. Repeat the above calculations for our motivating problems: what does the theorem say?

[11] The astute observer should recognize the similarity between this and the complementary function/particular integral method for linear differential equations: (x_0, y_0) is a 'particular solution' to the full equation $ax + by = c$, while $(\frac{b}{d}t, -\frac{a}{d}t)$ comprises all solutions to the 'homogeneous equation' $ax + by = 0$.

Exercises

3.2.1 Use the Euclidean Algorithm to compute the greatest common divisors indicated.

(a) $\gcd(20, 12)$ (b) $\gcd(100, 36)$ (c) $\gcd(207, 496)$

3.2.2 For each part of Question 3.2.1, find integers x, y for which $\gcd(m, n) = mx + ny$.

3.2.3 (a) Answer our motivating problems using the above process.

(i) Find all integer points on the line $9x - 21y = 6$.

(ii) Show that there are no integer points on the line $4x + 6y = 1$.

(b) Can you give an elementary proof as to why there are no integer points on the line $4x + 6y = 1$?

3.2.4 Find all the integer points on the following lines, or show that none exist.

(a) $16x - 33y = 2$.

(b) $122x + 36y = 3$.

(c) $324x - 204y = -12$.

3.2.5 Find all possible solutions to the motivating problem at the start of the notes: Five people each take the same number of candies from a jar. Then a group of seven does the same. The, now empty, jar originally contained 239 candies. How much candy did each person take?

3.2.6 Show that there exists no integer x such that $3x \equiv 5 \pmod 6$.

3.2.7 In Theorem 3.11 we claim that the Euclidean algorithm terminates with $r_{p+1} = 0$. Why? Show that the number of steps p is no more than n. *The algorithm is much faster than this in practice!*

3.2.8 Let $m = qn + r$ be the result of the division algorithm for integers m, n.

(a) Let d be a common positive divisor of m, n. Prove that $d \mid r$.

(b) Now suppose that c is a common divisor of n and r. Prove that $c \mid m$.

(c) Explain why parts (a) and (b) prove that $\gcd(m, n) = \gcd(n, r)$.

(d) Conclude that the final remainder r_p in the Euclidean algorithm really is $\gcd(m, n)$.

3.2.9 Prove the following:

$$\gcd(m, n) = 1 \iff \exists x, y \in \mathbb{Z} \text{ such that } mx + ny = 1.$$

One direction can be done by applying Corollary 3.12, but the other direction requires an argument.

3.2.10 In this question we prove the Theorem on integer solutions to linear equations. Let $a, b, c \in \mathbb{Z}$. Suppose that (x_0, y_0) and (x_1, y_1) are two integer solutions to the linear Diophantine equation $ax + by = c$.

(a) Show that $(x_0 - x_1, y_0 - y_1)$ satisfies the equation $ax + by = 0$.

(b) Suppose that $\gcd(a, b) = d$. Prove that $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$. (*Use Question 3.2.9)*

(c) Find all integer solutions (x, y) to $ax + by = 0$ (*Don't use the Theorem, it's what you're trying to prove! Think about part (b) and divide through by d first.*).

(d) Use (a) and (b) to conclude that (x, y) is an integer solution to $ax + by = c$ if and only if

$$x = x_0 + \frac{b}{d}t \qquad y = y_0 - \frac{a}{d}t, \qquad \text{where } t \in \mathbb{Z}.$$

3.2.11 Show that $\gcd(5n + 2, 12n + 5) = 1$ for every integer n. *There are two ways to approach this: you can try to use the Euclidean algorithm abstractly, or you can use the result of Exercise 3.2.9.*

3.2.12 The set of remainders $\mathbb{Z}_n = \{0, 1, 2, \ldots, n - 1\}$ is called a *ring* when equipped with addition and multiplication modulo n. For example $5 + 6 \equiv 3 \pmod 8$. We say that $b \in \mathbb{Z}_n$ is an *inverse* of $a \in \mathbb{Z}_n$ if

$$ab \equiv 1 \pmod n.$$

(a) Show that 2 has no inverse modulo 6.

(b) Show that if $n = n_1 n_2$ is composite ($\exists$ integers $n_1, n_2 \geq 2$) then there exist elements of the ring $\mathbb{Z}_n$ which have no inverses.

(c) Prove that a has an inverse modulo n if and only if $\gcd(a, n) = 1$. Conclude that the only sets $\mathbb{Z}_n$ for which all non-zero elements have inverses are those for which n is prime. *You will find Exercise 3.2.9 helpful.*

4 Sets and Functions

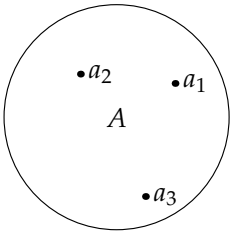
Sets are the fundamental building blocks of mathematics. In the sub-discipline of Set Theory, mathematicians define all basic notions, including number, addition, function, etc., purely in terms of sets. In such a system it can take over 100 pages of discussion to *prove* that $1 + 1 = 2$! We will not be anything like so rigorous. Indeed, before one can accept that such formality has its place in mathematics, a level of familiarity with sets and their basic operations is necessary.

4.1 Set Notation and Describing a Set

We start with a very naïve notion: a set is a collection of objects.[12]

Definition 4.1. If x is an object in a set A, we write $x \in A$ and say that x is an *element* or *member* of A. On the other hand, if x is a member of some other set B, but not of A, we write $x \notin A$.
Two sets are described as *equal* if they have exactly the same elements.

When thinking abstractly about sets, you may find *Venn diagrams* useful. A set is visualized as a region in the plane and, if necessary, members of the set can be thought of as dots in this region. This is most useful when one has to think about multiple, possibly over-lapping, sets. The graphic here represents a set A with at least three elements a_1, a_2, a_3.



Notation and Conventions

Use capital letters for sets, e.g. A, B, C, S, and lower-case letters for elements. It is conventional, though not required, to denote an abstract element of a set by the corresponding lower-case letter: thus $a \in A$, $b \in B$, etc.
Curly brackets $\{\ ,\ \}$ are used to bookend the elements of a set: for instance, if we wrote

$$S = \{3, 5, f, \alpha, \beta\}$$

then we'd say, 'S is the set whose elements are 3, 5, f, α and β.'
The order in which we list the elements in a set is irrelevant, thus

$$S = \{\beta, f, 5, \alpha, 3\} = \{f, \alpha, 3, \beta, 5\}.$$

Listing the elements in a set in this way is often known as *roster notation*.

By contrast, *set-builder notation* describes the elements of a set by starting with a larger set and restricting to those elements which satisfy some property. The symbols | or : are used as a short-hand for 'such that.' Which symbol you use depends partly on taste, although the context may make one clearer to read.[13] For example, if $S = \{3, 5, f, \alpha, \beta\}$ is the set defined above, we could write,

$$\{s \in S : s \text{ is a Greek letter}\} = \{\alpha, \beta\}$$

[12] Much thinking was required before mathematicians realized that this is indeed naïve. It eventually became clear that some collections of objects cannot be considered sets, and the search for a completely rigorous definition began. Thus was Axiomatic Set Theory born. For the present, our notion is enough.

[13] See Choice of Notation, below.

or

$$\{s \in S \mid s \text{ is a Greek letter}\} = \{\alpha, \beta\}.$$

We would read: 'The set of elements s in the set S such that s is a Greek letter is $\{\alpha, \beta\}$.'

Example. Let $A = \{2, 4, 6\}$ and $B = \{1, 2, 5, 6\}$. There are many options for how to write A and B in set-builder notation. For example, we could write

$$A = \{2n : n = 1, 2 \text{ or } 3\} \quad \text{and} \quad B = \{n \in \mathbb{Z} \mid 1 \leq n \leq 6 \text{ and } n \neq 3, 4\}.$$

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

$$S_1 = \{a \in A : a \text{ is divisible by } 4\} = \{4\}$$

$$S_2 = \{b \in B : b \text{ is odd}\} = \{1, 5\}$$

$$S_3 = \{a \in A \mid a \in B\} = \{2, 6\}$$

$$S_4 = \{a \in A : a \notin B\} = \{4\}$$

$$S_5 = \{b \in B \mid b \text{ is odd and } b - 1 \in A\} = \{5\}$$

Take your time getting used to this notation. Can you find an alternative description in set-builder notation of the sets $S_1, \ldots, S_5$ above? It is *crucial* that you can translate between set notations and English, or you will be incapable of understanding most higher-level mathematics.

Sets of Numbers

Common sets of numbers are written in the BLACKBOARD BOLD typeface.

$\mathbb{N} = \mathbb{Z}^+ =$ natural numbers $= \{1, 2, 3, 4, \ldots\}$
$\mathbb{N}_0 = \mathbb{W} = \mathbb{Z}_0^+ =$ whole numbers $= \{0, 1, 2, 3, 4, \ldots\}$
$\mathbb{Z} =$ integers $= \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$
$\mathbb{Q} =$ rational numbers $= \{\frac{m}{n} : m \in \mathbb{Z} \text{ and } n \in \mathbb{N}\} = \{\frac{a}{b} : a, b \in \mathbb{Z} \text{ and } b \neq 0\}$
$\mathbb{R} =$ real numbers
$\mathbb{R} \setminus \mathbb{Q} =$ irrational numbers (read '$\mathbb{R}$ minus $\mathbb{Q}$')
$\mathbb{C} =$ complex numbers $= \{x + iy : x, y \in \mathbb{R}, \text{ where } i = \sqrt{-1}\}$
$\mathbb{Z}_{\geq n} =$ Integers $\geq n = \{n, n+1, n+2, n+3, \ldots\}$
$n\mathbb{Z} =$ multiples of $n = \{\ldots, -3n, -2n, -n, 0, n, 2n, 3n, \ldots\}$

Where there are multiple choices of notation, we will tend to use the first in the list: for example $\mathbb{N}_0 = \mathbb{Z}_{\geq 0}$. The use of a subscript 0 to include zero and superscript $\pm$ to restrict to positive or negative numbers is standard.

Examples. $7 \in \mathbb{Z}$, $\pi \in \mathbb{R}$, $\pi \notin \mathbb{Q}$, $\sqrt{-5} \in \mathbb{C}$, $-e^2 \in \mathbb{R}^-$.

There are often many different ways to represent the same set in set-builder notation. For example, the set of even numbers may be written in multiple ways:

$$\begin{aligned}
2\mathbb{Z} &= \{2n : n \in \mathbb{Z}\} && \text{(The set of numbers of the form } 2n \text{ such that } n \text{ is an integer)}\\
&= \{n \in \mathbb{Z} : \exists k \in \mathbb{Z},\ n = 2k\} && \text{(The set of integers which are a multiple of 2)}\\
&= \{n \in \mathbb{Z} : n \equiv 0 \pmod 2\} && \text{(The set of integers congruent to 0 modulo 2)}\\
&= \{n \in \mathbb{Z} :\ 2 \mid n\} && \text{(The set of integers which are divisible by 2)}
\end{aligned}$$

Here we use both congruence and divisor notation to obtain suitable descriptions. Can you find any other ways to describe the even numbers using basic set notation?

The notation $n\mathbb{Z}$ is most commonly used when n is a natural number, but it can also be used for other n. For example

$$\tfrac{1}{2}\mathbb{Z} = \left\{\tfrac{1}{2}x : x \in \mathbb{Z}\right\} = \left\{m, m + \tfrac{1}{2} : m \in \mathbb{Z}\right\}$$

is the set of multiples of $\frac{1}{2}$ (comprising the integers and half-integers). The notation can also be extended: for example $2\mathbb{Z} + 1$ would denote the odd integers.

Aside: Choice of Notation

The two notations for 'such that' ($\mid$ and :) are to give you leeway in case of potential confusion. For example, the final expression (above) for the even numbers $2\mathbb{Z} = \{n \in \mathbb{Z} :\ 2 \mid n\}$ is much cleaner than the alternative

$$2\mathbb{Z} = \{n \in \mathbb{Z} \mid 2 \mid n\}.$$

In other situations the opposite is true. In Section 4.4 we shall consider functions. If you recall the concept of an odd function from calculus, we could denote the set of such with domain the real numbers as

$$\{f : \mathbb{R} \to \mathbb{R} : \forall x,\ f(x) = f(-x)\} \qquad \text{or} \qquad \{f : \mathbb{R} \to \mathbb{R} \mid \forall x,\ f(x) = f(-x)\}.$$

In this case the latter notation is superior. You may use whichever notation you prefer, provided the outcome is unambiguous.

Examples. 1. List the elements of the set $A = \{x \in \mathbb{R} : x^2 + 3x + 2 = 0\}$.
We are looking for the set of all real number solutions to the quadratic equation $x^2 + 3x + 2 = 0$. A simple factorization tells us that $x^2 + 3x + 2 = (x + 1)(x + 2)$, whence $A = \{-1, -2\}$.

2. Use the set $B = \{0, 1, 2, 3, \ldots, 24\}$ to describe $C = \{n \in \mathbb{Z} : n^2 - 3 \in B\}$ in roster notation.
We see that

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

Since n must be an integer, it follows that

$$C = \{\pm 2, \pm 3, \pm 4, \pm 5\}.$$

3. It is often harder to convert from roster to set-builder notation, as you might be required to spot a pattern, and many choice could be available. For example, if

$$D = \left\{ \frac{1}{6}, \frac{1}{20}, \frac{1}{42}, \frac{1}{72}, \frac{1}{110}, \frac{1}{156}, \dots \right\},$$

you might consider it reasonable to write

$$D = \left\{ \frac{1}{2n(2n+1)} : n \in \mathbb{N} \right\}.$$

Of course the ellipses $(\dots)$ might not indicate that the elements of the set continue in the way you expect. For larger sets, the concision and clarity of set-builder notation makes it much preferred!

4. Are the following sets equal?

$$E = \{n^2 + 2 : n \text{ is an odd integer}\}, \qquad F = \{n \in \mathbb{Z} : n^2 + 2 \text{ is an odd integer}\}.$$

It will help to first construct a table to list some of the values of $n^2 + 2$:

n	n^2	n^2+2
± 1	1	3
± 3	9	11
± 5	25	27
± 7	49	51
± 9	81	83
$\vdots$	$\vdots$	$\vdots$

The set E consists of those integers of the form $n^2 + 2$ where n is an odd integer. By the table,

$$E = \{3, 11, 27, 51, 83, \dots\}.$$

On the other hand, F includes all those integers n such that $n^2 + 2$ is odd. It is easy to see that

$$n^2 + 2 \text{ is odd} \iff n^2 \text{ is odd} \iff n \text{ is odd}.$$

Thus F is simply the set of all odd integers:

$$F = \{\pm 1, \pm 3, \pm 5, \pm 7, \dots\} = 2\mathbb{Z} + 1.$$

Plainly the two sets are not equal.

Intervals

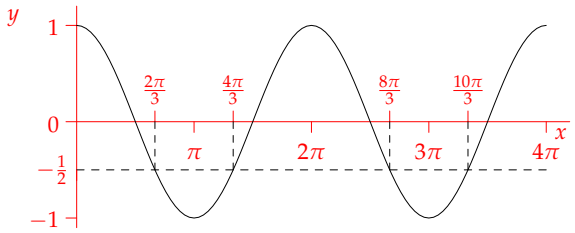
Interval notation is useful when discussing collections of *real numbers.* For example,

$$(0,1) = \{x \in \mathbb{R} : 0 < x < 1\},$$
$$[0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\},$$
$$(0,1] = \{x \in \mathbb{R} : 0 < x \leq 1\}.$$

When writing intervals with $\pm\infty$ use an open bracket at the infinite end(s): $[1,\infty) = \{x \in \mathbb{R} : x \geq 1\}$. This is since the symbols $\pm\infty$ do not represent real numbers and so are not members of any interval.

Example. Recall some basic trigonometry: the solutions of the equation $\cos x = -\frac{1}{2}$ on the interval $[0, 4\pi]$ can be written in set-builder and roster notation as

$$\left\{x \in [0,4\pi] : \cos x = -\frac{1}{2}\right\} = \left\{\frac{2\pi}{3}, \frac{4\pi}{3}, \frac{8\pi}{3}, \frac{10\pi}{3}\right\}$$



Cardinality and the Empty Set

Definition 4.2. A set A is *finite* if it contains a finite number of elements: this number is the set's *cardinality*, written $|A|$. A is said to be *infinite* otherwise.

Cardinality is a very simple concept for finite sets. For infinite sets, such as the natural numbers $\mathbb{N}$, the concept of cardinality is much more subtle. We cannot honestly talk about $\mathbb{N}$ having an 'infinite number' of elements, since infinity is not a number! In Chapter 8 we will consider what cardinality means for infinite sets and meet several bizarre and fun consequences. For the present, cardinality only has meaning for finite sets.

Examples.

1. Let $A = \{a, b, \alpha, \gamma, \sqrt{2}\}$, then $|A| = 5$.
2. Let $B = \left\{4, \{1,2\}, \{3\}\right\}$. It is important to note that the *elements/members* of B are 4, $\{1,2\}$ and $\{3\}$, two of which are themselves sets. Therefore $|B| = 3$. The set $\{1,2\}$ is an object in its own right, and can therefore be placed in a set along with other objects.[14]

To round things off we need a symbol to denote a set that contains nothing at all!

Axiom. There exists a set $\emptyset$ with no elements (cardinality zero: $|\emptyset| = 0$). We call $\emptyset$ the *empty set*.

There are many *representations* of the empty set. For example $\{x \in \mathbb{N} : x^2 + 3x + 2 = 0\}$ and $\{n \in \mathbb{N} : n < 0\}$ are both empty. Despite this, we will see in Theorem 4.4 that there is only one set with no elements, so that all such representations actually denote the *same set* $\emptyset$. Note also that $|A| \in \mathbb{N}$ for any *finite non-empty* set A.

Aside: Axioms

An axiom is a basic assumption; something that we need in order to do mathematics, but cannot prove. This is the cheat by which mathematicians can be 100% sure that something is true: a result is proved based on the assumption of several axioms. With regard to the empty set axiom, it probably seems bizarre that we can assume the existence of some set that has nothing in it. Regardless, mathematicians have universally agreed that we need the empty set in order to do the rest of mathematics.

[14] The fact that a set (containing objects) is also an object might seem confusing, but you should be familiar with the same problem in English. Consider the following sentences: 'UCI *are* constructing a laboratory' and 'UCI *is* constructing a laboratory.' In the first case we are thinking of UCI as a collection of individuals, in the latter case UCI is a single object. Opinions differ in various modes of English as to which is grammatically correct.

Exercises

4.1.1 Describe the following sets in roster notation, that is, list their elements.

(a) $\{x \in \mathbb{N} : x^2 \leq 3x\}$.

(b) $\{x^2 \in \mathbb{R} : x^2 - 3x + 2 = 0\}$.

(c) $\{n + 2 \in \{0, 1, 2, 3, \ldots, 19\} : n + 3 \equiv 5 \pmod 4\}$

(d) $\{n \in \{-2, -1, 0, 1, \ldots, 23\} : 4 \,|\, n^2\}$ (*does* : *or* | *denote the condition?*)

(e) $\{x \in \frac{1}{2}\mathbb{Z} : 0 \leq x \leq 4 \text{ and } 4x^2 \in 2\mathbb{Z} + 1\}$

4.1.2 Describe the following sets in set-builder notation (*look for a pattern*).

(a) $\{\ldots, -3, 0, 3, 6, 9, \ldots\}$

(b) $\{-3, 1, 5, 9, 13, \ldots\}$

(c) $\{1, \frac{1}{3}, \frac{1}{7}, \frac{1}{15}, \frac{1}{31}, \ldots\}$

4.1.3 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) $\{x \in \mathbb{R}^- : x^2 \geq 16 \text{ and } x^3 \leq 27\}$

4.1.4 Can you describe the set $\{x \in \mathbb{Z} : -1 \leq x < 43\}$ in interval notation? Why/why not?

4.1.5 Compare the sets $A = \{3x : x \in 2\mathbb{Z}\}$ and $B = \{x \in \mathbb{Z} : x \equiv 12 \pmod 6\}$. Are they equal?

4.1.6 What is the cardinality of the following set? What are the elements?

$$\Big\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\Big\}.$$

4.1.7 Let $A = \{\text{orange, banana, apple, mango}\}$, and let B be the set

$$B = \big\{\{x, y\} : x, y \in A\big\}.$$

(a) Describe B in roster notation.

(b) Now compute the cardinality of the sets

$$C = \big\{(x, y) : x, y \in A\big\}$$

and

$$D = \Big\{\big\{\{x, \{y\}\} : x, y \in A\big\}\Big\}.$$

Compare them to $|B|$.

4.2 Subsets

In this section we consider the most basic manner in which two sets can be related.

Definition 4.3. If A and B are sets such that every element of A is also an element of B, then we say that A is a *subset* of B and write $A \subseteq B$.
Sets A, B are *equal*, written $A = B$, if they have exactly the same elements. Equivalently

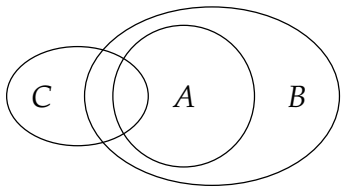
$$A = B \iff A \subseteq B \text{ and } B \subseteq A. \qquad (*)$$

A is a *proper subset* of B if it is a subset which is not equal. This can be written $A \subsetneq B$.[a]

[a] We will religiously stick to this notation. When reading other texts, note that some authors prefer $A \subset B$ for proper subset. Others use $\subset$ for any subset, whether proper or not.

The characterization $(*)$ of equality is *very* important. In order to *prove* that two sets are equal you will often have to show double-inclusion.

Venn diagrams are particularly useful for depicting subset relations. The graphic on the right depicts three sets A, B, C: it should be clear that the only valid subset relation between the three is $A \subseteq B$.



Set-builder notation implicitly uses the concept of subset: the notation $X = \{y \in Y : \ldots\}$ describes a set X as a subset of some larger set Y. The previous section contained many examples that were subsets of the set of real numbers $\mathbb{R}$. Here are some other examples of subsets.

Examples. 1. $\mathbb{N} = \{n \in \mathbb{Z} : n > 0\}$. This is clearly a subset of $\mathbb{Z}$.

2. $\{x \in \mathbb{R} : x^2 - 1 = 0\} \subseteq \{y \in \mathbb{R} : y^2 \in \mathbb{N}\}$.
   To make sense of this relationship, convert to roster notation: we obtain

   $$\{-1, 1\} \subseteq \{\pm\sqrt{1}, \pm\sqrt{2}, \pm\sqrt{3}, \pm\sqrt{4}, \ldots\}.$$

3. $m\mathbb{Z} \subseteq n\mathbb{Z} \iff n \mid m$. Make sure you're comfortable with this! For example, $4\mathbb{Z} \subseteq 2\mathbb{Z}$ since every multiple of 4 is also a multiple of 2.

Here we collect several results relating to subsets.

Theorem 4.4.
1. *If $|A| = 0$, then $A = \emptyset$* — *(Uniqueness of the empty set)*
2. *For any set A, we have $\emptyset \subseteq A$ and $A \subseteq A$* — *(Trivial and non-proper subsets)*
3. *If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$* — *(Transititvity of subsets)*

Proof. 1. Let A be a set with cardinality zero, i.e., with no elements. $\emptyset$ has no members, therefore $\emptyset \subseteq A$ is trivial: there is nothing to check to see that all elements of $\emptyset$ are also elements of A! The argument for $A \subseteq \emptyset$ is identical.

2. Let A be any set. $\emptyset \subseteq A$ follows by the argument in 1. To prove that $A \subseteq A$ we must show that all elements of A are also elements of A. But this is completely obvious!

3. Assume that A is a subset of B and that B is a subset of C. We must show that all elements of A are also elements of C. Let $a \in A$. Since $A \subseteq B$ we know that $a \in B$. Since $B \subseteq C$ and $a \in B$, we conclude that $a \in C$. This shows that every element of A belongs to C. Hence $A \subseteq C$. ■

As a final observation, to which we will return in Theorem 4.12 and in Chapter 8, your intutition should tell you that, for finite sets, subsets have smaller cardinatlity:

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

More generally, consider replacing the terms in Theorem 4.4 according to the following table:

$\subseteq$	$\leq$
$\emptyset$	0
sets A, B, C	non-negative integers
cardinality	absolute value

The results should seem completely natural! Recognizing the similarities between a new concept and a familiar one, essentially spotting patterns, is perhaps the most necessary skill in mathematics.

Exercises

4.2.1 Let A, B, C, D be the following sets.

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

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

4.2.2 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.

4.2.3 For which values of $x > 0$ is the following claim true?

$$[0, x] \subseteq [0, x^2]$$

Prove your assertion.

4.2.4 Given $A \subseteq \mathbb{Z}$ and $x \in \mathbb{Z}$, we say that x is A-mirrored if and only if $-x \in A$. We also define:

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

(a) What is the negation of 'x is A-mirrored.'

(b) Find M_B for $B = \{0, 1, -6, -7, 7, 100\}$.

(c) Assume that $A \subseteq \mathbb{Z}$ is closed under addition (i.e., $x + y \in A$, for all $x, y \in A$). Show that M_A is closed under addition.

(d) In your own words, under which conditions is $A = M_A$?

4.2.5 Define the set $[1]$ by:

$$[1] = \{x \in \mathbb{Z} \colon x \equiv 1 \pmod 5\}.$$

(a) Describe the set $[1]$ in roster notation.

(b) Compute the set $M_{[1]}$, as defined in Exercise 4.2.4

(c) Are the sets $[1]$ and $M_{[1]}$ equal? Prove/Disprove.

(d) Now consider the set $[10] = \{x \in \mathbb{Z} \colon x \equiv 10 \pmod 5\}$. Are the sets $[10]$ and $M_{[10]}$ equal? Prove/Disprove.

4.2.6 (a) Give a formal proof of the fact that $A \subseteq B \implies |A| \leq |B|$ for finite sets. *Resist the temptation to look at Theorem 4.12: it is far more technical than you need for this!*

(b) Explain why $|A| \leq |B| \not\Longrightarrow A \subseteq B$.

4.3 Unions, Intersections, and Complements

In the last section we compared nested sets. In this section we constuct new sets from old, modeled precisely on the logical concepts of *and*, *or*, and *not*. For the duration of this section, suppose that $\mathcal{U}$ is some *universal set*, of which every set mentioned subsequently is a subset.[15]

First we consider the set contruction modeled on *not*.

Definition 4.5. Let $A \subseteq \mathcal{U}$ be a set. The *complement* of A is the set

$$A^{\mathsf{C}} = \{x \in \mathcal{U} : x \notin A\}.$$

This can also be written $\mathcal{U} \setminus A$, $\mathcal{U} - A$, A', or $\overline{A}$.

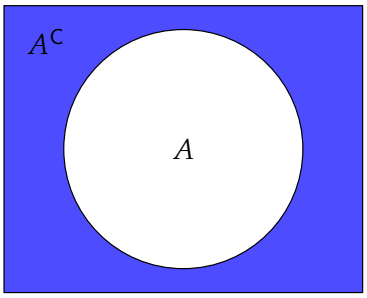
The Venn diagram is drawn on the right: A is represented by a circular region, while the rectangle represents the universal set $\mathcal{U}$. The complement A^{C} is the blue shaded region.

If $B \subseteq \mathcal{U}$ is some other set, then the *complement of A relative B* is

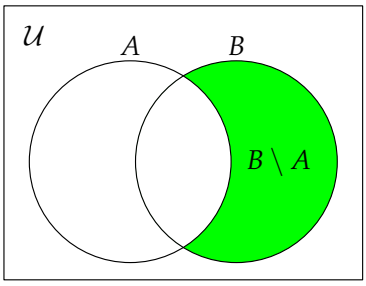
$$B \setminus A = \{x \in B : x \notin A\}.$$

The set $B \setminus A$ is also called B *minus* A. For its Venn diagram, we represent A and B as overlapping circular regions. The complement $B \setminus A$ is the green shaded region.

Note that $A^{\mathsf{C}} = \mathcal{U} \setminus A$, so that the two definitions correspond.



A^{C}: everything not in A



$B \setminus A$: everything in B but not in A

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

$$A^{\mathsf{C}} = \{4,5\}, \qquad B^{\mathsf{C}} = \{1,5\}, \qquad B \setminus A = \{4\}, \qquad A \setminus B = \{1\}.$$

Now we construct sets based on *or* and *and*.

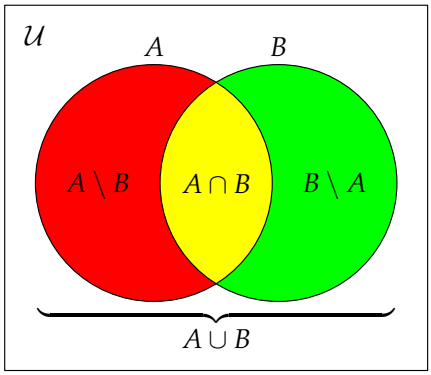
Definition 4.6. The *union* of A and B is the set

$$A \cup B = \{x \in \mathcal{U} : x \in A \text{ or } x \in B\}.$$

The *intersection* of A and B is the set

$$A \cap B = \{x \in \mathcal{U} : x \in A \text{ and } x \in B\}.$$

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



In the Venn diagram, the sets A and B are again depicted as overlapping circles. Although it doesn't constitute a proof, the diagram makes it clear that

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

[15] This is necessary so that the definitions in this section are legitimate.

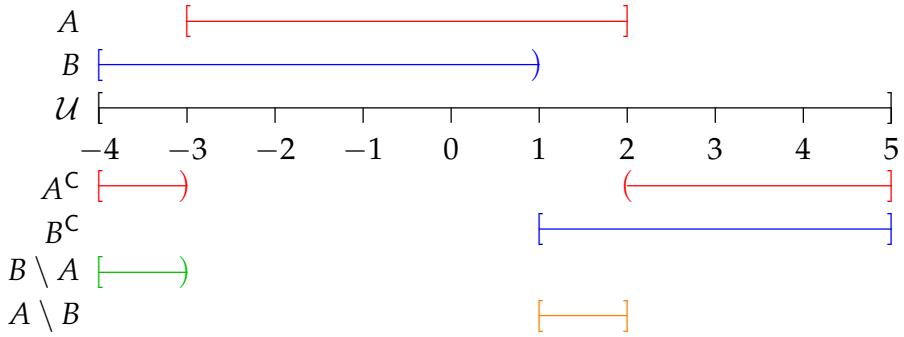
'Or' is used in the logical sense: $A \cup B$ is the collection of all elements that lie in A, in B, or in both. Now observe the notational pattern: $\cup$ looks very similar to the logic symbol $\vee$ from Chapter 2. The symbols $\cap$ and $\wedge$ are also similar.

Examples. 1. Let $\mathcal{U} = \{\text{fish, dog, cat, hamster}\}$, $A = \{\text{fish, cat}\}$, and $B = \{\text{dog, cat}\}$. Then,

$$A \cup B = \{\text{fish, dog, cat}\}, \qquad A \cap B = \{\text{cat}\}.$$

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

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



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

In the final example it seems reasonable to assume that $\mathcal{U} = \mathbb{R}$. The universal set is rarely made explicit in practice, and is often assumed to be the smallest suitable uncomplicated set. When dealing with sets of real numbers this typically means $\mathcal{U} = \mathbb{R}$. In other situations $\mathcal{U} = \mathbb{Z}$ or $\mathcal{U} = \{0, 1, 2, 3, \ldots, n-1\}$ might be more appropriate.

The next theorem comprises the basic rules of set algebra.

Theorem 4.7. *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* $A \cap C \subseteq B \cap C$.

You should be able to prove each of these properties directly from Definitions 4.3 and 4.6. Don't memorize the proofs: with a little practice working with sets, each of these results should feel completely obvious. It is more important that you are able to *vizualize* the laws using Venn diagrams. A Venn diagram does not constitute a formal proof, though it is extremely helpful for clarification. Here we prove only second result: think about how the Venn diagram in Defintion 4.6 illustrates the result. Some of the other proofs are in the Exercises.

Proof of 2. There are two results here: $A \cap B \subseteq A$ and $A \subseteq A \cup B$. We show each separately, along with some thinking.

Suppose that $x \in A \cap B$. (Must show $x \in A \cap B \Rightarrow x \in A$)
Then $x \in A$ and $x \in B$. (Definition of intersection)
But then $x \in A$, whence $A \cap B \subseteq A$ (Definition of subset)
Now let $y \in A$. (Must show $y \in A \Rightarrow y \in A \cup B$)
Then '$y \in A$ or $y \in B$' is true, from which we conclude that $y \in A \cup B$.
Thus $A \subseteq A \cup B$.

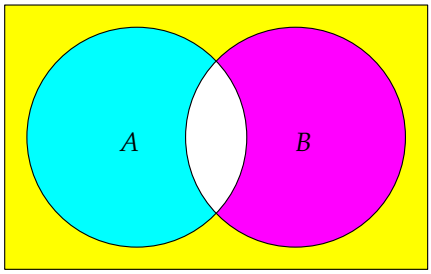
■

Once you get comfortable, you can strip away all the comments and write the proof more quickly.

The following theorem describes how complements interact with other set operations.

Theorem 4.8. *Let A, B be sets. Then:*

1. $(A \cap B)^{\mathsf{C}} = A^{\mathsf{C}} \cup B^{\mathsf{C}}$.
2. $(A \cup B)^{\mathsf{C}} = A^{\mathsf{C}} \cap B^{\mathsf{C}}$.
3. $(A^{\mathsf{C}})^{\mathsf{C}} = A$.
4. $A \setminus B = A \cap B^{\mathsf{C}}$.
5. $A \subseteq B \iff B^{\mathsf{C}} \subseteq A^{\mathsf{C}}$.



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

Again: don't memorize these laws! Draw Venn diagrams to help with visualization.

Proof of 1. We start by trying to show that the left hand side is a subset of the right hand side.

$$\begin{aligned} x \in (A \cap B)^{\mathsf{C}} &\implies x \notin A \cap B \\ &\implies x \text{ not a member of } \textit{both } A \text{ and } B \\ &\implies x \text{ not in } \textit{at least one} \text{ of } A \text{ and } B \\ &\implies x \notin A \text{ or } x \notin B \\ &\implies x \in A^{\mathsf{C}} \text{ or } x \in B^{\mathsf{C}} \\ &\implies x \in A^{\mathsf{C}} \cup B^{\mathsf{C}} \end{aligned}$$

With a little thinking, we realize that all of the $\implies$ arrows may be replaced with if and only if arrows $\iff$ without compromising the argument. We've therefore shown that the sets $(A \cap B)^{\mathsf{C}}$ and $A^{\mathsf{C}} \cup B^{\mathsf{C}}$ have the same elements, and are thus equal.

■

In the proof we were lucky. Showing that both sides are subsets of each other would have been tedious, but we found a quicker proof by carefully laying out one direction. This happens more often than you might think. Just be careful: you can't always make conditional connectives biconditional.

Parts 1. and 2. of the theorem are known as *De Morgan's laws*, just as the equivalent statements in logic: Theorem 2.9. Indeed, we could rephrase our proof in that language.

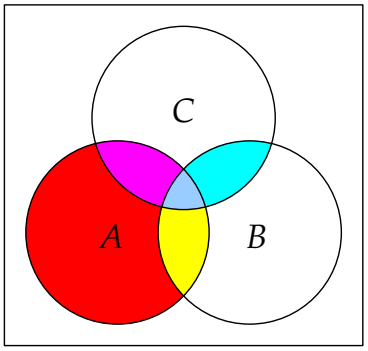
Alternative Proof of 1.

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

■

Theorem 4.9 (Distributive laws)**.** *For any sets A, B, C:*

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



We prove only the second result. The method is the standard approach: show that each side is a subset of the other. We do both directions this time, though with a little work and the cost of some clarity, you might be able to slim down the proof. The Venn diagram on the right illustrates the second result: simply add the colored regions.

Proof. ($\subseteq$) Let $x \in A \cup (B \cap C)$. Then $x \in A$ or $x \in B \cap C$. There are two cases:

(a) If $x \in A$, then $x \in A \cup B$ and $x \in A \cup C$ by Theorem 4.7, part 2.

(b) If $x \in B \cap C$, then $x \in B$ and $x \in C$. It follows that $x \in A \cup B$ and $x \in A \cup C$, again by Theorem 4.7.

In both cases $x \in (A \cup B) \cap (A \cup C)$.

($\supseteq$) Let $y \in (A \cup B) \cap (A \cup C)$. Then $y \in A \cup B$ and $y \in A \cup C$. There are again two cases:

(a) If $y \in A$, then we are done, for then $y \in A \cup (B \cap C)$.

(b) If $y \notin A$, then $y \in B$ and $y \in C$. Hence $y \in B \cap C$. In particular $y \in A \cup (B \cap C)$.

In both cases $y \in A \cup (B \cap C)$.

■

Exercises

4.3.1 Describe each of the following sets in as simple a manner as you can: e.g.,

$$\{x \in \mathbb{R} : (x^2 > 4 \text{ and } x^3 < 27) \text{ or } x^2 = 15\} = (-\infty, -2) \cup (2, 3) \cup \{\sqrt{15}, -\sqrt{15}\}.$$

(a) $\{x \in \mathbb{R} : x^2 \neq x\}$

(b) $\{x \in \mathbb{R} : x^3 - 2x^2 - 3x \leq 0 \text{ or } x^2 = 4\}$

(c) $\{x^2 \in \mathbb{R} : x \neq 1\}$

(d) $\{z \in \mathbb{Z} : z^2 \text{ is even and } z^3 \text{ is odd}\}$

(e) $\{y \in 3\mathbb{Z} + 2 : y^2 \equiv 1 \pmod 3\}$

4.3.2 Let $A = \{1, 3, 5, 7, 9, 11\}$ and $B = \{1, 4, 7, 10, 13\}$. What are the following sets?

(a) $A \cap B$

(b) $A \cup B$

(c) $A \setminus B$

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

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

$$\exists d > 0 \colon \text{ No element of } A \text{ belongs to the set } [x - d, x].$$

Equivalently, $A \cap [x - d, x] = \emptyset$. If this does not happen, we say that x is *close to* A.

(a) Draw a picture of a set A and an element x such that is *far away* from A.

(b) Draw a picture of a set A and an element x such that x is *close* to A.

(c) Compute the definition of "x is close to A". [So negate "x is far away from A".]

(d) Let $A = \{1, 2, 3\}$. Show that $x = 4$ is *far away* from A, by using definitions.

(e) Let $A = \{1, 2, 3\}$. Show that $x = 1$ is *close* to A, by using definitions.

(f) Show that if $x \in A$, then x is *close* to A.

(g) Let A be the open interval (a, b). Is the end-point *a far away* from A? What about the end-point b?

4.3.4 Consider Theorems 4.7 and 4.9. In all seven results, replace the symbols in the first row of the following table with those in the second. Which of the results seem familar? Which are false?

$\emptyset$	A, B, C sets	$\cup$	$\cap$	$\subseteq$
0	$A, B, C \in \mathbb{N}_0$	$+$	$\cdot$	$\leq$

4.3.5 Prove that $B \setminus A = B \iff A \cap B = \emptyset$.

4.3.6 Practice your proof skills by giving formal proofs of the following results from Theorems 4.7 and 4.8. With practice you should be able to prove *all* of parts of these theorems (and of Theorem 4.9) these *without* looking at the arguments in the notes!

(a) $\emptyset \cap A = \emptyset$.

(b) $A \cap (B \cap C) = (A \cap B) \cap C$.

(c) $(A^{\mathsf{C}})^{\mathsf{C}} = A$.

(d) $A \subseteq B \iff B^{\mathsf{C}} \subseteq A^{\mathsf{C}}$.

4.4 Introduction to Functions

You have been using functions for a long time. A formal definition in terms of relations will be given in Section 7.2. For the present, we will just use the following.

Definition 4.10. Let A and B be sets. A *function from A to B* is a rule f that assigns one (and only one) element of B to each element of A.
The *domain* of f, written $\text{dom}(f)$, is the set A. The *codomain* of f is the set B.
The *range* of f, written $\text{range}(f)$ or $\text{Im}(f)$, is the subset of B consisting of all the elements assigned by f.

You can think of the domain of f as the set of all inputs for the function, and the range of f as the set of all outputs. The codomain is the set of all potential values the function may take (of course, only the values in the range are actually achieved).

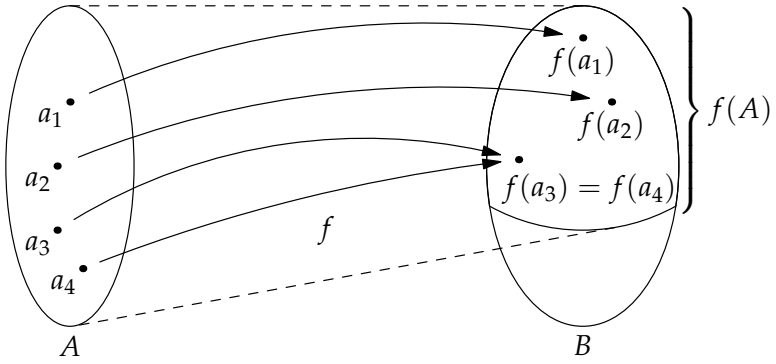
Notation

If f is a function from A to B we write $f : A \to B$.
If $a \in A$, we write $b = f(a)$ for the the element of B assigned to a by the function f.
We can also write $f : a \mapsto b$, which is read 'f maps a to b.'
If U is a subset of A then the *image* of U is the following subset of B,

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

The image of A is precisely the range of f, hence the notaion $\text{Im}(f)$,

$$f(A) = \text{range}(f) = \text{Im}(f) = \{f(a) : a \in A\}.$$

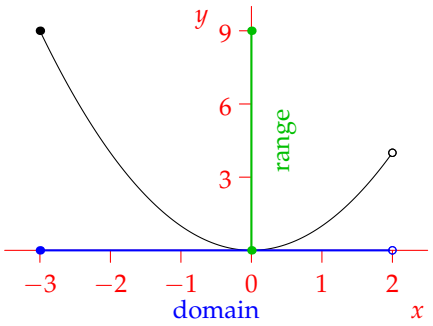


Examples. 1. Let $f : [-3, 2) \to \mathbb{R}$ be the square function

$$f : x \mapsto x^2.$$

We have $\text{dom}(f) = [-3, 2)$, and $\text{range}(f) = [0, 9]$, as shown in the picture. We could also calculate other images, for example,

$$f\big([-1, 2)\big) = [0, 4).$$



2. Define $f : \mathbb{Z} \to \{0,1,2\}$ by $f : n \mapsto n^2 \pmod 3$, where we take the remainder of n^2 modulo 3. Clearly $\text{dom}(f) = \mathbb{Z}$, but what is the range? Trying a few examples, we see the following:

n	0	1	2	3	4	5	6	7	8	9	10
$f(n)$	0	1	1	0	1	1	0	1	1	0	1

It looks like the range is simply $\{0,1\}$. We have already proved this fact in Theorem 2.17, although a faster proof can now be given by appealing to modular arithmetic (Section 3.1).

If $n \equiv 0$, then $n^2 \equiv 0 \pmod 3$.
If $n \equiv 1$, then $n^2 \equiv 1 \pmod 3$.
If $n \equiv 2$, then $n^2 \equiv 4 \equiv 1 \pmod 3$.

Thus $n^2 \equiv 0, 1 \pmod 3$, and $\text{range}(f) = \{0,1\}$.

3. Let $A = \{0,1,2,\ldots,9\}$ be the set of remainders modulo 10 and define $f : A \to A$ by $f : n \mapsto 3n \pmod{10}$. To help understand this function, list the elements: the domain only has 10 elements after all.

n	0	1	2	3	4	5	6	7	8	9
$f(n)$	0	3	6	9	2	5	8	1	4	7

It should be obvious that $\text{range}(f) = A$.

4. With the same notation as the previous example, let $g : A \to A : n \mapsto 4n \pmod{10}$. Now we have the following table:

n	0	1	2	3	4	5	6	7	8	9
$g(n)$	0	4	8	2	6	0	4	8	2	6

with $\text{range}(g) = \{0,2,4,6,8\}$.

Injections, surjections and bijections

Definition 4.11. A function $f : A \to B$ is *1–1* (one-to-one), *injective*, or an *injection* if it never takes the same value twice. Equivalently,[a]

$$\forall a_1, a_2 \in A,\ f(a_1) = f(a_2) \implies a_1 = a_2.$$

$f : A \to B$ is *onto*, *surjective*, or a *surjection* if it takes every value in the codomain: i.e., $B = \text{range}(f)$. Equivalently,

$$\forall b \in B,\ \exists a \in A \text{ such that } f(a) = b.$$

$f : A \to B$ is *invertible*, *bijective*, or a *bijection* if it is both injective and surjective.

[a] This is the contrapositive: if f never takes the same value twice, then $\forall a_1, a_2 \in A$ we have $a_1 \neq a_2 \implies f(a_1) \neq f(a_2)$.

Remark: Since the definitions of injective and surjective are both 'forall' statements, to show that a function is *not injective* or *not surjective* you will need *counterexamples.*

First we consider our examples above. The details are provided for 1 and 2. For the remaining examples, make sure you understand why the answer is correct.

1. $f : [-3, 2) \to \mathbb{R} : x \mapsto x^2$ is neither injective nor surjective. Indeed we have the following counterexamples:
   - $f(-1) = f(1)$. If f were injective, the values at 1 and -1 would have to be different.
   - $81 \in \mathbb{R}$, yet there is no $x \in [-3, 2)$ such that $f(x) = 81$. Thus f is not surjective.
2. $f : \mathbb{Z} \to \{0, 1, 2\} : n \mapsto n^2 \pmod 3$ is neither injective nor surjective.
   - If f were injective, then we could not have $f(1) = f(2)$.
   - 2 is in the codomain $\{0, 1, 2\}$ of f, yet $2 \notin \text{range}(f)$, so f is not surjective.
3. A bijection: this is an example of a *permutation*, a bijection from a set onto itself.
4. Neither injective, nor surjective.

Here is a more complicated example.

Example. Prove that $f : \mathbb{R} \setminus \{1\} \to \mathbb{R} \setminus \{2\}$ defined by $f(x) = 2 + \frac{1}{1-x}$ is bijective.

(Injectivity) Suppose that x_1 and x_2 are in $\mathbb{R} \setminus \{1\}$, and $f(x_1) = f(x_2)$. Then

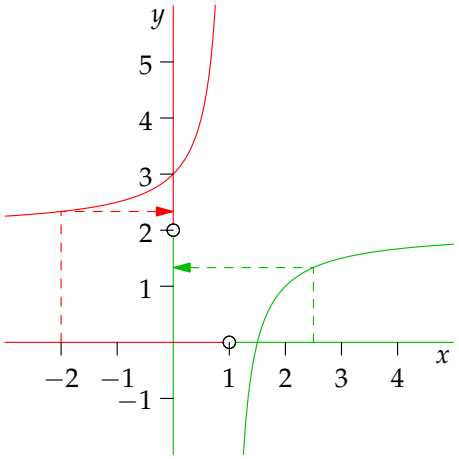
$$2 + \frac{1}{1 - x_1} = 2 + \frac{1}{1 - x_2}.$$

A little elementary algebra shows that $x_1 = x_2$, whence f is injective.

(Surjectivity) Let $y \in \mathbb{R} \setminus \{2\}$ and define $x = 1 - \frac{1}{y-2}$. This makes sense since $y \neq 2$. Then

$$f(x) = 2 + \frac{1}{1 - (1 - \frac{1}{y-2})} = y$$

whence f is surjective.



The graphic is colored so that you can see how the different parts of the range and domain correspond bijectively. The argument for surjectivity is sneaky: how did we know to choose $x = 1 - \frac{1}{y-2}$? The answer is scratch work: just solve $y = 2 + \frac{1}{1-x}$ for x. Essentially we've shown that f has the inverse function $f^{-1}(x) = 1 - \frac{1}{x-2}$.

Aside: Inverse Functions

The word *invertible* is a synonym for bijective because bijective functions really have inverses! Indeed, suppose that $f : A \to B$ is bijective. Since f is surjective, we know that $B = \text{range}(f)$ and so every element of B has the form $f(a)$ for some $a \in A$. Moreover, since f is injective, the a in question is unique. The upshot is that, when f is bijective, we can construct a new *function*

$$f^{-1} : B \to A : f(a) \mapsto a.$$

This may appear difficult at the moment but we will return to it in Chapter 7.

Instead, recall that in Calculus we saw that any injective function has an inverse. How does this fit with our definition? Consider, for example, $f : [0,3] \to \mathbb{R} : x \mapsto x^2$. This is injective but not surjective. To fix this, simply define a new function with the same formula but with codomain equal to the range of f. We obtain the bijective function

$$g : [0,3] \to [0,9] : x \mapsto x^2,$$

with inverse

$$g^{-1} : [0,9] \to [0,3] : x \mapsto \sqrt{x}.$$

In Calculus we didn't nitpick like this and would simply go straight to $f^{-1}(x) = \sqrt{x}$.
In general, if $f : A \to B$ is any injective function, then $g : A \to f(A) : x \mapsto f(x)$ is automatically bijective, since we are forcing the codomain of g to match its range.

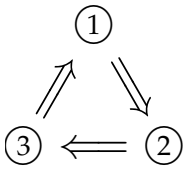
Functions and Cardinality

Injective and surjective functions are intimately tied to the notion of cardinality. Indeed, in Chapter 8, we will use such functions to give a *definition* of cardinality for infinite sets. For the present we stick to finite sets.

Theorem 4.12. *Let A and B be finite sets. The following are equivalent:*

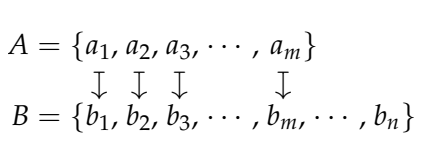
1. $|A| \leq |B|$.
2. $\exists f : A \to B$ *injective.*
3. $\exists g : B \to A$ *surjective.*

Read the theorem carefully. It is simply saying that, of the three statements, if *any* one is true then *all* are true. Similarly, if one is false then so are the others. It might appear that we require six arguments! Instead we illustrate an important technique: when showing that multiple statements are equivalent, it is enough to prove in a circle. E.g., if we prove the three implications indicated in the picture, then ① ⇒ ③ will be true because *both* ① ⇒ ② and ② ⇒ ③ are true.

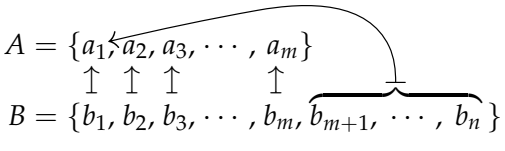


More generally, to show that n statements are equivalent, only n arguments are required.

The proof may appear very abstract, but it is motivated by two straightforward pictures. Don't be afraid to use pictures to illustrate your proofs if it's going to make them easier to follow! If $|A| = m$ and $|B| = n$, then the two functions can be displayed pictorially. Refer back to these pictures as you read through the proof.



The function f



The function g

Proof. The proof relies crucially on the fact that A, B are finite. Suppose that $|A| = m$ and $|B| = n$ throughout and list the elements of A and B as,

$$A = \{a_1, a_2, \ldots, a_m\}, \qquad B = \{b_1, b_2, \ldots, b_n\}.$$

$(① \Rightarrow ②)$ Assume that $m \leq n$. Define $f : A \to B$ by $f(a_k) = b_k$. This is injective since the elements $b_1, \ldots, b_m$ are distinct.

$(② \Rightarrow ③)$ Suppose that $f : A \to B$ is injective. Without loss of generality we may assume that the elements of A and B are labeled such that $f(a_k) = b_k$. Now define $g : B \to A$ by

$$g(b_k) = \begin{cases} a_k & \text{if } k \leq m, \\ a_1 & \text{if } k > m. \end{cases}$$

Then g is surjective since every element a_k is in the image of g.

$(③ \Rightarrow ①)$ Finally suppose that $g : B \to A$ is surjective. Without loss of generality we may assume that $a_k = g(b_k)$ for $1 \leq k \leq m$. Thus $n \geq m$. ■

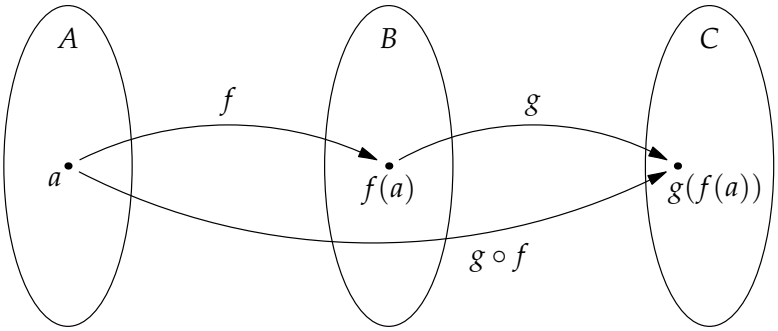
If you read the proof carefully, it should be clear that when $m = n$, the function f is actually a *bijection* (with inverse $f^{-1} = g$).

Corollary 4.13. *If A, B are finite sets, then $|A| = |B| \iff \exists f : A \to B$ bijective.*

Proof. Suppose that $m = n$. The argument ① $\Rightarrow$ ② creates an injective function $f : A \to B$. However every element $b_k \in B$ is in the image of f, so this function is also surjective. Hence f is a bijection. Conversely, if $f : A \to B$ is a bijection, then it is injective, whence $m \leq n$. It is also surjective, from which $n \leq m$. Therefore $m = n$. ■

Composition of functions

Definition 4.14. Suppose that $f : A \to B$ and $g : B \to C$ are functions. The *composition* $g \circ f : A \to C$ is the function defined by $(g \circ f)(a) = g(f(a))$. Note the order: to compute $(g \circ f)(x)$, you apply f first, then g.



Example. If $f(x) = x^2$ and $g(x) = \frac{1}{x-1}$, then

$$(g \circ f)(x) = \frac{1}{x^2 - 1}, \quad \text{and} \quad (f \circ g)(x) = \frac{1}{(x-1)^2}.$$

You should be extra careful of ranges and domains when composing functions. The domain and range are not always explicitly mentioned, and at times some restriction of the domain is implied. In this example, you might assume that $\mathrm{dom}(f) = \mathbb{R}$ and $\mathrm{dom}(g) = \mathbb{R} \setminus \{1\}$. This is perfectly good if we are considering f and g separately. However, it should be clear from the formulæ that the implied domains of the compositions are,

$$\mathrm{dom}(g \circ f) = \mathbb{R} \setminus \{\pm 1\}, \quad \text{and} \quad \mathrm{dom}(f \circ g) = \mathbb{R} \setminus \{1\}.$$

Finally we consider how injectivity and surjectivity interact with composition.

Theorem 4.15. *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. 1. Suppose that f and g are injective and let $a_1, a_2 \in A$ satisfy $(g \circ f)(a_1) = (g \circ f)(a_2)$. We are required to show that $a_1 = a_2$. However,

$$\begin{aligned} (g \circ f)(a_1) = (g \circ f)(a_2) &\implies g\big(f(a_1)\big) = g\big(f(a_2)\big) \\ &\implies f(a_1) = f(a_2) && \text{(since } g \text{ is injective)} \\ &\implies a_1 = a_2 && \text{(since } f \text{ is injective)} \end{aligned}$$

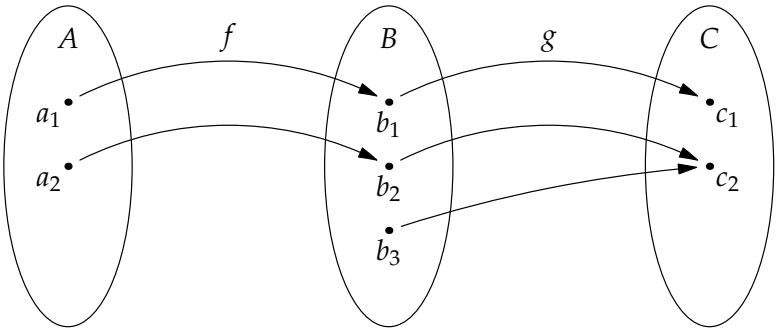
■

Part 2 is in the Exercises. It is an interesting to observe that the converse of this theorem is *false.* Assuming that a composition is injective or surjective only requires that *one* of the component functions be so.

Theorem 4.16. *Suppose that $f : A \to B$ and $g : B \to C$ are functions.*

1. *If $g \circ f$ is injective, then f is injective.*
2. *If $g \circ f$ is surjective, then g is surjective.*

Before showing the proof, consider the following picture of two functions f and g which simultaneously illustrate both parts of the theorem. It should be clear that $g \circ f$ is *bijective*, f is *only injective*, and g is *only surjective.*



Here is a formulaic example of the same thing. Make sure you're comfortable with the definitions and draw pictures or graphs to help make sense of what's going on.

$f : [0,2] \to [-4,4] : x \mapsto x^2$ (injective only)

$g : [-4,4] \to [0,16] : x \mapsto x^2$ (surjective only)

$g \circ f : [0,2] \to [0,16] : x \mapsto x^4$ (bijective!)

Proof. 2. Let $c \in C$ and assume that $g \circ f$ is surjective. We wish to prove that $\exists b \in B$ such that $g(b) = c$.
Since $g \circ f$ is surjective, $\exists a \in A$ such that $(g \circ f)(a) = c$. But this says that

$$g(f(a)) = c.$$

Hence $b = f(a)$ is an element of B for which $g(b) = c$. Thus g is surjective. ■

We leave part 1 for the Exercises.

Exercises

4.4.1 For each of the following functions $f : A \to B$ determine whether f is injective, surjective or bijective. Prove your assertions.

(a) $f : [0,3] \to \mathbb{R}$ where $f(x) = 2x$.
(b) $f : [3,12) \to [0,3)$ where $f(x) = \sqrt{x-3}$.
(c) $f : (-4,1] \to (-5,-3]$ where $f(x) = -\sqrt{x^2+9}$.

4.4.2 Suppose that $f : [-3, \infty) \to [-8, \infty)$ and $g : \mathbb{R} \to \mathbb{R}$ are defined by

$$f(x) = x^2 + 6x + 1, \qquad g(x) = 2x + 3.$$

Compute $g \circ f$ and show that $g \circ f$ is injective.

4.4.3 (*If you did Exercise 2.3.12 you should find this easy*) Let X be a subset of $\mathbb{R}$. A function $f : X \to \mathbb{R}$ is *strictly increasing* if

$$\forall\, a, b \in X, \quad a < b \implies f(a) < f(b).$$

For example, the function $f\colon [0, \infty) \to \mathbb{R}$, $x \mapsto x^2$ is increasing because

$$\forall a, b \in [0, \infty), \quad a < b \implies f(a) = a^2 < b^2 = f(b).$$

(a) Give another example of a function that is increasing. Draw its graph, and prove that the function is increasing.

(b) By negating the above definition, state what it means for a function *not to be strictly increasing*.

(c) Give an example of a function that is *not* strictly increasing. Draw its graph, and prove that the function is not stictly increasing.

(d) Let $f, g : \mathbb{R} \to \mathbb{R}$ be strictly increasing. Prove or disprove: The function $h = f + g$ is strictly increasing. Note that the formula for h is $h(x) = f(x) + g(x)$.

4.4.4 Find:

(a) A set A so that the function $f : A \to \mathbb{R} : x \mapsto \sin x$ is injective.

(b) A set B so that the function $f : \mathbb{R} \to B : x \mapsto \sin x$ is surjective.

4.4.5 A function $f : \mathbb{R} \to \mathbb{R}$ is *even* if

$$\forall\, x \in \mathbb{R},\ f(-x) = f(x).$$

For example, the function $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ is even because

$$\forall\, x \in \mathbb{R},\ f(-x) = (-x)^2 = x^2 = f(x).$$

Note that f is even if and only if the graph of f is symmetric with respect to the y axis.

(a) Give an example of a function that is even. Draw its graph, and prove that the function is even.

(b) Define what it means for a function *not to be even*, by negating the definition above.

(c) Give an example of a function that is *not* even. Draw its graph, and prove that the function is not even.

(d) Prove or disprove: for every $f, g\colon \mathbb{R} \to \mathbb{R}$ even, the composition $h = f \circ g$ is even. Here h is the function mapping x to $f(g(x))$.

4.4.6 Define $f : (-\infty, 0] \to \mathbb{R}$ and $g : [0, \infty) \to \mathbb{R}$ by

$$f(x) = x^2, \qquad g(x) = \begin{cases} \frac{x}{1-x} & x < 1, \\ 1 - x & x \geq 1. \end{cases}$$

Does $g \circ f$ map $(-\infty, 0]$ onto $\mathbb{R}$? Justify your answer.

4.4.7 Negate Definition 4.11 to find what it means for a function to be

(a) Not injective.

(b) Not surjective.

4.4.8 Prove that the composition of two surjective functions is surjective.

4.4.9 Suppose that $g \circ f$ is injective. Prove that f is injective.

4.4.10 In the proof of Theorem 4.12 we twice invoked *without loss of generality.* In both cases explain why the phrase applies.

4.4.11 Recall Examples 3 and 4 on page 67.

(a) Consider the nine functions $f_k : A \to A : x \mapsto kx \pmod{10}$, where $k = 1, 2, \ldots, 9$. Find the range of f_k for each k. Can you find a relationship between the cardinality of range(f_k) and k?

(b) More generally, let $A = \{0, 1, 2 \ldots, n-1\}$ be the set of remainders modulo n. If $f_k : A \to A : x \mapsto kx \pmod{n}$, conjecture a relationship between $|\text{range}(f_k)|$, k and n. You don't need to prove your assertions.

5 Mathematical Induction and Well-ordering

In Section 2.2 we discussed three methods of proof: direct, contrapositive, and contradiction. The fourth standard method of proof, *induction,* has a very different flavor. In practice it formalizes the idea of spotting a pattern. Before we give the formal definition of induction, we consider where induction fits into the investigative process.

5.1 Investigating Recursive Processes

In applications of mathematics, one often has a simple recurrence relation but no general formula. For instance, a process might be described by an expression of the form

$$x_{n+1} = f(x_n),$$

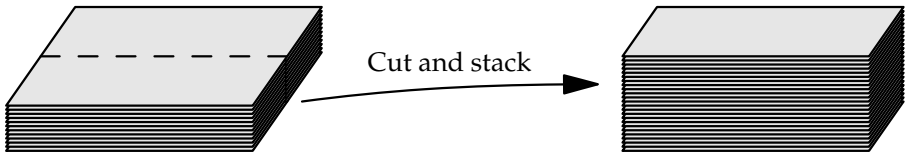
where some initial value x_1 is given. While investigating such recurrences, you might hypothesize a *general formula*

$$x_n = g(n).$$

Induction is a method of proof that allows us to *prove* the correctness of such general formulæ. Here is a simple example of the process.

Stacking Paper

Consider the operation whereby you take a stack of paper, cut all sheets in half, then stack both halves together.



If a single sheet of paper has thickness 0.1 mm, how many times would you have to repeat the process until the stack of paper reached to the sun? ($\approx$ 150 million kilometers).

The example is describing a recurrence relation. If h_n is the height of the stack after n operations, then we have a sequence $(h_n)_{n=0}^{\infty}$ satisfying

$$\begin{cases} h_{n+1} = 2h_n \\ h_0 = 0.1\,\text{mm}. \end{cases}$$

It is easy to compute the first few terms of the sequence:

n	0	1	2	3	4	5	6	7	8	$\cdots$
h_n (mm)	0.1	0.2	0.4	0.8	1.6	3.2	6.4	12.8	25.6	$\cdots$

It is not hard to hypothesize that, after n such operations, the stack of paper will have height

$$h_n = 2^n \times 0.1\,\text{mm}.$$

All we have done is to spot a pattern. We can reassure ourselves by checking that the first few terms of the sequence satisfy the formula: certainly $h_0 = 2^0 \times 0.1\,\text{mm}$ and $h_1 = 2^1 \times 0.1\,\text{mm}$, etc. Unfortunately the sequence has *infinitely many* terms, so we need a trick which confirms *all of them at once.* Unless we can *prove* that our formula is correct *for all* $n \in \mathbb{N}_0$ it will remain just a guess. This is where induction steps in.

The trick is called the *induction step*. We *assume* that we have already confirmed the formula for some fixed, but unspecified, value of n and then use what we know (the recurrence relation $h_{n+1} = 2h_n$) to confirm the formula for the *next value* $n+1$. Here it goes:

Induction Step Suppose that $h_n = 2^n \times 0.1$ mm, for some fixed $n \in \mathbb{N}_0$. Then

$$h_{n+1} = 2h_n = 2(2^n \times 0.1) = 2^{n+1} \times 0.1 \text{ mm}.$$

This is exactly the expression we hoped to find for the $(n+1)$th term of the sequence. Think about what the induction step is doing. By leaving n unspecified, we have proved an *infinite collection of implications at once!* Each implication has the form

$$h_n = 2^n \times 0.1 \implies h_{n+1} = 2^{n+1} \times 0.1.$$

Since the implications have been proved for all $n \in \mathbb{N}_0$, we can string them together:

$$h_0 = 2^0 \times 0.1 \implies h_1 = 2^1 \times 0.1 \implies h_2 = 2^2 \times 0.1 \implies h_3 = 2^3 \times 0.1 \implies \cdots$$

We have already checked that the first formula $h_0 = 2^0 \times 0.1$ in the implication chain is true. By the induction step, the *entire infinite collection of formulæ must be true.* We have therefore *proved* that

$$h_n = 2^n \times 0.1\,\text{mm} = 2^n \times 10^{-4}\,\text{m}, \quad \forall n \geq 0.$$

Now that we've proved the formula for every h_n, finishing the original problem is easy: we need to find $n \in \mathbb{N}_0$ such that

$$h_n = 2^n \times 10^{-4} \geq 150 \times 10^9\,\text{m} \iff 2^n \geq 15 \times 10^{14}.$$

Since logorithms are increasing functions, they preserve inequalities and we may easily solve to see that

$$n \geq \log_2(15 \times 10^{14}) = \log_2 15 + 14 \log_2 10 \approx 50.4.$$

Thus 51 iterations of the cut-and-stack process are sufficient for the pile of paper to reach the sun!

We will formalize the discussion of induction in the next section so that you will never have to write as much as we've just done. However, it is important to remember how induction fits into a practical investigation. It is the missing piece of logic that turns a *guess* into a justified formula. Before we do so, here is a famous and slightly more complicated problem.

The Tower of Hanoi

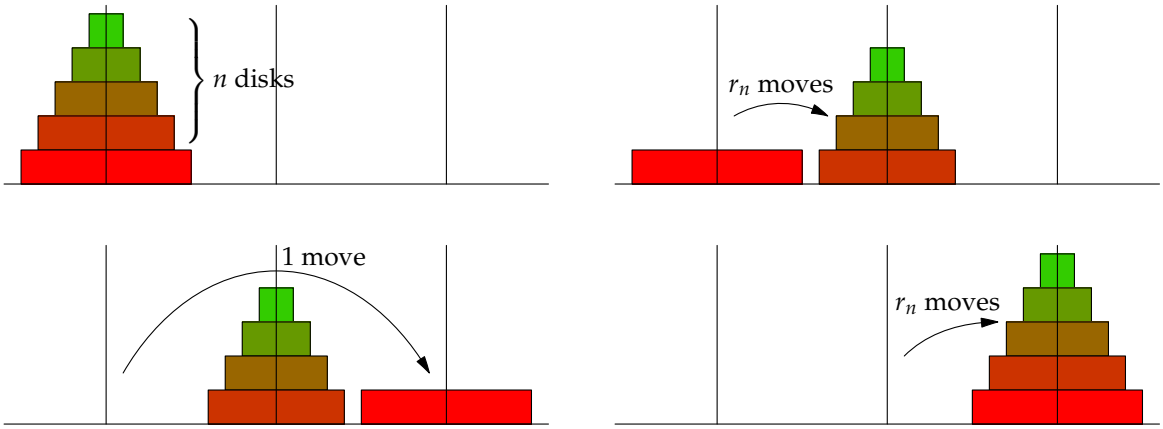
The *Tower of Hanoi* is a game involving circular disks of decreasing radii stacked on three pegs. A 'move' consists of transferring the top disk in any stack onto a larger disk or an empty peg. If we start with n disks on the first peg, how many moves are required to transfer all the disks to one of the other pegs?

The challenge here is that we have no formula to play with, only the variable n for the number of disks. The first thing to do is to play the game. If the variable r_n represents the number of moves required when there are n disks, then it should be immediately clear that $r_1 = 1$: one disk only requires one move! The picture below shows that $r_2 = 3$.

With more disks you can keep experimenting and find that $r_3 = 7$, etc. At this point you may be ready to hypothesize a general formula.

Conjecture 5.1. *The Tower of Hanoi with n disks requires* $r_n = 2^n - 1$ *moves.*

Certainly the conjecture is true for $n = 1, 2$ and 3. To see that it is true in general, we need to think about how to move a stack of $n + 1$ disks. Since the largest disk can only be moved onto an empty peg, it follows that the n smaller disks must already be stacked on a single peg *before* the $(n + 1)$th disk can move. From the starting position this requires r_n moves.



The largest disk can now be moved to the final peg, before the original n disks are moved on top of it. In total this requires $r_n + 1 + r_n$ moves, as illustrated in the picture. We therefore have a recurrence relation for r_n:

$$\begin{cases} r_{n+1} = 2r_n + 1 \\ r_1 = 1. \end{cases}$$

We are now in a position to prove our conjecture. Again we know that the conjecture is true for $n = 1$ and we assume that the formula $r_n = 2^n - 1$ is true for some fixed but unspecified n. Now we

use the recurrence relation to prove that $r_{n+1} = 2^{n+1} - 1$.

Induction Step Suppose that $r_n = 2^n - 1$ for some fixed $n \in \mathbb{N}$. Then

$$r_{n+1} = 2r_n + 1 = 2(2^n - 1) + 1 = 2^{n+1} - 2 + 1 = 2^{n+1} - 1.$$

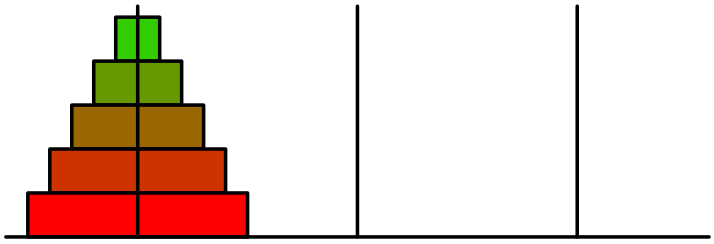
Exactly as in the paper-stacking example, we have simultaneously proved an *infinite collection of implications*:

$$r_1 = 2^1 - 1 \implies r_2 = 2^2 - 1 \implies r_3 = 2^3 - 1 \implies r_4 = 2^4 - 1 \implies \cdots$$

Since the first of these statements is true, it follows that *all of the others are true.* Hence Conjecture 5.1 is true, and becomes a theorem.

As an illustration of how ridiculously time-consuming the Tower becomes, the following table gives the time taken to complete the Tower if you were able to move one disk per second.

Disks	Time
5	31sec
10	17min 3sec
15	9hr 6min 7sec
20	12days 3hrs 16min 15sec
25	$\sim$ 1yr 23days
30	$\sim$ 34yrs 9days



Animation of five disks (click)

Exercises

5.1.1 A room contains n people. Everybody wants to shake everyone else's hand (but not their own).

(a) Suppose that n people require h_n handshakes. If an $(n+1)$th person enters the room, how many *additional* handshakes are required? Obtain a recurrence relation for h_{n+1} in terms of h_n.

(b) Hypothesize a general formula for h_n, and prove it using the method in this section.

5.1.2 Skippy the Kangaroo is playing jump rope, but he tires as the day goes on. The heights h_n (inches) of successive jumps are related by the recurrence

$$h_{n+1} = \frac{8}{9}h_n + 1.$$

(a) Suppose that Skippy's initial jump has height $h_1 = 100$ in. Show that Skippy fails to jump above 10in for the first time on the 40th jump.

(b) Find the *total* height jumped by Skippy in the first n jumps.

You may find it useful to define $H_n = h_n - 9$ and think about the recurrence for H_n. Now guess and prove a general formula for H_n. Finally, remind yourself about geometric series.)

5.2 Proof by Induction

The previous section motivated the need for induction and helped us see where induction fits into a logical investigation. In this section we formally lay out several induction proofs.

Induction is the mathematical equivalent of a domino rally; toppling the nth domino causes the $(n+1)$th domino to fall, hence to knock all the dominos over it is enough merely to topple the first. Instead of dominoes, in mathematics we consider a sequence of *propositions:* $P(1)$, $P(2)$, $P(3)$, etc. Induction demonstrates the truth of *every* proposition $P(n)$ by doing two things:

1. Proving that $P(1)$ is true (*Base Case*)
2. Proving that $\forall n \in \mathbb{N},\ P(n) \implies P(n+1)$ (*Induction Step*)

You could think of the base case as knocking over the first domino, and the induction step as the nth domino knocking over the $(n+1)$th, *for all* n. Both of the examples in the previous section followed this pattern. Unpacking the induction step gives an infinite chain of implications:

$$P(1) \implies P(2) \implies P(3) \implies P(4) \implies P(5) \implies \cdots.$$

The base case says that $P(1)$ is true, and so *all* of the remaining propositions $P(2)$, $P(3)$, $P(4)$, $P(5)$, ... are also true.

All induction proofs have the same formal structure:

(*Set-up*)	Define $P(n)$, set-up notation and orient the reader as to what you are about to prove.
(*Base Case*)	Prove $P(1)$.
(*Induction Step*)	Let $n \in \mathbb{N}$ be fixed and assume that $P(n)$ is true. This assumption is the *induction hypothesis.* Perform calculations or other reasonings to conclude that $P(n+1)$ is true.
(*Conclusion*)	Remind the reader what it is you have proved.

As you read more mathematics, you will find that the induction step is often the most involved part of the proof. The *set-up* stage is often no more than a sentence: 'We prove by induction,' and the explicit definition of $P(n)$ is commonly omitted. These are the only shortcuts that it is sensible to take until you are extremely comfortable with induction. Practice making it completely clear what you are doing at each juncture.

Here is a straightforward theorem, where we write the proof in the above language.

Theorem 5.2. *The sum of the first n positive integers is given by the formula*

$$\sum_{i=1}^{n} i = \frac{1}{2}n(n+1).$$

Proof. (*Set-up*) We prove by induction. For each $n \in \mathbb{N}$, let $P(n)$ be the proposition

$$\sum_{i=1}^{n} i = \frac{1}{2}n(n+1).$$

(*Base Case*) Clearly $\sum_{i=1}^{1} i = 1 = \frac{1}{2}1(1+1)$, and so $P(1)$ is true.
(*Induction Step*) Assume that $P(n)$ is true for some fixed $n \geq 1$. We compute the sum of the first $n+1$ positive integers using our induction hypothesis $P(n)$ to simplify:

$$\begin{aligned}\sum_{i=1}^{n+1} i &= (n+1) + \sum_{i=1}^{n} i = (n+1) + \frac{1}{2}n(n+1) && \text{(by assumption of } P(n)\text{)}\\ &= \left(1 + \frac{1}{2}n\right)(n+1) = \frac{1}{2}(n+2)(n+1)\\ &= \frac{1}{2}(n+1)\big[(n+1)+1\big].\end{aligned}$$

This last says that $P(n+1)$ is true.
(*Conclusion*) By mathematical induction, we conclude that $P(n)$ is true for all $n \in \mathbb{N}$. That is

$$\forall n \in \mathbb{N}, \quad \sum_{i=1}^{n} i = \frac{1}{2}n(n+1).$$

■

Note how we grouped $\frac{1}{2}(n+1)\big[(n+1)+1\big]$ so that it is obviously the right hand side of $P(n+1)$.

Here is another example in the same vein, but done a little faster.[16]

Theorem 5.3. *Prove that $n(n+1)(2n+1)$ is divisible by 6 for all natural numbers n.*

Proof. We prove by induction. For each $n \in \mathbb{N}$, let $P(n)$ be the proposition

$n(n+1)(2n+1)$ is divisible by 6.

(*Base Case*) Clearly $1 \cdot (1+1) \cdot (2 \cdot 1 + 1) = 6$ is divisible by 6, hence $P(1)$ is true.
(*Induction Step*) Assume that $P(n)$ is true for some fixed $n \in \mathbb{N}$. Then

$$\begin{aligned}(n+1)(n+2)\big[2(n+1)+1\big] - n(n+1)(2n+1) &= (n+1)\big[(n+2)(2n+3) - n(2n+1)\big]\\ &= (n+1)(2n^2+7n+6-2n^2-n)\\ &= 6(n+1)^2.\end{aligned}$$

This is divisible by 6. Since, by the induction hypothesis, $n(n+1)(2n+1)$ is also divisible by 6, it

[16] The most common question after reading this proof is, 'How would I know to do that calculation?' It is better to think on how much scratch work was done before the originator stumbled on exactly this argument. Read more proofs and practice writing them, and you'll soon find that strategies like these will suggest themselves!

follows that

$$(n+1)(n+2)\big[2(n+1)+1\big] = n(n+1)(2n+1) + 6(n+1)^2$$

is divisible by 6, as required. Thus $P(n+1)$ is true.
By mathematical induction, $P(n)$ is true for all $n \geq 1$.

■

Theorem 5.3 is also true for $n = 0$, and indeed for *all* integers n. As we shall see in the next section, induction works perfectly well with any base case (say $n = 0$): you are not tied to $n = 1$. We can even modify induction to prove the result for the negative integers!

Here is another example, written in a more advanced style: we don't explicitly name $P(n)$, and the reader is expected to be familiar enough with induction to realize when we are covering the base case and induction step. If you find this proof a challenge, you should rewrite it in the same style as we used previously. Some assistance in this is given below.

Theorem 5.4. *For all $n \in \mathbb{N}$, $2 + 5 + 8 + \cdots + (3n-1) = \frac{1}{2}n(3n+1)$.*

Proof. For $n = 1$ we have $2 = 2$, hence the proposition holds. Now suppose the proposition holds for some fixed $n \in \mathbb{N}$. Then

$$\begin{aligned} 2 + 5 + \cdots + [3(n+1)-1] &= [2 + 5 + \cdots + (3n-1)] + 3n + 2 \\ &= \frac{1}{2}n(3n+1) + 3n + 2 = \frac{1}{2}(3n^2 + 7n + 4) \\ &= \frac{1}{2}(n+1)(3n+4) = \frac{1}{2}(n+1)\big[3(n+1)+1\big]. \end{aligned}$$

This says that the proposition holds for $n+1$. By mathematical induction the proposition holds for all $n \in \mathbb{N}$.

■

Scratch work is your friend! Once you are comfortable with the structure of an induction proof, the challenge is often in finding a clear argument for the induction step. Don't dive straight into the proof! First try some scratch calculations. Be creative, since the same approach will not work for all proofs.

One of the benefits of explicitly stating $P(n)$ is that it helps you to isolate what you know and to identify your goal. When stuck, write down both expressions $P(n)$ and $P(n+1)$ and you will often see how to proceed. Consider, for example, the proof of Theorem 5.4. We have:

$$\begin{aligned} P(n): &\quad 2 + 5 + 8 + \cdots + (3n-1) = \frac{1}{2}n(3n+1). \\ P(n+1): &\quad 2 + 5 + 8 + \cdots + [3(n+1)-1] = \frac{1}{2}(n+1)\big[3(n+1)+1\big] \end{aligned}$$

Simply by writing these down, we know that our goal is to somehow convert the left hand side of $P(n+1)$ into the right hand side, using $P(n)$.

As a final comment on scratch work, remember that it is *very unlikely* to constitute a proof. Here is a typical attempt at a proof of Theorem 5.4 by someone who is new to induction.

False Proof. $P(n+1)$: $\underbrace{2+5+\cdots+(3n-1)}_{=\frac{1}{2}n(3n+1) \text{ by } P(n)}+[3(n+1)-1] = \frac{1}{2}(n+1)\big[3(n+1)+1\big]$

$$= \frac{1}{2}(n+1)(3n+4)$$

$$\implies \quad \frac{3}{2}n^2+\frac{1}{2}n+3n+3-1 = \frac{1}{2}(3n^2+7n+4)$$

$$\implies \quad \frac{3}{2}n^2+\frac{7}{2}n+2 = \frac{3}{2}n^2+\frac{7}{2}n+2$$

■

Such an approach is likely to score zero in an exam! Here are some of the reasons why.

- $P(n+1)$ is the *goal*, the conclusion of the induction step. You cannot prove $P(n) \implies P(n+1)$ by *starting* with $P(n+1)$!
- More logically: the false proof says that something we don't know ($P(n) \wedge P(n+1)$) implies something true (the trivial final line). Since the implications $T \implies T$ and $F \implies T$ are both true, this tells us *nothing* about whether $P(n+1)$ is true.
- Reversing the arrows and turning the false proof upside down would be a start. However there is no explanation as to *why* the calculation is being done. The induction step is only part of an induction proof and it need to be placed and explained in context. More concretely:
  - There is no set-up. $P(n)$ has not been defined, neither indeed has n. You cannot use symbols in a proof unless they have been properly defined.
  - The base case is missing.
  - There is no conclusion. Indeed the word *induction* isn't mentioned: is the reader supposed to guess that we're doing induction?!

For all this negativity, there are some good things here. If you remove the $\implies$ symbols, you are left with an excellent piece of scratch work. By simplifying both sides of your goal you can more easily see how to calculate. For example, the expression $\frac{1}{2}(n+1)(3n+4)$ is an easier target to aim for when manipulating the left hand side of $P(n+1)$.

Your scratch work may make perfect sense to you, but if a reader cannot follow it without your assistance then it isn't a proof. The moral of the story is to do your scratch work for the induction step *then* lay out the structure of the proof (set-up, base case, etc.) before incorporating your calculation into a coherent and convincing argument.

Exercises

5.2.1 (a) Complete Gauss' direct proof of Theorem 5.2.

(b) Give a direct proof of Theorem 5.3.

(c) In Theorem 5.3, what is the proposition $P(n+1)$?

(d) In the Induction Step of Theorem 5.3, explain why it would be incorrect to write

$$P(n+1)-P(n) = (n+1)\big[(n+2)(2n+3)-n(2n+1)\big]$$

$$= (n+1)(2n^2+7n+6-2n^2-n)$$
$$= 6(n+1)^2.$$

5.2.2 Prove by induction that for each natural number n, we have $\sum_{j=0}^{n} 2^j = 2^{n+1} - 1$.

5.2.3 Consider the following Theorem:

If n is a natural number, then $\sum_{k=1}^{n} k^3 = \frac{1}{4}n^2(n+1)^2$.

(a) What explicitly is the meaning of $\sum_{k=1}^{4} k^3$?

(b) What would be meant by the expression $\sum_{k=1}^{n} n^3$, and why is it different to $\sum_{k=1}^{n} k^3$?

(c) If the Theorem is written in the form $\forall n \in \mathbb{N},\ P(n)$, what is the proposition $P(n)$?

(d) Give as many reasons as you can as to why the following 'proof' of the induction step is incorrect.

$$\begin{aligned} P(n+1) = \sum_{k=1}^{n+1} k^3 &= \frac{1}{4}(n+1)^2((n+1)+1)^2 \\ &= \sum_{k=1}^{n} k^3 + (n+1)^3 = \frac{1}{4}(n+1)^2(n+2)^2 \\ &= \frac{1}{4}n^2(n+1)^2 + (n+1)^3 = \frac{1}{4}(n+1)^2(n+2)^2 \\ &= \frac{1}{4}(n+1)^2\left[n^2+4(n+1)\right] = \frac{1}{4}(n+1)^2(n+2)^2 \\ &= \frac{1}{4}(n+1)^2(n+2)^2 = \frac{1}{4}(n+1)^2(n+2)^2 \end{aligned}$$

(e) Give a correct proof of the Theorem by induction.

5.2.4 (a) Prove by induction that $\forall n \in \mathbb{N}$ we have $3 \mid (2^n + 2^{n+1})$.

(b) Give a direct proof that $3 \mid (2^n + 2^{n+1})$ for all integers $n \geq 1$ *and* for $n = 0$.

(c) Look carefully at your proof for part (a). If you had started with the base case $n = 0$ instead of $n = 1$, would your proof still be valid?

5.2.5 Show by induction, that for every $n \in \mathbb{N}$ we have: $n \equiv 5 \pmod 3$ or $n \equiv 6 \pmod 3$ or $n \equiv 7 \pmod 3$.

5.2.6 Show, by induction, that for all $n \in \mathbb{N}$, 4 divides the integer $11^n - 7^n$.

5.2.7 (a) Find a formula for the sum of the first n odd natural numbers. Prove your assertion by induction.

(b) Give an alternative direct proof of your formula from part (a). You may use results such as $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$.

5.3 Well-ordering and the Principle of Mathematical Induction

Before seeing more examples of induction, it is worth thinking more carefully about the logic behind induction. The fact that induction really proves statements of the form $\forall n \in \mathbb{N},\ P(n)$ depends on a fundamental property of the natural numbers.

Definition 5.5. A set of real numbers A is *well-ordered* if every non-empty subset of A has a minimum element.

The definition is delicate: to test if a set A is well-ordered, we need to check *all* of its non-empty subsets. The definition could be written as follows:

$\forall B \subseteq A$ such that $B \neq \emptyset$, we have that $\min(B)$ exists.

Consequently, to show that a set A is *not* well-ordered, we need only exhibit a non-empty subset B which has *no minimum.*

Examples.

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?
2. The interval $[3, 10)$ is not well-ordered. Indeed $(3, 4)$ is a non-empty subset which has no minimum element.
3. The integers $\mathbb{Z}$ are not well-ordered, since there is no minimum integer.

More generally, every finite set of numbers is well-ordered, and intervals are not. Are there any *infinite* sets which are well-ordered? The answer is yes. Indeed it is part of the standard definition (Peano's Axioms) of the natural numbers that $\mathbb{N}$ is such a set.

Axiom. $\mathbb{N}$ is well-ordered.

Armed with this axiom, we can justify the method of proof by induction.

Theorem 5.6 (Principle of Mathematical Induction)**.** *Let $P(n)$ be a proposition for each $n \in \mathbb{N}$. Suppose:*

(a) $P(1)$ *is true.*

(b) $\forall n \in \mathbb{N},\ P(n) \implies P(n+1)$.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Proof. We argue by contradiction. Assume that conditions (a) and (b) hold and that $\exists n \in \mathbb{N}$ such that $P(n)$ is *false.* Then the set

$$S = \{k \in \mathbb{N} : P(k) \text{ is false}\}$$

is a non-empty subset of the well-ordered set $\mathbb{N}$. It follows that S has a minimum element $m = \min(S)$. Note that $P(m)$ is *false.*
Clearly $m \neq 1$, since $P(1)$ is true (condition (a)). Therefore $m \geq 2$ and so $m - 1 \in \mathbb{N}$.
Since $m = \min(S)$ it follows that $m - 1 \notin S$ and so $P(m-1)$ must be *true.*

Now condition (b) forces $P(m)$ to be *true.* A contradiction.
We conclude that $P(n)$ is true for all $n \in \mathbb{N}$.

■

Different Base Cases

An induction argument need not begin with the case $n = 1$. By proving Theorem 5.6 it should be clear where we used the well-ordering of $\mathbb{N}$. Now fix an integer m (positive, negative or zero) and consider the set

$$\mathbb{Z}_{\geq m} = \{n \in \mathbb{Z} : n \geq m\} = \{m, m+1, m+2, m+3, \ldots\}.$$

This set is well-ordered, whence the following modification of the induction principle is immediate.

Corollary 5.7. *Fix $m \in \mathbb{Z}$. Let $P(n)$ be a proposition for each integer $n \geq m$. Suppose:*

(a) $P(m)$ is true.

(b) $\forall n \geq m,\ P(n) \implies P(n+1)$.

Then $P(n)$ is true for all $n \geq m$.

We are simply changing the base case. The induction concept is exactly the same as before:

$$P(m) \implies P(m+1) \implies P(m+2) \implies P(m+3) \implies \cdots$$

As long as you explicitly prove the first claim in the sequence, and you show the induction step, then all the propositions are true.

Here is an example where we begin with $n = 4$.

Theorem 5.8. *For all integers $n \geq 4$, we have $3^n > n^3$.*

Proof. We prove by induction. The first case of interest is $n = 4$, so we choose this to be our base case.
(*Base Case*) If $n = 4$ we have $3^n = 81 > 64 = n^3$. The proposition is therefore true for $n = 4$.
(*Induction Step*) Fix $n \in \mathbb{Z}_{\geq 4}$ and suppose that $3^n > n^3$. Then

$$3^{n+1} = 3 \cdot 3^n > 3n^3.$$

To finish the proof, we want to see that this right hand side is at least $(n+1)^3$. Now

$$3n^3 \geq (n+1)^3 \iff 3 \geq \left(1 + \frac{1}{n}\right)^3$$

This is true for $n = 3$ and, since the right hand side is decreasing as n increases, it is certainly true when $n \geq 4$. We therefore conclude that

$$3^n > n^3 \implies 3^{n+1} > (n+1)^3$$

which is the induction step.
By induction, we have shown that $3^n > n^3$ whenever $n \in \mathbb{Z}_{\geq 4}$. ∎

Our next example is reminiscent of sequences and series from elementary calculus. If you follow the derivation of such a formula given in an elementary calculus text, you'll probably see liberal use of ellipsis dots (...). When you see ellipses in a proof, it is often because the author is hiding an induction argument.

Theorem 5.9. *For all integers $n \geq 3$, we have*

$$\sum_{i=3}^{n} \frac{1}{i(i-2)} = \frac{3}{4} - \frac{2n-1}{2n(n-1)}. \qquad (*)$$

Proof. We prove by induction.

(*Base Case*) When $n = 3$, $(*)$ reads $\sum_{i=3}^{3} \frac{1}{i(i-2)} = \frac{3}{4} - \frac{5}{12}$. Both sides are equal to $\frac{1}{3}$, and so $(*)$ is true.

(*Induction Step*) Assume that $(*)$ is true for some fixed $n \geq 3$. Then

$$\begin{aligned}
\sum_{i=3}^{n+1} \frac{1}{i(i-2)} &= \sum_{i=3}^{n} \frac{1}{i(i-2)} + \frac{1}{(n+1)(n-1)} \\
&= \frac{3}{4} - \frac{2n-1}{2n(n-1)} + \frac{1}{(n+1)(n-1)} && \text{(by the induction hypothesis)} \\
&= \frac{3}{4} - \left[\frac{(2n-1)(n+1) - 2n}{2(n+1)n(n-1)}\right] = \frac{3}{4} - \left[\frac{1+n-2n^2}{2(n+1)n(n-1)}\right] \\
&= \frac{3}{4} + \frac{(2n+1)(1-n)}{2(n+1)n(n-1)} = \frac{3}{4} - \frac{2n+1}{2(n+1)n}
\end{aligned}$$

which is exactly $(*)$ when n is replaced by $n+1$.
By induction $(*)$ holds for all integers $n \geq 3$. ∎

Our final example involves a little abstraction.

Theorem 5.10. *The interior angles of an n-gon (n-sided polygon) sum to $(n-2)\pi$ radians.*

The challenge here is to set up the induction step properly. We will take the initial case ($n = 3$) that the angles of a triangle sum to π radians as given,[17] and merely prove the induction step. The main logical difficulty comes from the fact that we must consider *all* n-gons simultaneously. If we were to write the induction step in the form

$$\forall n \in \mathbb{Z}_{\geq 3},\ P(n) \implies P(n+1),$$

[17] Can you supply a direct proof of this fact?

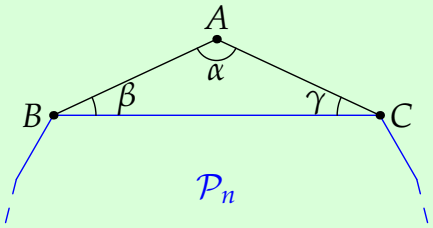
then the proposition $P(n)$ would be

$$P(n): \quad \forall n\text{-gons } \mathcal{P}_n, \text{ the sum of the interior angles of } \mathcal{P}_n \text{ is } (n-2)\pi \text{ radians.}$$

To prove our induction step for a *fixed* integer n, we must show that *all* $(n+1)$-gons have the correct sum of interior angles. We therefore assume that we are given some $(n+1)$-gon $\mathcal{P}_{n+1}$ and proceed to compute its interior angles in terms of a related n-gon.

Proof. Fix an integer $n \geq 3$, and suppose that *all* n-gons have interior angles summing to $(n-2)\pi$ radians. Suppose we are given an $(n+1)$-gon $\mathcal{P}_{n+1}$. Select a vertex A, and label the adjacent vertices B and C. Delete A, and join B and C with a straight edge. The result is an n-gon $\mathcal{P}_n$. There are two cases to consider.[18]

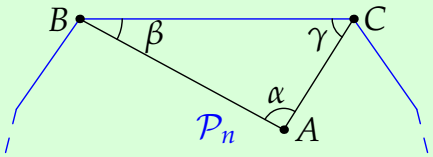
Case 1: The deleted point A is *outside* $\mathcal{P}_n$. The sum of the interior angles of $\mathcal{P}_{n+1}$ exceeds those of P_n by the $\alpha + \beta + \gamma = \pi$ radians of the triangle $\triangle ABC$. Therefore $\mathcal{P}_{n+1}$ has interior angles summing to $(n-2)\pi + \pi = [(n+1)-2]\pi$ radians.



Case 1: A outside $\mathcal{P}_n$

Case 2: The deleted point A is *inside* $\mathcal{P}_n$. To obtain the sum of the interior angles of $\mathcal{P}_{n+1}$, we take the sum of the interior angles of $\mathcal{P}_n$ and do three things:

- *Subtract* β
- *Subtract* γ
- *Add the reflex angle* $2\pi - \alpha$ *at* A



Case 2: A inside $\mathcal{P}_n$

We are therefore adding an additional

$$-\beta - \gamma + (2\pi - \alpha) = 2\pi - (\alpha + \beta + \gamma) = 2\pi - \pi = \pi$$

radians. $\mathcal{P}_{n+1}$ again has interior angles summing to $[(n+1)-2]\pi$ radians.
Note that if A was on the edge of $\mathcal{P}_n$, then our original polygon $\mathcal{P}_{n+1}$ would have had only n sides. ∎

[18] We are obscuring two subtleties here. It is a fact, though not an obvious one, that it is always possible to choose a vertex A so that the new polygon $\mathcal{P}_n$ doesn't cross itself. Read about 'ears' and 'mouths' of polygons and triangulation if you're interested. There are also two other, less likely, cases, where deleting a point from an $(n+1)$-gon leaves you with an $(n-1)$-gon, or even an $(n-2)$-gon. To think it out, try drawing a 12-gon in the shape of a Star of David. Deleting one of the outer corners creates a 9-gon! Dealing with these cases strictly requires strong induction, so we return to them later.

Aside: Well-ordering more generally

Well-ordering is a fundamental concept whose implications are far beyond what we're discussing here. Informally speaking, *well-ordering* a set A involves listing the elements of A in some order so that every non-empty subset of A has a first element *with respect to that order.*
Consider, for example, the set of negative integers $\mathbb{Z}^-$. For the purposes of these notes we will always consider the standard ordering:

$$\cdots < -4 < -3 < -2 < -1.$$

Written in the standard order, $\mathbb{Z}^- = \{\ldots, -4, -3, -2, -1\}$ is *not* a well-ordered set. In more advanced logic course one could consider alternative orderings, and the definition of well-ordered would change accordingly. If we choose the alternative ordering

$$\mathbb{Z}^- = \{-1, -2, -3, -4, \cdots\}, \qquad (*)$$

then $\mathbb{Z}^-$ would be well-ordered: if $B \subseteq \mathbb{Z}^-$ is non-empty and has its elements listed in the same order as $(*)$, then B has a first element. Since the principle of mathematical induction depends only on us having a well-ordered set, we are now permitted to prove theorems of the form $\forall n \in \mathbb{Z}^-,\ P(n)$, by induction. The base case is $n = -1$ and the induction step justifies the chain

$$P(-1) \implies P(-2) \implies P(-3) \implies \cdots$$

An extremely important theorem in advanced set theory states that it is possible to well-order *every* set. With a slight modification of the process, this massively increases the applicability of induction. In these notes we keep things simple: well-ordering is always in the sense of Definition 5.5, where we list the elements of a set in the usual increasing order.

Exercises

5.3.1 Consider the following Theorem. 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 Theorem is written in the form $\forall n \in \mathbb{N}_{\geq 2},\ P(n)$, what is $P(n)$?

(b) Π-notation is used for products in the same way as Σ-notation for sums: for example

$$\prod_{k=1}^{5}(k+1)^k = 2^1 \cdot 3^2 \cdot 4^3 \cdot 5^4 \cdot 6^5$$

Rewrite the statement of the Theorem using Π-notation.

(c) Prove the Theorem by induction (you may use whatever notation you wish).

5.3.2 Recall the geometric series formula from calculus: if $r \neq 1$ is constant, and $n \in \mathbb{N}_0$, then

$$\sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r} \qquad (*)$$

(a) Here is an incorrect proof by induction. Explain why it is incorrect.

Proof. Let $P(n) = \sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r}$.

(*Base Case* $n = 0$) $P(0) = \sum_{k=0}^{0} r^k = r^0 = 1 = \frac{1-r^{0+1}}{1-r}$ is true.

(*Induction Step*) Fix $n \in \mathbb{N}_0$ and assume that $P(n)$ is true. Then

$$P(n+1) = \sum_{k=0}^{n+1} r^k = \sum_{k=0}^{n} r^k + r^{n+1} = \frac{1 - r^{n+1}}{1 - r} + r^{n+1}$$

$$= \frac{1-r^{n+1}}{1-r} + \frac{r^{n+1}-r^{n+2}}{1-r} = \frac{1-r^{n+2}}{1-r}, \text{ is true.}$$

By induction, $(*)$ is true for all $n \in \mathbb{N}_0$. ■

(b) Give a correct proof of $(*)$.

5.3.3 Here is an argument attempting to justify $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1) + 7$. What is wrong with it?

Assume that the statement is true for some fixed n. Then

$$\sum_{i=1}^{n+1} i = \sum_{i=1}^{n} i + (n+1) = \frac{1}{2}n(n+1) + 7 + (n+1) = \frac{1}{2}(n+1)[(n+1)+1] + 7,$$

hence the statement is true for $n+1$ and, by induction, for all $n \in \mathbb{N}$.

5.3.4 Consider the following 'proof' that all human beings have the same age. Where is the flaw in the argument.

Proof. (*Base case* $n = 1$) Clearly, in a set with only 1 person, all the people in the set have the same age.
(*Inductive hypothesis*) Suppose that for some integer $n \geq 1$ and for all sets with n people, it is true that all of the people in the set have the same age.
(*Inductive step*) Let A be a set with $n+1$ people, say $A = \{a_1, \ldots, a_n, a_{n+1}\}$, and let

$$A' = \{a_1, \ldots, a_n\} \quad \text{and} \quad A'' = \{a_2, \ldots, a_{n+1}\}.$$

The inductive hypothesis tells us that all the people in A' have the same age and all the people in A'' have the same age. Since a_2 belongs to both sets, then all the people in A have the same age as a_2. We conclude that all the people in A have the same age.
(*Conclusion*) By induction, the claim holds for all $n \geq 1$. ■

5.3.5 Let $P(n)$ and $Q(n)$ be propositions for each $n \in \mathbb{N}$.

(a) Assume that m is the smallest natural number such that $P(m)$ is false. Let

$$A = \{n \in \mathbb{N} : n < m\}.$$

What can you say about the elements in the set A, with respect to the property P?

(b) Assume that a is the smallest natural number such that $P(a) \vee Q(a)$ is false. Let

$$B = \{n \in \mathbb{N} : n < a\}.$$

What can you say about the elements in the set B, with respect to the properties P and Q?

(c) Assume that u is the smallest natural number such that $P(u) \wedge Q(u)$ is false. Let

$$C = \{n \in \mathbb{N} : n < u\}.$$

What can you say about the elements in the set C, with respect to the properties P and Q?

(d) Assume that $P(1)$ is true, but that '$\forall n \in \mathbb{N},\ P(n)$' is false. Show that there exists a natural number k such that the implication $P(k) \implies P(k+1)$ is false.

5.3.6 Prove that if $A \subseteq \mathbb{R}$ is a *finite* set, then A is well-ordered.

5.3.7 In this question we use the fact that $\mathbb{N}_0$ is well-ordered to prove the Division Algorithm (Theorem 3.2).

If $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, then $\exists$ unique $q, r \in \mathbb{Z}$ such that $m = qn + r$ and $0 \leq r < n$.

Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ be given, and define $S = \{k \in \mathbb{N}_0 : k = m - qn \text{ for some } q \in \mathbb{Z}\}$.

(a) Show that S is a *non-empty* subset of $\mathbb{N}_0$.

(b) $\mathbb{N}_0$ is well-ordered. By part (a), S has a minimal element r. Prove that $0 \leq r < n$.

(c) Suppose that there are two pairs of integers (q_1, r_1) and (q_2, r_2) which satisfy $m = q_i n + r_i$. Prove that $r_1 = r_2$ and, consequently, that the division algorithm is true.

5.3.8 In this question we consider Peano's Axioms for the natural numbers:

Initial element: $1 \in \mathbb{N}$

Successor elements: There is a *successor function* $f : \mathbb{N} \to \mathbb{N}$. For each $n \in \mathbb{N}$, the successor $f(n)$ is also a natural number.

No predecessor of 1 $\forall n \in \mathbb{N}, f(n) = 1$ is false.

Unique predecessor: f is injective: $f(n) = f(m) \implies m = n$.

Induction: If $A \subseteq \mathbb{N}$ has the following properties:

- $1 \in A$,
- $\forall a \in A, f(a) \in A$,

then $A = \mathbb{N}$.

The successor function f is simply 'plus one' in disguise: $f(n) = n + 1$.

(a) Suppose you replace $\mathbb{N}$ with $\mathbb{Z}$ in each of the above axioms. Which axioms are still true and which are false?

(b) Here we use the notation (m, n) to represent a pair of natural numbers. Let T be the set of all pairs

$$T = \{(m, n) : m, n \in \mathbb{N}\}.$$

Let $f : T \to T$ be the function $f(m, n) = (m + 1, n)$. Letting the pair $(1, 1)$ play the role of '1' in Peano's axioms, and f be the successor function, decide which of the above axioms are satisfied by T.

(c) (Hard!) With the same set T as in part (b), take the successor function $f : T \to T$ to be

$$f(m, n) = \begin{cases} (m - 1, n + 1) & \text{if } m \geq 2, \\ (m + n, 1) & \text{if } m = 1. \end{cases}$$

Which of the above axioms are satisfied by T and f?

5.3.9 (*Ignore this question if you haven't studied matrices*) Suppose that $A = \left(\begin{smallmatrix} 7 & 12 \\ -2 & -3 \end{smallmatrix}\right)$. We prove that

$$\forall n \in \mathbb{Z}, \quad A^n = \begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix} + 3^n \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}. \tag{†}$$

Here $A^{-n} = (A^n)^{-1}$ is the inverse of A^n, and we follow the convention that $A^0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is the identity matrix.

(a) Prove by induction that (†) holds $\forall n \in \mathbb{N}_0$.

(b) Modify your argument in part (a) to prove that (†) holds $\forall n \in \mathbb{Z}_0^-$. (*Use the fact that, when written in reverse order,* $\mathbb{Z}_0^- = \{0, -1, -2, -3, -4, \ldots\}$ *is a well-ordered set.)*

(c) Using what you know about matrix inverses, give a direct proof that (†) holds $\forall n \in \mathbb{Z}_0^-$. (*If* C *and* D *are* 2×2 *matrices such that* $CD = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$*, then* $D = C^{-1}$.)

(d) Diagonalize the matrix A and thereby give a direct proof of (†) for all integers n.

5.4 Strong Induction

The principle of mathematical induction as represented in Theorem 5.6 is sometimes known as *weak* induction. In weak induction, the induction step requires only that one proposition $P(n)$ is true to demonstrate the truth of $P(n+1)$. By contrast, the induction step in *strong* induction additionally requires that some, perhaps *all*, of the propositions coming before $P(n)$ are also true.

Theorem 5.11 (Principle of Strong Induction)**.** *Let m be an integer and suppose that $P(n)$ is a proposition for each $n \in \mathbb{Z}_{\geq m}$. Also fix an integer $l > m$. Suppose:*

(a) $P(m), P(m+1), \ldots, P(l)$ *are true.*

(b) $\forall n \geq l,\ (P(m) \wedge P(m+1) \wedge \cdots \wedge P(n)) \implies P(n+1)$.

Then $P(n)$ is true for all $n \in \mathbb{Z}_{\geq m}$.

The statement is a little complicated: what matters is that $\mathbb{Z}_{\geq m}$ is a well-ordered set. In the simplest examples, we have $m = 1$ and $\mathbb{Z}_{\geq 1} = \mathbb{N}$. The challenge in strong induction is identifying how many base cases $l - m + 1$ are needed.

To see this in action, consider the Fibonacci numbers: an excellent source of strong induction examples.

Definition 5.12. The *Fibonacci numbers* are the sequence $(f_n)_{n=1}^{\infty}$ defined by the recurrence relation

$$\begin{cases} f_{n+1} = f_n + f_{n-1} & \text{if } n \geq 2, \\ f_1 = f_2 = 1 \end{cases} \qquad (*)$$

Theorem 5.13. $\forall n \in \mathbb{N}$, $f_n < 2^n$.

Proof. For each natural number n, let $P(n)$ be the proposition $f_n < 2^n$.
(*Base cases* $n = 1, 2$) $f_1 = 1 < 2^1$ and $f_2 = 1 < 2^2$, whence $P(1)$ and $P(2)$ are true.
(*Induction step*) Fix $n \geq 2$ and suppose that $P(1), \ldots, P(n)$ are true. Then

$$f_{n+1} = f_n + f_{n-1} < 2^n + 2^{n-1} < 2^n + 2^n = 2^{n+1}$$

which says that $P(n+1)$ is true.
By strong induction $P(n)$ is true for all $n \in \mathbb{N}$, and so $f_n < 2^n$. ∎

In terms of Theorem 5.11, we have $m = 1$ and $l = 2$ with $m - l + 1 = 2$ base cases. Te reason we need $m = 1$ is because the first claim in the Theorem is about the integer 1, namely $f_1 < 2^1$. We need two base cases because the recurrence relation $(*)$ defining the Fibonacci numbers requires the previous *two* terms of the sequence to construct the next.

To help understand strong induction, it is instructive to see why a proof by weak induction would fail in this setting.

Wrong Proof A. We show, by weak induction, that $\forall n \in \mathbb{N},\ f_n < 2^n$.
(*Base Case* $n = 1$) By definition, $f_1 = 1 < 2^1$, whence the claim is true for $n = 1$.
(*Induction Step*) Fix $n \in \mathbb{N}$ and assume that $f_n < 2^n$. We want to show that $f_{n+1} < 2^{n+1}$. By the recurrence relation, we can write

$$f_{n+1} = f_n + f_{n-1}. \tag{$*$}$$

The inductive hypothesis tells us that $f_n < 2^n$, but what can we say about f_{n-1}? Absolutely nothing! We are stuck: weak induction fails to prove the theorem.

■

The incorrect proof tells us why we need strong induction: the recurrence relation defines each Fibonacci number (except f_1 and f_2) in terms of *the previous two.* To make use of the recurrence, our induction hypothesis must assume something about *at least* f_n *and* f_{n-1}. Assuming something about only f_n is not enough.

From *Wrong Proof A* we learned that we needed to prove by strong induction. Now suppose that we try the following, which looks almost identical to the correct proof.

Wrong Proof B. For each $n \in \mathbb{N}$, let $P(n)$ be the proposition $f_n < 2^n$. We prove that $P(n)$ is true for all $n \in \mathbb{N}$ by strong induction.
(*Base Case* $n = 1$) By definition, $f_1 = 1 < 2^1$, whence $P(1)$ is true.
(*Induction Step*) Fix $n \in \mathbb{N}$ and assume that $P(1), \ldots, P(n)$ are all true. We want to show that $f_{n+1} < 2^{n+1}$. By the recurrence relation, we can write

$$f_{n+1} = f_n + f_{n-1} < 2^n + 2^{n-1} < 2 \cdot 2^n = 2^{n+1}. \tag{$\dagger$}$$

Hence $P(n)$ is true for all $n \geq 1$.

■

Where is the problem with this second incorrect proof? The recursive formula $f_{n+1} = f_n + f_{n-1}$ *only* applies if $n \geq 2$. If we take $n = 1$, then it reads $f_2 = f_1 + f_0$, but f_0 is not defined! In the induction step of *Wrong Proof B*, we are letting n be any integer ≥ 1. When $n = 1$ the step ($\dagger$) is not justified, and so the proof fails. For ($\dagger$) to be legitimate, we must have $n \geq 2$. This is why, in our correct proof, we had to prove $P(1)$ *and* $P(2)$ separately.

The moral here is to try the induction step as scratch work. Your attempt will tell you *if* you need strong induction and, if you do, *how many* base cases are required.

Strong Induction on Well-ordered Sets

In the next example the first term is suffixed by $n = 0$. In the language of Theorem 5.11, we have $m = 0$ and $l = 1$ with $m - l + 1 = 2$ base cases. Just like the Fibonacci example, two base cases are required because the defining recurrence relation constructs the next term in the sequence from the two previous terms.

Theorem 5.14. *A sequence of integers $(a_n)_{n=0}^{\infty}$ is defined by*

$$\begin{cases} a_n = 5a_{n-1} - 6a_{n-2}, & n \geq 2, \\ a_0 = 0,\ a_1 = 1. \end{cases}$$

Then $a_n = 3^n - 2^n$ for all $n \in \mathbb{N}_0$.

Proof. We prove by strong induction.
(*Base cases* $n = 0, 1$) The formula is true in both cases: $a_0 = 0 = 3^0 - 2^0$ and $a_1 = 1 = 3^1 - 2^1$.
(*Induction step*) Fix an integer $n \geq 1$ and suppose that $a_k = 3^k - 2^k$ for all $k \leq n$. Then

$$\begin{aligned} a_{n+1} &= 5a_n - 6a_{n-1} = 5(3^n - 2^n) - 6(3^{n-1} - 2^{n-1}) \\ &= (15 - 6)3^{n-1} + (10 - 6)2^{n-1} = 3^{n+1} - 2^{n+1}. \end{aligned}$$

By strong induction $a_n = 3^n - 2^n$ is true for all $n \in \mathbb{N}_0$. ∎

Think about why we wrote $a_{n+1} = 5a_n - 6a_{n-1}$ in the induction step, whereas the statement in the Theorem reads $a_n = 5a_{n-1} - 6a_{n-2}$. Does it matter? What does it mean to say that n is a 'dummy variable'?

In the two previous examples, it might seem that strong induction is something of a logical overkill. In the induction step we are assuming far more than we need. In both examples, establishing the truth of $P(n+1)$ required only the truth of $P(n)$ and $P(n-1)$. We assumed that the earlier propositions were also true, but we never used them. Depending on the proof, you might need two, three or even all of the propositions prior to $P(n+1)$ to complete the induction step. Once you are used to strong induction you may feel comfortable slimming a proof down so that you only mention precisely what you need. For the present, the way we've stated the principle is maximally safe! For some practice with this, see Exercise 5.4.3 where *three* base cases are needed, and the induction step requires the *three* previous propositions $P(n), P(n-1), P(n-2)$ to $P(n+1)$.

In order to see strong induction in all its glory, where the induction step requires *all* of the previous propositions, we prove part of the famous Fundamental Theorem of Arithmetic which states that all natural numbers may be factored into a product of primes: for example $3564 = 2^2 \times 3^4 \times 11$.

Definition 5.15. $p \in \mathbb{N}_{\geq 2}$ is *prime* if its only positive divisors are itself and 1. If $q \in \mathbb{N}_{\geq 2}$ is not prime, then it is *composite:* $\exists a, b \in \mathbb{N}_{\geq 2}$ such that $q = ab$.

As you read the proof, think carefully about why *only one* base case is required.

Theorem 5.16. *Every natural number $n \geq 2$ is either prime, or a product of primes.*

Proof. We prove by strong induction.
(*Base case* $n = 2$) The only positive divisors of 2 are itself and 1, hence 2 is prime.
(*Induction step*) Fix $n \in \mathbb{N}_{\geq 2}$ and assume that *every* natural number k satisfying $2 \leq k \leq n$ is either prime or a product of primes. There are two possibilities:

- $n+1$ is prime. In this case we are done.
- $n+1$ is composite. Thus $n+1 = ab$ for some natural numbers $a, b \geq 2$. Clearly $a, b \leq n$, and so, by the induction hypothesis, *both* are prime or the product of primes. Therefore $n+1$ is also the product of primes.

By strong induction we see that all natural numbers $n \geq 2$ are either prime, or a product of primes. ∎

Exercises

5.4.1 Define a sequence $(b_n)_{n=1}^{\infty}$ as follows:

$$\begin{cases} b_n = b_{n-1} + b_{n-2}, & n \geq 3, \\ b_1 = 3,\ b_2 = 6. \end{cases}$$

Prove: $\forall\, n \in \mathbb{N},\ b_n$ is divisible by 3.

5.4.2 Consider the proof of Theorem 5.16.

(a) If the Theorem is written in the form $\forall n \in \mathbb{N}_{\geq 2}, P(n)$, what is the proposition $P(n)$?

(b) Explicitly carry out the induction step for the three situations $n+1 = 9$, $n+1 = 106$ and $n+1 = 45$. How many different ways can you perform the calculation for $n+1 = 45$? Explain why it is only necessary in the induction step to assume that all integers k satisfying $2 \leq k \leq \frac{n+1}{2}$ are prime or products of primes.

(c) Rewrite the proof in the style of Theorem 5.13, explicitly mentioning the propositions $P(n)$, and thus making the logical flow of strong induction absolutely clear.

5.4.3 Define a sequence $(c_n)_{n=0}^{\infty}$ as follows:

$$\begin{cases} c_{n+1} = \frac{49}{8} c_n - \frac{225}{8} c_{n-2}, & n \geq 2, \\ c_0 = 0, c_1 = 2, c_2 = 16. \end{cases}$$

Prove that $c_n = 5^n - 3^n$ for all $n \in \mathbb{N}_0$. *Hint: you need three base cases!*

5.4.4 Prove that the nth Fibonacci number f_n is given by the formula

$$f_n = \frac{\phi^n - \hat{\phi}^n}{\sqrt{5}}, \quad \text{where} \quad \phi = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \hat{\phi} = \frac{1-\sqrt{5}}{2}.$$

ϕ is the famous Golden ratio. ϕ and $\hat{\phi}$ are the two solutions to the equation $\phi = 1 + \phi^{-1}$.

5.4.5 In this question we use an alternative definition of prime.[19]

Definition. $p \in \mathbb{N}_{\geq 2}$ is *prime* if $\forall a, b \in \mathbb{N},\ p \,|\, ab \implies p \,|\, a$ or $p \,|\, b$.

Let p be prime, let $n \in \mathbb{N}$, and let $a_1, \ldots, a_n$ be natural numbers such that p divides the product $a_1 a_2 \cdots a_n$. Prove by induction that,

$$\exists i \in \{1, 2, \ldots, n\} \text{ such that } p \,|\, a_i.$$

Hint: you need to cover two *base cases. Why? Think about the induction step first and it will help you decide how many base cases you need.*

5.4.6 Show that for every positive integer n, $(3+\sqrt{5})^n + (3-\sqrt{5})^n$ is an even integer.
Hints: Prove simultaneously that $(3+\sqrt{5})^n - (3-\sqrt{5})^n$ *is an even multiple of* $\sqrt{5}$.
Subtract the nth expression from the $(n+1)$*th in both cases...*

5.4.7 (Hard!) Return to the proof of Theorem 5.10. Can you make a watertight argument using strong induction that also covers the two missing cases? Draw a picture to illustrate each case.

[19] Strictly this is what it means for p to be *irreducible.* In the ring of integers, *prime* and *irreducible* are synonymous. For the details, take a Number Theory course.

6 Set Theory, Part II

In this chapter we return to set theory, where we consider 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). The basic idea is that each of the co-ordinates x and y is a member of the set $\mathbb{R}$.

Definition 6.1. Let A and B be sets. The *Cartesian product* of A and B is the set

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

$A \times B$ is exactly the set of *ordered pairs* (a, b).

Examples. 1. The Cartesian product of the real line $\mathbb{R}$ with itself is the xy-plane: rather than writing $\mathbb{R} \times \mathbb{R}$ which is unwieldy, we write $\mathbb{R}^2$.

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) : x, y \in \mathbb{R}\}.$$

More generally, $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \mathbb{R}}_{n \text{ times}}$ is the set of n-tuples of real numbers:

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{R}\}.$$

2. Suppose you go to a restaurant where you have a choice of one main course and one side. The menu might be summarized set-theoretically: consider the sets

$$\text{Mains} = \{\text{fish, steak, eggplant, pasta}\}$$
$$\text{Sides} = \{\text{asparagus, salad, potatoes}\}$$

The Cartesian product Mains$\times$Sides is the set of all possible meals made up of one main and one side. It should be obvious that there are $4 \times 3 = 12$ possible meal choices.

This last example illustrates the following theorem. Indeed it partly explains the use of the word *product* in the definition.

Theorem 6.2. *If A and B are finite sets, then $|A \times B| = |A| \cdot |B|$.*

Proof. Label the elements of each set and list the elements of $A \times B$ lexicographically. If $|A| = m$ and $|B| = n$, then we have:

$$\begin{array}{cccc} (a_1,b_1) & (a_1,b_2) & (a_1,b_3) & \cdots & (a_1,b_n) \\ (a_2,b_1) & (a_2,b_2) & (a_2,b_3) & \cdots & (a_2,b_n) \\ \vdots & \vdots & \vdots & & \vdots \\ (a_m,b_1) & (a_m,b_2) & (a_m,b_3) & \cdots & (a_m,b_n) \end{array}$$

It should be clear that every element of $A \times B$ is listed exactly once. There are m rows and n columns, thus $|A \times B| = mn$.

■

Before we go any further, consider the complement of a Cartesian product $A \times B$. If you had to guess an expression for $(A \times B)^C$, you might well try $A^C \times B^C$. Let us think more carefully.

$$\begin{aligned} (x,y) \in (A \times B)^C &\iff (x,y) \notin A \times B \\ &\iff \neg((x,y) \in A \times B) \\ &\iff \neg(x \in A \text{ and } y \in B) \\ &\iff x \notin A \text{ or } y \notin B \end{aligned}$$

Since the definition of Cartesian product involves *and*, its negation, by De Morgan's laws, involves *or*. It follows that the complement of a Cartesian product is *not a Cartesian product!*

As an example of a basic set relationship involving Cartesian products, we prove a theorem.

Theorem 6.3. *Let A, B, C, D be sets. Then $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.*

Proof. Since we are dealing with Cartesian products, the general element has the form (x, y).

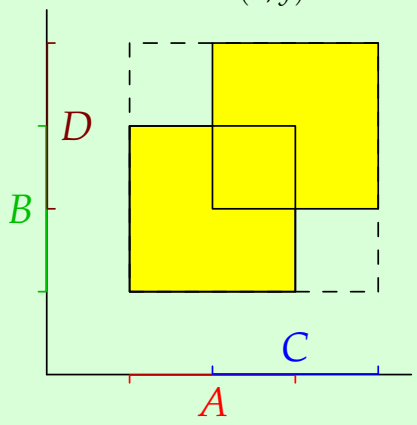
Let $(x,y) \in (A \times B) \cup (C \times D)$. Then

$$(x,y) \in A \times B \qquad \text{or} \qquad (x,y) \in C \times D.$$

But then

$$(x \in A \text{ and } y \in B) \qquad \text{or} \qquad (x \in C \text{ and } y \in D).$$

Clearly $x \in A$ or $x \in C$, so $x \in A \cup C$.
Similarly $y \in B$ or $y \in D$, so $y \in B \cup D$.
Therefore $(x,y) \in (A \cup C) \times (B \cup D)$, as required.



■

The picture is an imagining of the theorem, where we assume that the sets A, B, C and D are all intervals of real numbers. $(A \times B) \cup (C \times D)$ is the yellow shaded region, while $(A \cup C) \times (B \cup D)$

is the larger dashed square. Be careful with pictures! The theorem is a statement about *any* sets, whereas the picture implicitly assumes that these sets are intervals. While helpful, the picture is *not* a proof!

Either by carefully reading the proof or by thinking about the picture, you should be convinced that the two sets in the theorem are not equal (in general): if $x \in (A \setminus C)$ and $y \in D$, then (x, y) is an element of the right hand side, but not the left. Is it clear where the point (x, y) lives in the picture?

Exercises

6.1.1 Consider the following subintervals of the real line: $A = [2, 5],\ B = (0, 4)$.

(a) Express the set $(A \setminus B)^{\mathsf{C}}$ in interval notation, as a disjoint union of intervals.

(b) Draw a picture of the set $(A \setminus B)^{\mathsf{C}} \times (B \setminus A)$.

6.1.2 Rewrite the condition

$$(x, y) \in (A^{\mathsf{C}} \cup B) \times (C \setminus D)$$

in terms of (some of) the following propositions:

$$x \in A, \quad x \notin A, \quad x \in B, \quad x \notin B, \quad y \in C, \quad y \notin C, \quad y \in D, \quad y \notin D.$$

6.1.3 Let $A = [1, 3]$, $B = [2, 4]$ and $C = [2, 3]$. *Prove or disprove* that

$$(A \times B) \cap (B \times A) = C \times C.$$

Hint: Draw the sets $A \times B$, $B \times A$ and $C \times C$ in the Cartesian plane. The picture will give you a hint on whether or not the statement is true, but it does not constitute a proof.

6.1.4 A straight line subset of the plane $\mathbb{R}^2$ is a subset of the form

$$A_{a,b,c} = \{(x, y) : ax + by = c\}, \quad \text{for some constants } a, b, c, \text{ with } ab \neq 0.$$

(a) Draw the set $A_{1,2,3}$. Is it a Cartesian product?

(b) Which straight line subsets in the plane $\mathbb{R}^2$ are Cartesian products? Otherwise said, find a condition on the constants a, b, c for which the set $A_{a,b,c}$ is a Cartesian product.

6.1.5 Draw a picture, similar to that in Theorem 6.3, which illustrates the fact that

$$(A \times B)^{\mathsf{C}} \neq A^{\mathsf{C}} \times B^{\mathsf{C}}.$$

Using your picture, write the set $(A \times B)^{\mathsf{C}}$ in the form

$$(C_1 \times D_1) \cup (C_2 \times D_2) \cup \cdots$$

where each of the unions are *disjoint:* that is $i \neq j \implies (C_i \times D_i) \cap (C_j \times D_j) = \emptyset$. You don't have to prove your assertion.

6.1.6 Let $E \subseteq \mathbb{N} \times \mathbb{N}$ be the smallest subset which satisfies the following conditions:

- Base case: $(1, 1) \in E$

- Generating Rule I: If $(a, b) \in E$ then $(a, a + b) \in E$
- Generating Rule II: If $(a, b) \in E$ then $(b, a) \in E$

(a) Show in detail that $(4, 3) \in E$.

(b) Show by induction that for every $n \in \mathbb{N}$, $(1, n) \in E$.

(c) (Very hard!!!) Show that $E = \{(a, b) \in \mathbb{N} \times \mathbb{N} : \gcd(a, b) = 1\}$. *Think carefully about how the Euclidean algorithm works, and what the generating rules might have to do with it...*

6.1.7 A strict set-theoretic definition requires you to build the ordered pair (a, b) as a set: typically $(a, b) = \{a, \{a, b\}\}$. One then proves that $(a, b) = (c, d) \iff a = c$ and $b = d$.

(a) One of the axioms of set theory (*regularity*) says that there is no set a for which $a \in a$. Use this to prove that the cardinality of $(a, b) = \{a, \{a, b\}\}$ is two.

(b) Prove that $(a, b) = (c, d) \implies \begin{cases} a = c \text{ and } b = d, \\ \quad \textit{or} \\ a = \{c, d\} \text{ and } c = \{a, b\}. \end{cases}$

(c) In the second case, prove that there exists a set S such that $a \in S \in a$. The axiom of regularity also says that this is illegal. Conclude that $(a, b) = (c, d) \iff a = c$ and $b = d$.

6.2 Power Sets

Given a set A, it is often useful to consider the collection of *all* of the subsets of A. Indeed, we want to call this collection a set.

Definition 6.4. The *power set* of A is the set $\mathcal{P}(A)$ of all subsets of A. That is,

$$\mathcal{P}(A) = \{B : B \subseteq A\}.$$

Otherwise said: $B \in \mathcal{P}(A) \iff B \subseteq A$.

Examples. 1. Let $A = \{1,3,7\}$. Then A has the following subsets, listed by how many elements are in each subset.

0-elements: $\emptyset$
1-element: $\{1\}$, $\{3\}$, $\{7\}$
2-elements: $\{1,3\}$, $\{1,7\}$, $\{3,7\}$
3-elements: $\{1,3,7\}$

Gathering these together, we have the power set:

$$\mathcal{P}(A) = \Big\{\emptyset, \{1\}, \{3\}, \{7\}, \{1,3\}, \{1,7\}, \{3,7\}, \{1,3,7\}\Big\}.$$

2. Consider $B = \Big\{1, \big\{\{2\},3\big\}\Big\}$. It is essential that you use different size set brackets to prevent confusion. B has only *two* elements, namely 1 and $\big\{\{2\},3\big\}$. We can gather the subsets of B in a table.

0-elements: $\emptyset$
1-element: $\{1\}$, $\Big\{\big\{\{2\},3\big\}\Big\}$
2-elements: $\Big\{1, \big\{\{2\},3\big\}\Big\}$

In the second line, remember that to make a subset out of a single element you must surround the element with set brackets. Thus $1 \in B \implies \{1\} \subseteq B$ and

$$\big\{\{2\},3\big\} \in B \implies \Big\{\big\{\{2\},3\big\}\Big\} \subseteq B.$$

The power set of B is therefore

$$\mathcal{P}(B) = \Big\{\emptyset,\ \{1\},\ \Big\{\big\{\{2\},3\big\}\Big\},\ \Big\{1, \big\{\{2\},3\big\}\Big\}\Big\}.$$

Notation

Be absolutely certain that you understand the difference between $\in$ and $\subseteq$. It is easy to become confused when considering power sets. In the context of the previous example, here are eight propositions. Which are true and which are false?[20]

(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)	$\{1\} \subseteq \mathcal{P}(A)$

[20] Only (a), (d), and (g) are true. Make sure you understand why!

As a further exercise in being careful with notation, consider the following theorem.

Theorem 6.5. *If $A \subseteq B$, then $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.*

Proof. Suppose that $A \subseteq B$ and let $C \in \mathcal{P}(A)$. We must show that $C \in \mathcal{P}(B)$.
By definition, $C \in \mathcal{P}(A) \implies C \subseteq A$. Since subset inclusion is transitive (Theorem 4.4), we have

$$C \subseteq A \subseteq B \implies C \subseteq B.$$

This says that $C \in \mathcal{P}(B)$. Therefore $\mathcal{P}(A) \subseteq \mathcal{P}(B)$. ∎

It is very easy to get confused by this theorem. Exercises 6.2.4 and 6.2.5 discuss things further.

Cardinality and Power Sets

Let's investigate how the cardinality of a set and its power set are related. Consider a few basic examples where we list all of the subsets, grouped by cardinality.

Set A	0-elements	1-element	2-elements	3-elements	$\lvert\mathcal{P}(A)\rvert$
$\emptyset$	$\emptyset$				1
$\{a\}$	$\emptyset$	$\{a\}$			$1+1=2$
$\{a,b\}$	$\emptyset$	$\{a\},\{b\}$	$\{a,b\}$		$1+2+1=4$
$\{a,b,c\}$	$\emptyset$	$\{a\},\{b\},\{c\}$	$\{a,b\},\{a,c\},\{b,c\}$	$\{a,b,c\}$	$1+3+3+1=8$

You should have seen this pattern before: we are looking at the first few lines of Pascal's Triangle. It should be no surprise that if $|A| = 4$, then $|\mathcal{P}(A)| = 1+4+6+4+1 = 16$. The progression $1, 2, 4, 8, 16, \ldots$ in the final column immediately suggests the following theorem.

Theorem 6.6. *Suppose that A is a finite set. Then $|\mathcal{P}(A)| = 2^{|A|}$.*

How are we supposed to prove such a theorem for all sets at once? The trick is to think about all n-element sets simultanously, and prove by induction on the cardinality of A. The basic idea is that every set with $n+1$ elements is the disjoint union of a set with n elements and a single-element set. The induction step is essentially the observation that an $n+1$-element set B has *twice* the number of subsets of some n-element set A. It is instructive to see an example of this before writing the proof.

Example. Let $B = \{1, 2, 3\}$. Now choose the element $3 \in B$ and delete it to create the smaller set

$$A = \{1, 2\} = B \setminus \{3\}.$$

We can split the subsets of B into two groups: those which contain 3 and those which do not. In the following table we list all of the subsets of B. In the first column are those subsets X which do not

contain 3. These are exactly the subsets of A. In the second column are the subsets $Y = X \cup \{3\}$ of B which do contain 3.

X	$X \cup \{3\}$
$\emptyset$	$\{3\}$
$\{1\}$	$\{1,3\}$
$\{2\}$	$\{2,3\}$
$\{1,2\}$	$\{1,2,3\}$

It is clear that B has twice the number of subsets of A.

This method of pairing is exactly mirrored in the proof.

Proof. We prove by induction. For each $n \in \mathbb{N}_0$, let $Q(n)$ be the proposition

$$|A| = n \implies |\mathcal{P}(A)| = 2^n.$$

(*Base Case*) If $n = 0$, then $A = \emptyset$ (Theorem 4.4). But then $\mathcal{P}(A) = \{\emptyset\}$, whence $|\mathcal{P}(A)| = 1 = 2^0$. Therefore $Q(0)$ is true.
(*Induction Step*) Fix $n \in \mathbb{N}_0$ and assume that $Q(n)$ is true. That is, assume that any set with n elements has 2^n subsets. Now let B be *any* set with $n+1$ elements. Choose one of the elements $b \in B$ and define $A = B \setminus \{b\}$. Subsets of B are of 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 really a subset of A. Since $|A| = n$, the induction hypothesis $Q(n)$ tells us that there are 2^n subsets X of this type. In the second case, we can write $Y = X \cup \{b\}$, where X is again a subset of A. Since there are 2^n subsets X, it follows that there are 2^n subsets $Y \subseteq B$ of this form. Therefore

$$|\mathcal{P}(B)| = 2^n + 2^n = 2^{n+1}.$$

By induction, $Q(n)$ is true for all $n \in \mathbb{N}_0$. ∎

Once you understand the proof, you should compare it to the proof of Theorem 5.10 on the interior angles of a polygon. The idea is very similar. Exercise 6.2.8 gives an alternative proof of this result.

As a final example, we consider the interaction of power sets and Cartesian products. Suppose that $A = \{a\}$ and $B = \{b, c\}$. Then

$$A \times B = \{(a,b), (a,c)\}.$$

The power set $\mathcal{P}(A \times B)$ therefore contains $2^2 = 4$ elements: indeed

$$\mathcal{P}(A \times B) = \Big\{\emptyset,\ \{(a,b)\},\ \{(a,c)\},\ \{(a,b),(a,c)\}\Big\}.$$

The power sets of A and B have 2 and 4 elements respectively:

$$\mathcal{P}(A) = \{\emptyset, \{a\}\}, \qquad \mathcal{P}(B) = \{\emptyset, \{b\}, \{c\}, \{b,c\}\}.$$

The Cartesian product of the power sets therefore has $2 \times 4 = 8$ elements:

$$\mathcal{P}(A) \times \mathcal{P}(B) = \Big\{ (\emptyset,\emptyset),\ (\emptyset,\{b\}),\ (\emptyset,\{c\}),\ (\emptyset,\{b,c\}),\\ (\{a\},\emptyset),\ (\{a\},\{b\}),\ (\{a\},\{c\}),\ (\{a\},\{b,c\}) \Big\}.$$

It should be clear from this example not only that $\mathcal{P}(A \times B) \neq \mathcal{P}(A) \times \mathcal{P}(B)$, but that the elements of the two sets are completely different. The elements of $\mathcal{P}(A \times B)$ are *sets of ordered pairs,* while the elements of $\mathcal{P}(A) \times \mathcal{P}(B)$ are *ordered pairs of sets.*

Exercises

6.2.1 Find $\mathcal{P}(A)$ and $|\mathcal{P}(A)|$ for the following:

(a) $A = \{1,2\}$.

(b) $A = \{1,2,3\}$.

(c) $A = \big\{(1,2),(2,3)\big\}$.

(d) $A = \{\emptyset, 1, \{a\}\}$.

(e) $A = \Big\{\{1,2\}, 3, \{4,\{5\}\}\Big\}$.

(f) $A = \Big\{(1,2),\ 3,\ \big(4,\{5\}\big)\Big\}$.

6.2.2 Let $A = \{1,3\}$ and $B = \{2,4\}$.

(a) Draw a picture of the set $A \times B$.

(b) Compute $\mathcal{P}(A \times B)$.

(c) What is the cardinality of $\mathcal{P}(A) \times \mathcal{P}(B)$? *Don't compute the set!*

6.2.3 Determine whether the following statements are true or false (*in (b), the symbol $\subsetneq$ means 'proper subset'*). Justify your answers.

(a) If $\{7\} \in \mathcal{P}(A)$, then $7 \in A$ and $\{7\} \notin A$.

(b) Suppose that A, B and C are sets such that $A \subsetneq \mathcal{P}(B) \subsetneq C$ and $|A| = 2$. Then $|C|$ can be 5, but $|C|$ cannot be 4.

(c) If a set B has one more element than a set A, then $\mathcal{P}(B)$ has at least two more elements than $\mathcal{P}(A)$.

(d) Suppose that the sets A, B, C and D are all subsets of $\{1,2,3\}$ with cardinality two. Then at least two of these sets are equal.

6.2.4 Here are three incorrect proofs of Theorem 6.5. Explain why each fails.

(a) Let $x \in \mathcal{P}(A)$. Then $x \in A$. Since $A \subseteq B$, we have $x \in B$. Therefore $x \in \mathcal{P}(B)$, and so $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

(b) Let $A = \{1,2\}$ and $B = \{1,2,3\}$. Then $\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, A\}$, and $\mathcal{P}(B) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, B\}$. Thus $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

(c) Let $x \in A$. Since $A \subseteq B$, we have $x \in B$. Since $x \in A$ and $x \in B$, we have $\{x\} \in \mathcal{P}(A)$, and $\{x\} \in \mathcal{P}(B)$.

6.2.5 Consider the converse of Theorem 6.5. Is it true or false? Prove or disprove your conjecture.

6.2.6 (a) Prove that $\mathcal{P}(A) \cup \mathcal{P}(B) \subseteq \mathcal{P}(A \cup B)$. Provide a counter-example to show that we do not expect equality.

(b) Does anything change if you replace $\cup$ with $\cap$ in part (a)? Justify your answer.

6.2.7 Consider the proof of Theorem 6.6. Let B be a set with $n+1$ elements, let $b \in B$ and let $A = B \setminus \{b\}$. Prove that the function $f : \mathcal{P}(A) \times \{1,2\} \to \mathcal{P}(B)$ defined by

$$f(X,1) = X, \qquad f(X,2) = X \cup \{b\}$$

is a bijection, and that consequently, by Theorem 4.12, $|\mathcal{P}(A) \times \{1,2\}| = |\mathcal{P}(B)|$.

6.2.8 We use the following notation for the binomial coefficient:[21] $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. This symbol denotes the number of distinct ways one can choose r objects from a set of n objects.

(a) Prove directly, use the definition of the binomial coefficient, that

$$\text{If } 1 \leq r \leq n, \text{ then } \binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}.$$

(b) Prove by induction that $\forall n \in \mathbb{N}$, $\sum_{r=0}^{n} \binom{n}{r} = 2^n$. *You will need part (a) in the induction step.*

(c) Explain why part (b) provides an alternative proof of Theorem 6.6.

If you found this easy, try proving the binomial theorem: $\forall n \in \mathbb{N}$, $(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}$.

[21] You may have seen this written ${}^nC_r$, or ${}_nC_r$, where the C stands for *combination.*

6.3 Indexed Collections of Sets

An indexed family of sets is a collection of sets A_n, one for each n in some indexing set I. It is very often the case that $I = \mathbb{N}$ or $\mathbb{Z}$. If I is some other set, for example the real numbers $\mathbb{R}$, the label for the index may be chosen accordingly: e.g. $A_x \subseteq \mathbb{R}$.

Definition 6.7. Given a family of indexed sets $\mathcal{A} = \{A_n : n \in I\}$, we may form the *union* and *intersection* of the collection:

$$\cup\mathcal{A} = \bigcup_{n\in I} A_n = \{x : x \in A_n \text{ for some } n \in I\},$$
$$\cap\mathcal{A} = \bigcap_{n\in I} A_n = \{x : x \in A_n \text{ for all } n \in I\}.$$

Otherwise said,

$$x \in \bigcup_{n\in I} A_n \iff \exists n \in I \text{ such that } x \in A_n$$
$$x \in \bigcap_{n\in I} A_n \iff \forall n \in I \text{ we have } x \in A_n$$

A collection $\mathcal{A} = \{A_n : n \in I\}$ is *pairwise disjoint* if $A_m \cap A_n = \emptyset$ whenever $m \neq n$.

When the indexing set is $\mathbb{N}$ or $\mathbb{Z}$, it is also common to write, for example, $\bigcup_{n\in\mathbb{N}} A_n$ as $\bigcup_{n=1}^{\infty} A_n$.

The following Theorem is almost immediate given the definitions of union and intersection: can you supply a formal proof?

Theorem 6.8. *Let $\mathcal{A} = \{A_n : n \in I\}$ and let $m \in I$. Then*

$$A_m \subseteq \bigcup_{n\in I} A_n \quad \text{and} \quad \bigcap_{n\in I} A_n \subseteq A_m.$$

Examples. 1. For each $n \in \mathbb{N}$, let $A_n = [-n, n]$. Each of the sets A_n is a closed interval. E.g.,

$$A_1 = [-1, 1], \qquad A_2 = [-2, 2], \qquad A_3 = [-3, 3].$$

It should be clear that $n \leq m \implies A_n \subseteq A_m$. We therefore have a *nested* sequence of sets:

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$$

It follows immediately that $\bigcap_{n\in\mathbb{N}} A_n = A_1 = [-1, 1]$.
The union is a little harder. With a little thinking you might hypothesize $\bigcup_{n\in\mathbb{N}} A_n = \mathbb{R}$. This is indeed the case, but to prove it we need to return to the definition. Since every interval A_n is a subset of $\mathbb{R}$, we automatically have $\bigcup_{n\in\mathbb{N}} A_n \subseteq \mathbb{R}$. All that remains is to see that $\mathbb{R} \subseteq \bigcup_{n\in\mathbb{N}} A_n$.
Let $x \in \mathbb{R}$. We must show that $\exists n \in \mathbb{N}$ such that $x \in A_n$. We construct n explicitly using the

ceiling function.[22] If $x \geq 0$, then $x \leq \lceil x \rceil$, whence $x \in A_{\lceil x \rceil}$. Similarly, if $x < 0$, then $x \in A_{\lceil -x \rceil}$. For example,

$$-3.124 \in A_{\lceil 3.124 \rceil} = A_4.$$

It follows that all real numbers x are in at least one of the sets A_n, and so $\bigcup_{n\in\mathbb{N}} A_n = \mathbb{R}$.

2. Let $A_n = (n, n+1] \subseteq \mathbb{R}$, for each $n \in \mathbb{Z}$. For example,

$$A_3 = (3, 4], \quad \text{and} \quad A_{-17} = (-17, -16].$$

In this case the sets A_n are pairwise disjoint, and we have

$$\bigcup_{n\in\mathbb{Z}} A_n = \mathbb{R}, \quad \text{and} \quad \bigcap_{n\in\mathbb{Z}} A_n = \emptyset.$$

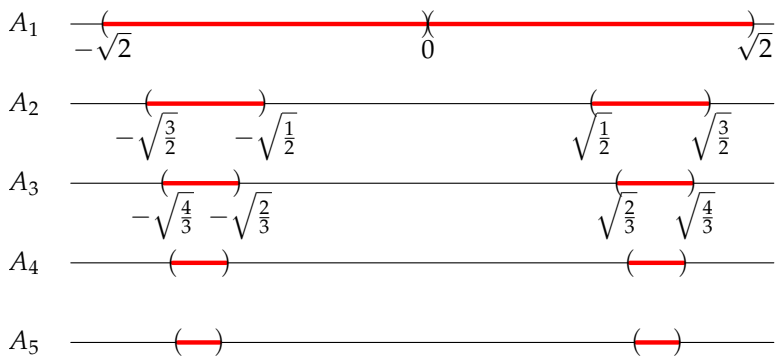
3. For each $n \in \mathbb{N}$, let $A_n = \{x \in \mathbb{R} : \left|x^2 - 1\right| < \frac{1}{n}\}$. Before computing the union and intersection of these sets, it is helpful to write each set as a pair of intervals. Note that

$$\left|x^2 - 1\right| < \frac{1}{n} \iff -\frac{1}{n} < x^2 - 1 < \frac{1}{n} \iff \sqrt{1 - \frac{1}{n}} < |x| < \sqrt{1 + \frac{1}{n}}.$$

Therefore

$$A_n = \left(-\sqrt{1+\tfrac{1}{n}}, -\sqrt{1-\tfrac{1}{n}}\right) \cup \left(\sqrt{1-\tfrac{1}{n}}, \sqrt{1+\tfrac{1}{n}}\right).$$

As the picture suggests, the sets A_n are nested: $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$



Since A_1 is the largest of the nested sets, we see that $\bigcup_{n\in\mathbb{N}} A_n = A_1 = (-\sqrt{2}, 0) \cup (0, \sqrt{2})$. For the intersection, note that

$$\forall n \in \mathbb{N},\ x \in A_n \iff \forall n \in \mathbb{N},\ \left|x^2 - 1\right| < \frac{1}{n} \iff x^2 - 1 = 0.$$

It follows that $\bigcap_{n\in\mathbb{N}} A_n = \{1, -1\}$.

[22] The *ceiling* $\lceil x \rceil$ is the smallest integer greater than or equal to x. For example $\lceil 3.124 \rceil = 4$. The ceiling function is simply the concept of 'rounding up' written in mathematical language. The corresponding function for 'rounding down' is the *floor:* $\lfloor x \rfloor$ is the greatest integer less than or equal to x.

Don't take Limits!

Here we dissect an extremely important example. For each $n \in \mathbb{N}$, define the interval $A_n = \left[0, \frac{1}{n}\right)$. Let us analyze the collection $\{A_n : n \in \mathbb{N}\}$. First observe that $m \leq n \implies \frac{1}{n} \leq \frac{1}{m} \implies A_n \subseteq A_m$, so that the sets are nested:

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$$

The union is therefore the largest interval A_1,

$$\bigcup_{n=1}^{\infty} A_n = A_1 = [0,1).$$

Before considering the full intersection, we first compute a finite intersection. Since the sets A_n are nested, it follows that any *finite* intersection is simply the smallest of the listed sets: i.e., for any constant $m \in \mathbb{N}$ we have

$$\bigcap_{n=1}^{m} A_n = A_m = \left[0, \frac{1}{m}\right).$$

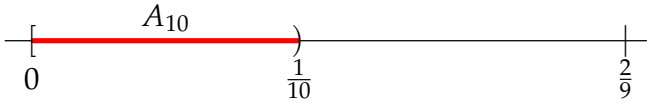
Observe that this is non-empty *for every* m. What about the infinite intersection? You might be tempted to take a limit and make an argument such as

$$\bigcap_{n=1}^{\infty} A_n = \lim_{m\to\infty} \bigcap_{n=1}^{m} A_n = \lim_{m\to\infty} \left[0, \frac{1}{m}\right) = [0,0).$$

Quite apart from the issue that $[0,0)$ is ugly and could only mean the empty set, we should worry about whether this is a legitimate use of limits. It isn't! Moreover, the attempt to use limits produces an incorrect conclusion: the intersection is in fact non-empty, and we claim the following.

Theorem 6.9. $\bigcap_{n=1}^{\infty} A_n = \{0\}$.

Before we give a formal proof, it is instructive to see a calculation. Let us show, for example, that $\frac{2}{9} \notin \bigcap_{n=1}^{\infty} A_n$. To prove that $\frac{2}{9}$ is not in the intersection of *all* the A_n, it is enough to exhibit a single integer m such that $\frac{2}{9} \notin A_m$. The picture shows that we can choose $m = 10$: since $\frac{1}{10} < \frac{2}{9}$, we have $\frac{2}{9} \notin [0, \frac{1}{10}] = A_{10}$. Since $\frac{2}{9} \notin A_{10}$, we conclude that $\frac{2}{9} \notin \bigcap_{n=1}^{\infty} A_n$.



Proof. We will prove that $x \in \bigcap_{n=1}^{\infty} A_n \implies x = 0$.

Suppose that $x \in \bigcap_{n=1}^{\infty} A_n$. Then $x \in \left[0\,,\,\frac{1}{n}\right)$ for all n. Otherwise said,

$$\forall n \in \mathbb{N}, \text{ we have } 0 \leq x < \frac{1}{n}.$$

Certainly $x = 0$ satisfies these inequalities. Suppose, for a contradiction, that $x > 0$. Since $\lim_{n\to\infty} \frac{1}{n} = 0$, we can certainly choose[a] N large enough so that $\frac{1}{N} \leq x$. A contradiction. Thus the intersection contains no positive elements, and we conclude that

$$\bigcap_{n=1}^{\infty} A_n = \{0\}.$$

■

[a] Explicitly, you may choose choose $N = \lceil \frac{1}{x} \rceil$, or anything larger.

The outcome of this discussion depends crucially on whether the ends of the intervals A_n are open or closed. Consider each of the following modifications in turn. How would the argument for computing each intersection differ from what we did above?

- If $B_n = \left(0\,,\,\frac{1}{n}\right)$, then $\bigcap_{n=1}^{\infty} B_n = \emptyset$.
- If $C_n = \left(0\,,\,\frac{1}{n}\right]$, then $\bigcap_{n=1}^{\infty} C_n = \emptyset$.
- If $D_n = \left[0\,,\,\frac{1}{n}\right]$, then $\bigcap_{n=1}^{\infty} D_n = \{0\}$.

The moral of these examples is that you cannot naïvely apply limits to sets. Be very careful with infinite unions and intersections, for your intuition can easily lead you astray.

Finite Decimals

Here is another example where 'taking the limit' is the incorrect thing to do. This time it is the union that forces us to be careful.

For each $n \in \mathbb{N}$, let $A_n = \{\text{decimals } 0.a_1a_2\ldots a_n \text{ of length } n\}$, where each $a_i \in \{0, 1, 2, \ldots, 9\}$. For example $0.134 \in A_3$. Since $0.134 = 0.1340$, we also have $0.134 \in A_4$. Once again we have nested intervals

$$m \leq n \implies A_m \subseteq A_n,$$

whence the infinite intersection is simply

$$\bigcap_{n\in\mathbb{N}} A_n = A_1 = \{0, 0.1, \ldots, 0.9\}.$$

Consider first a finite union: if $m \in \mathbb{N}$, then

$$\bigcup_{n=1}^{m} A_n = A_m = \{x \in [0,1) : x \text{ has a decimal representation of length} \leq m\}.$$

If one were to take the limit as $m \to \infty$ of the *property* 'length m decimal,' it seems like the infinite union should be the whole[23] interval $[0,1]$. This is another incorrect application of limits: one cannot take the limit of a property! Instead we use the definiton:

$$\begin{aligned} x \in \bigcup_{n\in\mathbb{N}} A_n &\iff \exists n \in \mathbb{N} \text{ such that } x \in A_n \\ &\iff \exists n \in \mathbb{N} \text{ such that } x \text{ is a decimal of length } n. \end{aligned}$$

It follows that

$$\bigcup_{n\in\mathbb{N}} A_n = \big\{x \in [0,1) : x \text{ has a } \textit{finite} \text{ decimal representation}\big\}$$

Not only does this mean that there are no irrational numbers in $\bigcup_{n\in\mathbb{N}} A_n$, but many rational numbers are also excluded. For example $\frac{1}{3} = 0.3333\cdots$ is not in any set A_n and is therefore not in the union.

Indexed Unions: Don't Confuse Sets and Elements

It is easy to confuse, but important to distinguish between the sets

$$\mathcal{A} = \{A_n : n \in I\} \qquad \text{and} \qquad \cup\mathcal{A} = \bigcup_{n\in I} A_n.$$

$\mathcal{A}$ is a set whose *elements* are themselves sets. The second is the collection of all elements in *any* set A_n. Consider the following examples.

Examples. 1. For each $n \in \{1,2,3\}$, let A_n be the plane $\{(x,y,z) : x + ny + n^2z = 1\} \subseteq \mathbb{R}^3$.
$\mathcal{A} = \{A_1, A_2, A_3\}$ has *three* elements: each of the planes A_1, A_2, A_3 is an object in its own right.
The union $\cup\mathcal{A} = A_1 \cup A_2 \cup A_3$ is an *infinite* set consisting of all the *points* on the three planes.
For the intersection, a little work with simultaneous equations should convince you that

$$(x,y,z) \in \bigcap_{n\in\{1,2,3\}} A_n \iff \begin{cases} x+y+z=1 \\ x+2y+4z=1 \\ x+3y+9z=1 \end{cases} \iff (x,y,z) = (1,0,0).$$

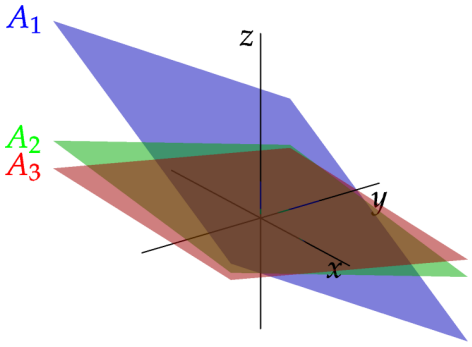
The planes are drawn below.

2. For each $m \in \mathbb{R} \cup \{\infty\}$, let A_m be the line[24] through the origin in $\mathbb{R}^2$ with gradient m.
Each element of $\mathcal{A}$ is a *line*: there is one for each possible direction through the origin.
$\cup\mathcal{A}$ is all of the *points* that lie on *any* line through the origin. Since every point can be joined to the origin with a straight line, the set $\cup\mathcal{A} = \mathbb{R}^2$ consists of all points in the plane.

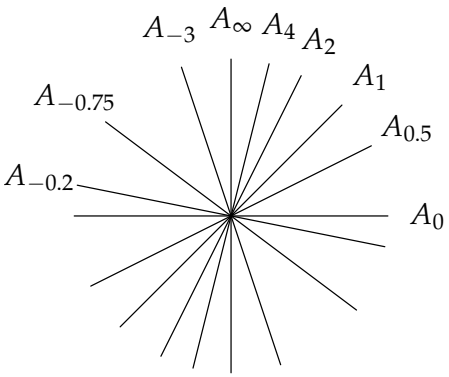
[23] We would include $1 = 0.9999\cdots$

[24] We include the vertical line A_∞.

It should be clear that all the lines intersect at the origin, and so $\cap\mathcal{A} = \{(0,0)\}$.
The collection of lines $\mathcal{A} = \{A_m : m \in \mathbb{R} \cup \{\infty\}\}$ is the famous *projective space* $\mathbb{P}(\mathbb{R}^2)$; this is a very different set from $\mathbb{R}^2$!
This example also shows that indexing sets don't have to be simple sets of integers. It is also possible to index the same set using $I = [0, \pi)$. If we define B_θ to be the line through the origin making an angle θ with the positive x-axis, we would then have $B_\theta = A_{\tan\theta}$.



Example 1: Three elements, or an infinite number?



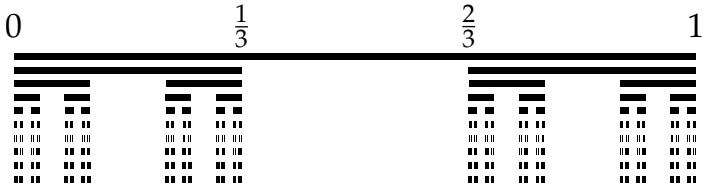
Example 2: Elements in $\mathbb{P}(\mathbb{R}^2)$

Aside: The Cantor Set

For a bit of fun, we can use infinite intersections to create self-similar sets, or *fractals.* Here is a famous example: the *Cantor middle-third set.*

Construct a sequence of sets C_n for $n \in \mathbb{N}_0$ by repeatedly removing the middle third of each of the intervals at each step, starting with $[0,1]$.

$C_0 = [0,1]$,
$C_1 = [0,\frac{1}{3}] \cup [\frac{2}{3},1]$,
$C_2 = [0,\frac{1}{9}] \cup [\frac{2}{9},\frac{1}{3}] \cup [\frac{2}{3},\frac{7}{9}] \cup [\frac{8}{9},1]$, etc.



The sequence is drawn up to C_9, with an animation below. To see the detail for the last few sets, try zooming in as far as you can.

Define the *Cantor set* $\mathcal{C}$ to be the infinte intersection $\mathcal{C} = \bigcap_{n=0}^{\infty} C_n$. This set has several interesting properties.

Zero Measure (length) Intuitively, the *length* of a set of real numbers is the sum of the lengths of all the intervals contained in the set. Since we start with the interval $[0,1]$ and remove a third of the set each time, it should be clear that

$$\text{length}(C_0) = 1, \qquad \text{length}(C_1) = \frac{2}{3}, \qquad \text{length}(C_2) = \left(\frac{2}{3}\right)^2, \quad \text{etc.}$$

Induction then gives us

$$\text{length}(C_n) = \left(\frac{2}{3}\right)^n.$$

As $n \to \infty$ this goes to zero, so the Cantor set contains no intervals: it is purely made up of individual points. This at least seems reasonable from the picture.

Non-emptiness The Cantor set $\mathcal{C}$ contains the endpoints of every interval removed at any stage of its construction. In particular, $\frac{1}{3^n} \in \mathcal{C}$ for all $n \in \mathbb{N}_0$, and so $\mathcal{C}$ is an *infinite* set. Indeed it is more than merely infinite, it is *uncountably* so, as we shall see in Chapter 8.

Self-similarity If $\frac{\mathcal{C}}{3}$ means 'take all the numbers in the set $\mathcal{C}$ and divide them by three,' and $\frac{\mathcal{C}}{3} + \frac{2}{3}$ means 'take all the numbers in $\frac{\mathcal{C}}{3}$ and add $\frac{2}{3}$ to them,' then

$$\mathcal{C} = \frac{\mathcal{C}}{3} \cup \left(\frac{\mathcal{C}}{3} + \frac{2}{3} \right). \tag{$*$}$$

Otherwise said, $\mathcal{C}$ is made up of two shrunken copies of itself, a classic property of fractals. If you were to zoom into the Cantor set far enough that you couldn't see the whole set, you would not know what the scale was. In the following animation we are repeatedly zooming in on the second (of four) groups of points.

To get further with the Cantor set, it is necessary to understand exactly what the elements of the set are. This can be accomplished using the *ternary representation.* It can be shown that every number $x \in [0,1]$ may be written in the form[25]

$$x = [0.a_1a_2a_3\cdots]_3 = \sum_{n=1}^{\infty} a_n \cdot 3^{-n} = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \cdots$$

where each $a_n \in \{0,1,2\}$. For example:

$$[0.12]_3 = \frac{1}{3} + \frac{2}{3^2} = \frac{5}{9}, \qquad \frac{64}{243} = \frac{2}{3^2} + \frac{1}{3^3} + \frac{1}{3^5} = [0.02101]_3, \qquad 1 = [0.22222\cdots]_3.$$

For this last, use the formula for the sum of a geometric series to calculate $\sum_{n=1}^{\infty} 2\left(\frac{1}{3}\right)^n = 2 \cdot \frac{1/3}{1-1/3} = 1$. The only possibility whereby x can have two ternary expansions is if one of them terminates. The other will eventually become a sequence of repeating 2's. For example:[26]

$$[0.0222222\cdots]_3 = [0.1]_3 = \frac{1}{3} \quad \text{and} \quad [0.10122222\cdots]_3 = [0.102]_3 = \frac{1}{3} + \frac{2}{27} = \frac{11}{27}.$$

Theorem 6.10. *C_n is the set of all numbers $x \in [0,1]$ with a ternary expansion whose first n digits are only 0 or 2. It follows that $\mathcal{C}$ is exactly the set of $x \in [0,1]$ with a ternary expansion containing only 0 and 2.*

[25] Analogous to a decimal representation $x = \sum_{n=1}^{\infty} a_n \cdot 10^{-n} = \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \cdots$ where $a_n \in \{0,1,2,\ldots,9\}$.

[26] This is ticklish to prove, as is the corresponding result for decimals: consider $1 = 0.99999999\cdots$

Proof. We prove by induction.
(*Base Case*) The proposition is clearly true for $C_0 = [0,1]$, as there is nothing to check.
(*Induction Step*) Assume that the proposition is true for some fixed $n \in \mathbb{N}_0$. Analogously to $(*)$ above, observe that C_{n+1} is built from two shrunken copies of C_n:

$$C_{n+1} = \frac{1}{3}C_n \cup \left(\frac{1}{3}C_n + \frac{2}{3}\right).$$

Multiplication by $\frac{1}{3}$ shifts a ternary representation one position to the right.[a]
Addition of $\frac{2}{3}$ adds $[0.2]_3$ to the representation, inserting 2 in the (now empty) first ternary place.
Thus if C_n contains only 0's and 2's in its first n entries, C_{n+1} contains only 0's and 2's in its first $n+1$ entries.
By induction the proposition is true for all $n \in \mathbb{N}$. ∎

[a] Compare to multiplication of a decimal by $\frac{1}{10}$.

As the Theorem shows, the Cantor set contains a lot of elements. For example:

$$[0.020202020\cdots]_3 = 2\sum_{n=1}^{\infty} 3^{-2n} = \frac{2/9}{1-1/9} = \frac{1}{4} \in \mathcal{C}.$$

What is strange is that $\frac{1}{4}$ is not the endpoint of any of the open intervals deleted during the construction of $\mathcal{C}$, and yet we've already established that $\mathcal{C}$ contains no intervals! Cantor introduced his set precisely because it was so challenging to the traditional concept of size: $\mathcal{C}$ seems to simultaneously have very few elements and enormously many.

Generalizations and related concepts include Cantor dust $\mathcal{C} \times \mathcal{C}$, the Sierpiński carpet and gasket, and the von Koch snowflake.

Exercises

6.3.1 For each integer n, consider the set $B_n = \{n\} \times \mathbb{R}$.

(a) Draw a picture of $\bigcup_{n=2}^{4} B_n$ (in the Cartesian plane).

Hint: $\bigcup_{n=2}^{4} B_n = B_2 \cup B_3 \cup B_4$.

(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 $\bigcup_{n=2}^{4} (B_n \cap C)$.

(e) Compare $\left(\bigcup_{n=2}^{4} B_n\right) \cap C$ and $\bigcup_{n=2}^{4} (B_n \cap C)$. What do you notice?

6.3.2 For each real number r, define the interval $S_r = [r-1, r+3]$. Let $I = \{1, 3, 4\}$. Determine $\bigcup_{r \in I} S_r$ and $\bigcap_{r \in I} S_r$.

6.3.3 Give an example of four different subsets A, B, C and D of $\{1, 2, 3, 4\}$ such that all intersections of two subsets are different.

6.3.4 For each of the following collections of intervals, define an interval A_n for each $n \in \mathbb{N}$ such that indexed collection $\{A_n\}_{n \in \mathbb{N}}$ is the given collection of sets. Then find both the union and intersection of the indexed collections of sets.

(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) $\{(\frac{1}{4}, 1), (\frac{1}{8}, \frac{1}{2}), (\frac{1}{16}, \frac{1}{4}), (\frac{1}{32}, \frac{1}{8}), (\frac{1}{64}, \frac{1}{16}), \ldots\}$

6.3.5 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 $\bigcap_{x \in \mathbb{R}} A_x$.

6.3.6 In Example 2 on page 107, give a formal proof using the ceiling function that $\bigcup_{n \in \mathbb{Z}} A_n = \mathbb{R}$.

6.3.7 Use Definition 6.7 to prove the following results about nested sets.

(a) $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \implies \bigcup_{n \in \mathbb{N}} A_n = A_1$.

(b) $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \implies \bigcap_{n \in \mathbb{N}} A_n = A_1$.

6.3.8 Let $C_0(\mathbb{R})$ denote the set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ which satisfy $f(0) = 0$. Let $A_f = \{x \in [0, 1] : f(x) = 0\}$ (so, for example, if $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x(2x-1)$, then $A_f = \{0, \frac{1}{2}\}$). *Prove* that

$$\bigcup_{f \in C_0(\mathbb{R})} A_f = [0, 1] \qquad \text{and} \qquad \bigcap_{f \in C_0(\mathbb{R})} A_f = \{0\}.$$

6.3.9 Let A_n be the set of decimals of length n, as described on page 109.

(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) Prove by contradiction that $\frac{1}{3} \notin \bigcup_{n=1}^{\infty} A_n$.

6.3.10 Suppose that the following are true:

- $\forall n \in \mathbb{N}$, $A_n \neq \emptyset$.
- $m \geq n \implies A_m \subseteq A_n$.

Prove or disprove the following conjectures:

(a) $\bigcup_{n=1}^{293} A_n \neq \emptyset$

(b) $\bigcap_{n=1}^{293} A_n \neq \emptyset$

(c) $\bigcup_{n\in\mathbb{N}} A_n \neq \emptyset$

(d) $\bigcap_{n\in\mathbb{N}} A_n \neq \emptyset$

6.3.11 (Hard) Let $A_n = \{\frac{m}{n} \in \mathbb{Q} : 0 < m < n, m \in \mathbb{N}\}$, for each $n \in \mathbb{N}$.

(a) Write down A_1, A_2, A_3, A_4 explicitly.

(b) Prove that $A_m \subseteq A_{pm}$ for any $p \in \mathbb{N}$.

(c) Argue that $\bigcup_{n\in\mathbb{N}} A_n = \mathbb{Q} \cap (0,1)$.

(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 $p \in \mathbb{N}$, $\bigcup_{n\in\mathbb{N}} A_{pn} = \mathbb{Q} \cap (0,1)$.

7 Relations and Partitions

The mathematics of sets is rather basic, at least until one has a notion of how to relate elements of sets with each other. We are already familiar with examples of this:

1. The usual *order* of the natural numbers (e.g. $3 < 7$) is a way of relating/comparing two elements of $\mathbb{N}$. Recall that, as sets, order doesn't matter: $\{3,7\} = \{7,3\}$. As *ordered pairs* however, $(3,7) \neq (7,3)$.

2. A *function* $f : A \to B$ relates elements in the set A with those in B.

It turns out that the concept of ordered pair is essential to relating elements.

7.1 Relations

Definition 7.1. Let A and B be sets. A *(binary) relation* R from A to B is a set of ordered pairs

$$R \subseteq A \times B.$$

A *relation on* A is a relation from A to itself.
If $(x,y) \in R$ we can also write $x\,R\,y$, and say 'x is related to y.' Similarly $x \not R y$ means $(x,y) \notin R$.

Examples. 1. $R = \{(1,3),(2,2),(2,3),(3,2),(4,1),(5,2)\}$ is a relation from $\mathbb{N}$ to $\mathbb{N}$. It is also a relation from $\{1,2,3,4,5\}$ to $\{1,2,3\}$.

2. $R = \big([1,3) \times (3,4]\big) \cup \big\{(2t+1, t^2) : t \in [\tfrac{1}{2}, 2]\big\}$ is a relation from $\mathbb{R}$ to $\mathbb{R}$. Be careful: it is easy to confuse interval notation with the notation for ordered pair!

3. The diagonal $R = \{(a,a) : a \in A\}$ is a relation on A, indeed

   $$(x,y) \in R \iff x = y$$

   defines a relation on *any* set A. This example is where the term *equivalence relation* comes from. $x\,R\,y \iff x = y$ simply says that R is 'equals.'

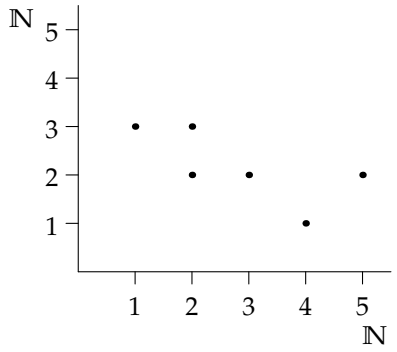
4. If $A = \{\text{all humans}\}$, we may define $R \subseteq A \times A$ by

   $$(a_1, a_2) \in R \iff a_1, a_2 \text{ have a parent-child, or a sibling relationship.}$$

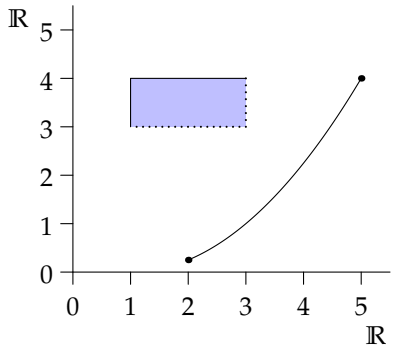
   In this example, the mathematical use of the word relation is identical to that in English. For example, I am related to my sister, and my mother is related to me.

5. If A is a set, then $\subseteq$ is a relation on the power set $\mathcal{P}(A)$.

When R is a relation between sets of numbers, we can often graph the relation. Examples 1 and 2 above would be graphed as follows:



Example 1.



Example 2.

Not all relations between sets of numbers can be graphed: for example, graphing the relation $\mathbb{Q} \times \mathbb{Q}$ is impossible!

Definition 7.2. If $R \subseteq A \times B$ is a relation, then its *inverse* $R^{-1} \subseteq B \times A$ is the set

$$R^{-1} = \{(y, x) \in B \times A : (x, y) \in R\}.$$

To find the elements of R^{-1}, you simply switch the components of each ordered pair in R. Suppose $A = B$. We say that R is *symmetric* if $R = R^{-1}$.

The following results should seem natural, even if some of the proofs may not be obvious.

Theorem 7.3. *Given any relations $R, S \subseteq A \times A$:*

1. $(R^{-1})^{-1} = R$
2. $R \subseteq S \iff R^{-1} \subseteq S^{-1}$
3. $(R \cup S)^{-1} = R^{-1} \cup S^{-1}$
4. $(R \cap S)^{-1} = R^{-1} \cap S^{-1}$
5. $R \cup R^{-1}$ *is symmetric*
6. $R \cap R^{-1}$ *is symmetric*

Proof. Here are two of the arguments. Try the others yourself.

2. Assume that $R \subseteq S$, and suppose that $(x, y) \in R^{-1}$. We must prove that $(x, y) \in S^{-1}$. By the definition of inverse,

$$\begin{aligned}(x, y) \in R^{-1} &\implies (y, x) \in R \implies (y, x) \in S \\ &\implies (x, y) \in S^{-1}.\end{aligned}$$

Therefore $R^{-1} \subseteq S^{-1}$. For the converse, suppose that $R^{-1} \subseteq S^{-1}$. Then, by an argument similar

to the above, we see that $(R^{-1})^{-1} \subseteq (S^{-1})^{-1}$. Now use 1. to see that

$$R^{-1} \subseteq S^{-1} \implies R \subseteq S.$$

5. By 3, $(R \cup R^{-1})^{-1} = R^{-1} \cup (R^{-1})^{-1} = R^{-1} \cup R = R \cup R^{-1}$, and so $R \cup R^{-1}$ is symmetric. ∎

Be careful! Several parts of Theorem 7.3 look suspiciously similar to earlier results and it is easy to get confused. For example, 3. and 4. look almost like De Morgan's laws, except that $\cup$ and $\cap$ do not switch over. This is why it is important to be able to prove and come up with examples of such statements. Suppose that you forget which result is correct: you might expect that

$$(R \cup S)^{-1} = \begin{cases} R^{-1} \cup S^{-1} \\ \quad \text{or} \\ R^{-1} \cap S^{-1}. \end{cases}$$

Now that you have two sensible guesses, you should be able to decide the correct one by thinking about examples and, if necessary, proving it!

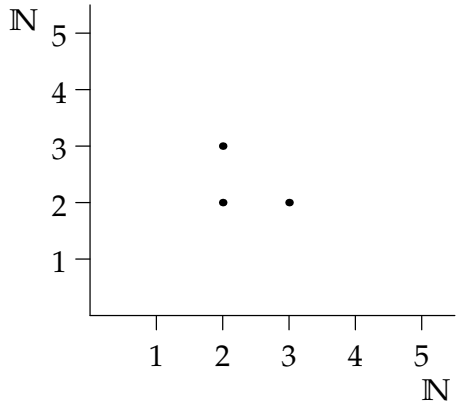
Example. Consider the example $R = \{(1,3),(2,2),(2,3),(3,2),(4,1),(5,2)\}$ from earlier. This is clearly not symmetric since $(1,3) \in R$ but $(3,1) \notin R$. We compute

$$R^{-1} = \{(3,1),(2,2),(3,2),(2,3),(1,4),(2,5)\},$$

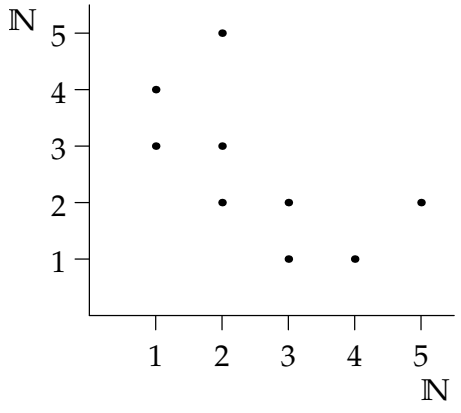
and observe that

$$R \cap R^{-1} = \{(2,2),(2,3),(3,2)\} \quad \text{and}$$
$$R \cup R^{-1} = \{(1,3),(3,1),(2,2),(2,3),(3,2),(4,1),(1,4),(5,2),(2,5)\}$$

are both symmetric.



The relation $R \cap R^{-1}$



The relation $R \cup R^{-1}$

The above pictures should confirm something intuitive: if you are able to graph a symmetric relation, then the graph will have symmetry about the line $y = x$.

Exercises

7.1.1 Draw pictures of the following relations on $\mathbb{R}$.

(a) $R = \{(x, y) : y \leq x \text{ and } y \leq 2 \text{ and } y \leq 2 - x\}$.

(b) $S = \{(x, y) : (x - 4)^2 + (y - 1)^2 \leq 9\}$.

Also draw the inverse of each relation.

7.1.2 A relation is defined on $\mathbb{N}$ by $a \, R \, b \iff \frac{a}{b} \in \mathbb{N}$. Let $c, d \in \mathbb{N}$. Under what conditions is it permissable to write $c \, R^{-1} \, d$?

7.1.3 Let $R \subseteq \{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$ be the relation

$$R = \{(1, 3), (1, 4), (2, 2), (2, 4), (3, 1), (3, 2), (4, 4)\}.$$

(a) Compute R^{-1}.

(b) Compute the relations $R \cup R^{-1}$ and $R \cap R^{-1}$, and check that they are symmetric.

7.1.4 For the relation $R = \{(x, y) : x \leq y\}$ defined on $\mathbb{N}$, what is R^{-1}?

7.1.5 Let A be a set with $|A| = 4$. What is the maximum number of elements that a relation R on A can contain such that $R \cap R^{-1} = \emptyset$?

7.1.6 Give formal proofs of the remaining cases (1, 3, 4 & 6) of Theorem 7.3.

7.2 Functions revisited

Now that we have the language of relations, we can properly define functions. Recall that a function $f : A \to B$ is a rule that assigns one, and only one, element of B to each element of A. We may therefore view f as a collection of ordered pairs in $A \times B$:

$$\{(a, f(a)) : a \in A\}.$$

This set is nothing more than the *graph* of the function, and, being a set of ordered pairs, it is a relation.

Definition 7.4. Let $R \subseteq A \times B$ be a relation from A to B. The *domain* and *range* of R are the sets

$$\text{dom}(R) = \{a \in A : (a, b) \in R \text{ for some } b \in B\},$$
$$\text{range}(R) = \{b \in B : (a, b) \in R \text{ for some } a \in A\}.$$

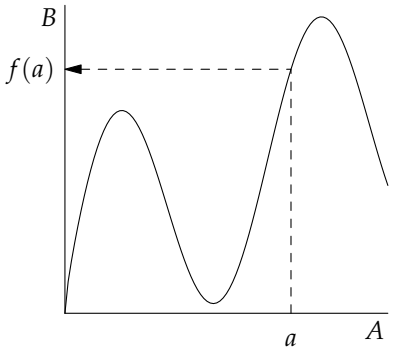
A *function* from A to B is a relation $f \subseteq A \times B$ satisfying the following conditions:

1. $\text{dom}(f) = A$,
2. $(a, b_1), (a, b_2) \in f \implies b_1 = b_2$.

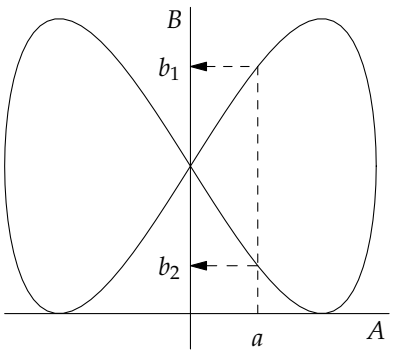
The two conditions can be thought of as saying:

1. Every element of A is related to *at least one* element of B.
2. Every element of A is related to *at most one* element of B.

Putting these together, we see that a relation $R \subseteq A \times B$ is a function if *every* $a \in A$ is the first entry of one (and only one) ordered pair $(a, b) \in R$. The second condition is the vertical line test, familiar from calculus.



$b_1 = b_2 = f(a)$: a function



$b_1 \neq b_2$: not a function

We can also think about injectivity and surjectivity (recall Definition 4.11) in this context. A function $f \subseteq A \times B$ is:

- *Injective* if no two pairs in f share the same second entry.
- *Surjective* if every $b \in B$ appears as the second entry of at least one pair in f.
- *Bijective* if every $b \in B$ appears as the second entry of one (and only one) ordered pair $(a, b) \in f$.

Definition 7.5. The *inverse* of a function $f \subseteq A \times B$ is the inverse relation $f^{-1} \subseteq B \times A$.

Since to compute the inverse relation we simply switch the components of each ordered pair, it should be clear that

$$\text{dom}(f^{-1}) = \text{range}(f) \quad \text{and} \quad \text{range}(f^{-1}) = \text{dom}(f).$$

In general, you should expect the inverse of a function to be merely a relation and not a function in its own right. Theorem 7.6 will discuss when the inverse relation is a function. The inverse of a function is usually written in set notation. If $V \subseteq B$, then we defined the *inverse image of* V (or *pull-back* of V) by

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

In particular, if $b \in B$, then

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

Both are *subsets* of A. When f^{-1} is a function, each set $f^{-1}(\{b\})$ consists of a single point of A (one for each $b \in B$). *Only* in this case are we entitled to write $f^{-1}(b) = a$.

Examples. 1. Let $A = B = \{1, 2, 3\}$ and $f = \{(1, 3), (2, 1), (3, 3)\}$.

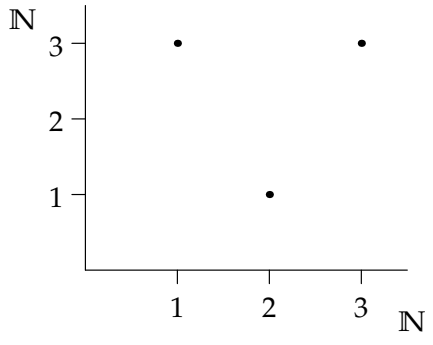
Note that $\text{dom}(f) = \{1, 2, 3\} = A$, and that each element of A appears exactly once as the first element in a pair $(a, b) \in f$. This relation therefore satisfies both conditions necessary to be a function. In more elementary language we would write $f(1) = 3$, $f(2) = 1$, and $f(3) = 3$.

f is not injective, since 3 appears twice as a second entry of an ordered pair in f.
f is not surjective, since 2 never appears as the second entry of an ordered pair in f.

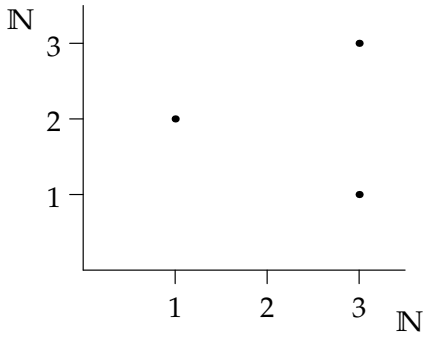
The inverse relation $f^{-1} = \{(3, 1), (1, 2), (3, 3)\} \subseteq B \times A$ is not a function by dint of failing *both* conditions in Definition 7.4.

- $\text{dom}(f^{-1}) = \{1, 3\}$ is not the whole of B.
- $(3, 1) \in f^{-1}$ and $(3, 3) \in f^{-1}$, but $1 \neq 3$.

The graphs of f and f^{-1} are shown below.



$f : A \to B$



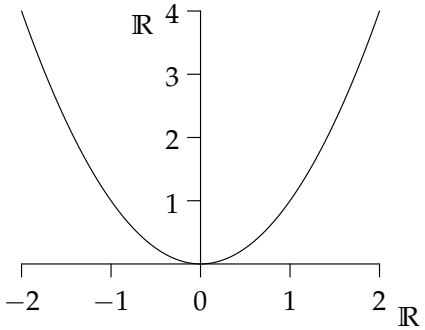
$f^{-1} \subseteq B \times A$: not a function

2. Let $A = B = \mathbb{R}$ and $f = \{(x, x^2) : x \in \mathbb{R}\}$. This is just the function $f(x) = x^2$.

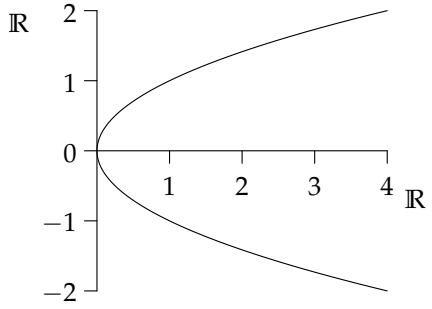
   The inverse is not a function:

   $$f^{-1} = \{(x^2, x) : x \in \mathbb{R}\} = \{(y, \pm\sqrt{y}) : y \geq 0\},$$

   since, for example, $f^{-1}(\{4\}) = \{-2, 2\}$ is not a single-element set.



$f : A \to B$

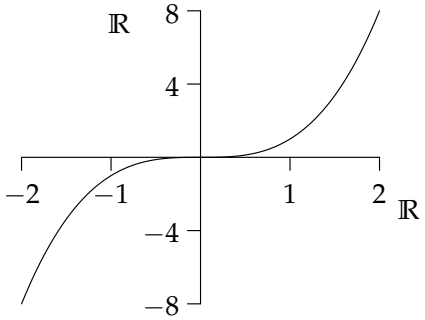


$f^{-1} \subseteq B \times A$: not a function

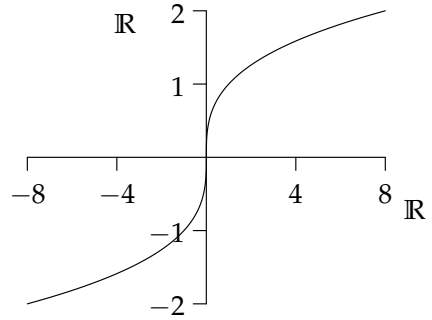
3. Let $A = B = \mathbb{R}$ and $f = \{(x, x^3) : x \in \mathbb{R}\}$. This is the function $f(x) = x^3$.

   This time, the inverse is also a function, $f^{-1}(y) = \sqrt[3]{y}$:

   $$f^{-1} = \{(x^3, x) : x \in \mathbb{R}\} = \{(y, \sqrt[3]{y}) : y \in \mathbb{R}\}.$$



$f : A \to B$



$f^{-1} : B \to A$ is a function

4. Let $A = \mathbb{R}$, $B = \mathbb{Q}$ and $f = \{(x, x) : x \in \mathbb{Q}\} \cup \{(x, 0) : x \notin \mathbb{Q}\}$. Then f is a function

   $$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

   This is a surjective function since every element of $B = \mathbb{Q}$ appears as the second entry in an ordered pair $(a, b) \in f$.
   It is not injective since zero appears more than once as the second entry of an ordered pair. For example,

   $$(\sqrt{2}, 0),\ (\sqrt{3}, 0) \in f.$$

   Intuitively this is simply $f(\sqrt{3}) = f(\sqrt{2})$.
   The inverse relation f^{-1} is not a function; for example $f^{-1}(\{0\})$ is the *set* $\{0\} \cup (\mathbb{R} \setminus \mathbb{Q})$, not a single value.

These examples help to illustrate the following important theorem.

Theorem 7.6. *A relation $f^{-1} \subseteq B \times A$ is a function $\iff$ f is bijective (both injective and surjective).*

Proof. Recalling Definition 7.4, we see that

$$f^{-1} \text{ is a function } \iff \begin{cases} \operatorname{dom}(f^{-1}) = B, \\ \quad \textit{and} \\ (b, a_1), (b, a_2) \in f^{-1} \implies a_1 = a_2. \end{cases}$$

The first of these is equivalent to $\operatorname{range}(f) = B$, and says that f is surjective.
The second is equivalent to $(a_1, b), (a_2, b) \in f \implies a_1 = a_2$, which says that f is injective. ∎

Equality of functions

There are two competing notions of equality of functions, dependent on what definition you take as fundamental.

Same domain, same graph, same codomain $f = g$ means that f and g are the same subset of the *same* $A \times B$. This notion is preferred by set theorists because it sticks rigidly to the idea that a function is a *relation,* and it requires both the domain A and codomain B to be explicit.

Same domain, same graph $f = g$ means that $f \subseteq A \times B$, $g \subseteq A \times C$, and $(a, b) \in f \iff (a, b) \in g$. This notion considers fundamental the notion of what a function *does,* rather than its strict status as a relation; if two functions do the same thing to elements of the same domain then they are the same. This looser notion of equality is used more often.

Unfortunately the second notion, while intuitive, has a problem. For example, let

$$f : \mathbb{R} \to \mathbb{R}, \quad \text{and} \quad g : \mathbb{R} \to [-1, 1] \quad \text{satisfy} \quad f(x) = g(x) = \sin x.$$

Although f and g have the same graph, the different codomains of f and g mean that these are *different functions* with respect to the first notion. Under the second notion, they are the *same.* However, g is surjective while f is not, so don't we want f and g to be different?!

The same problem does not arise when considering domains. For example, in calculus you might have compared functions such as

$$f(x) = x^2 + 2, \quad \text{and} \quad g(x) = \frac{(x^2 + 2)(x - 1)}{x - 1}.$$

The implied domains of these functions are $\operatorname{dom}(f) = \mathbb{R}$ and $\operatorname{dom}(g) = \mathbb{R} \setminus \{1\}$. Even though these functions have the same graph whenever *both* are defined, regardless of which notion you choose we have $f \neq g$, since the functions have *different domains.*

Exercises

7.2.1 Suppose that $f \subseteq \{1,2,3,4\} \times \{1,2,3,4,5,6,7\}$ is the relation

$$f = \{(1,1),(2,3),(3,5),(4,7)\}.$$

(a) Show that f is a function $f : \{1,2,3,4\} \to \{1,2,3,4,5,6,7\}$. Can you find a concise formula $f(x)$ to describe f?

(b) Is f injective? Justify your answer.

(c) Suppose that $g \subseteq \{1,2,3,4\} \times B$ is another relation so that the *graphs* of f and g are identical: i.e.

$$\big\{(a, f(a)) : a \in \{1,2,3,4\}\big\} = \big\{(a, g(a)) : a \in \{1,2,3,4\}\big\}.$$

as sets. If g is a bijective function, what is B?

7.2.2 Decide whether each of the following relations are functions. For those which are, decide whether the function is injective and/or surjective.

(a) $R = \{(x,y) \in [-1,1] \times [-1,1] : x^2 + y^2 = 1\}$

(b) $S = \{(x,y) \in [-1,1] \times [0,1] : x^2 + y^2 = 1\}$

(c) $T = \{(x,y) \in [0,1] \times [-1,1] : x^2 + y^2 = 1\}$

(d) $U = \{(x,y) \in [0,1] \times [0,1] : x^2 + y^2 = 1\}$

7.2.3 In Example 2 on page 122, explain why the functon f is neither injective nor surjective in the same manner as we did for Example 1.

7.2.4 (a) Express the function $f : \mathbb{R} \to \mathbb{R} : x \mapsto x^4 + 3$ as a relation.

(b) What is the inverse relation f^{-1}?

(c) Use Definition 7.4 to prove that the relation f^{-1} is *not* a function.

(d) Prove directly from Definition 4.11 that f is not injective and not surjective. Compare your arguments with your answer to part (c).

7.3 Equivalence Relations

In mathematics, the notion of *equality* is not as simple as one might think. The idea of two numbers being equal is straightforward, but suppose we want to consider two paths between given points as 'equal' if and only if they have the same length? Since two 'equal' paths might look very different, is this a good notion of equality? Mathematicians often want to gather together objects that have a common property and then treat them as if they were a single object. This is done using equivalence relations and equivalence classes.

First recall the alternative notation for a relation on a set A: if $R \subseteq A \times A$ is a relation on A, then $x\,R\,y$ has the same meaning as $(x, y) \in R$. We might read $x\,R\,y$ as 'x is R-related to y.'

Definition 7.7. A relation R on a set A may be described as *reflexive, symmetric* or *transitive* if it satisfies the following properties:

Reflexivity	$\forall x \in A,\ x\,R\,x$	(every element of A is related to itself)
Symmetry	$\forall x, y \in A,\ x\,R\,y \implies y\,R\,x$	(if x is related to y, then y is related to x)
Transitivity	$\forall x, y, z \in A,\ x\,R\,y$ and $y\,R\,z \implies x\,R\,z$	(if x is related to y, and y is related to z, then x is related to z)

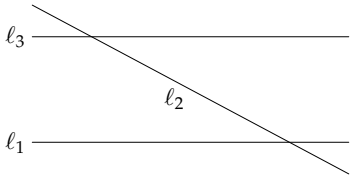
Symmetry is exactly the same notion as in Definition 7.2.

Examples. 1. Let $A = \mathbb{R}$ and let R be $\leq$. Thus $2 \leq 3$, but $7 \not\leq 4$. We check whether R satisfies the above properties.

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 $x \leq z$.

2. Let A be the set of lines in the plane and define $\ell_1\,R\,\ell_2 \iff \ell_1$ and ℓ_2 intersect.

Reflexivity	True. Every line intersects itself, so $\ell\,R\,\ell$ for all $\ell \in A$.
Symmetry	True. For all lines $\ell_1, \ell_2 \in A$, if ℓ_1 intersects ℓ_2, then ℓ_2 intersects ℓ_1..
Transitivity	False. As the picture illustrates, we may let ℓ_1 and ℓ_3 be parallel lines, and ℓ_2 cross both of these. Then $\ell_1\,R\,\ell_2$ and $\ell_2\,R\,\ell_3$, but $\ell_1 \not R\, \ell_3$.



Definition 7.8. An *equivalence relation* is a relation $\sim$ which is reflexive, symmetric and transitive.

The symbol $\sim$ is almost universally used for an abstract equivalence relation. It can be read as 'related to,' 'tilde,' or 'twiddles.' The two examples above are *not* equivalence relations because they fail one of the three conditions. Here is the simplest equivalence relation.

Example. Equals '=' is an equivalence relation on any set, hence the name!

Read the definitions of reflexive, symmetric and transitive until you are certain of this fact. There are countless other equivalence relations: here are a few.

Examples. 1. For all $x, y \in \mathbb{Z}$, let $x \sim y \iff x - y$ is even. We claim that $\sim$ is an equivalence relation on $\mathbb{Z}$.

Reflexivity $\forall x \in \mathbb{Z},\ x - x = 0$ is even, hence $x \sim x$.

Symmetry $\forall x, y \in \mathbb{Z},\ x \sim y \implies x - y$ is even $\implies y - x$ is even $\implies y \sim x$.

Transitivity $\forall x, y, z \in \mathbb{Z}$, if $x \sim y$ and $y \sim z$, then $x - y$ and $y - z$ are even. But the sum of two even numbers is even, hence $x - z = (x - y) + (y - z)$ is even, and so $x \sim z$.

2. Let $A = \{\text{all students taking this course}\}$. For all $x, y \in A$, let $x \sim y \iff x$ achieves the same letter-grade as y. Then $\sim$ is an equivalence relation on A.

Reflexivity $\forall x \in A,\ x \sim x$ since everyone scores the same as themself!

Symmetry $\forall x, y \in A,\ x \sim y \implies x$ achieves the same letter-grade as y
$\implies y$ achieves the same letter-grade as x
$\implies y \sim x$

Transitivity $\forall x, y, z \in A$, if $x \sim y$ and $y \sim z$, then x achieves the same as y who achieves the same as z, whence x achieves the same as z. Thus $x \sim z$.

3. For all $x, y \in \mathbb{Z}$, let $x \sim y \iff x^2 \equiv y^2 \pmod 5$. Then $\sim$ is an equivalence relation on $\mathbb{Z}$.

Reflexivity $\forall x \in \mathbb{Z},\ x \sim x$ since x^2 is always congruent to itself!

Symmetry $\forall x, y \in \mathbb{Z},\ x \sim y \implies x^2 \equiv y^2 \pmod 5$
$\implies y^2 \equiv x^2 \pmod 5$
$\implies y \sim x$

Transitivity $\forall x, y, z \in \mathbb{Z}$, if $x \sim y$ and $y \sim z$, then $x^2 \equiv y^2$ and $y^2 \equiv z^2 \pmod 5$. But then $x^2 \equiv z^2 \pmod 5$ and so $x \sim z$.

The most important thing to observe with each of these examples is that **an equivalence relation separates elements of a set into subsets where elements share a common property** (even/oddness, letter-grade, etc.). The next definition formalizes this idea.

Definition 7.9. Let $\sim$ be an equivalence relation on X. The *equivalence class* of x is the set

$$[x] = \{y \in X : y \sim x\}.$$

$X/_\sim$ is the set of all equivalence classes: the *quotient* of X by $\sim$, or 'X mod $\sim$.'

Let us think about the definition in the context of our examples.

Examples. 1. $[0] = \{y \in \mathbb{Z} : y \sim 0\} = \{y \in \mathbb{Z} : y \text{ is even}\}$ is the set of even numbers. Note that $[0]$ is also equal to $[2]$, $[4]$, $[6]$, etc.
The other equivalence class is $[1] = \{y \in \mathbb{Z} : y - 1 \text{ is even}\}$, which is the set of odd numbers.
The quotient set is $\mathbb{Z}/_\sim = \{[0], [1]\} = \{\{\text{even numbers}\}, \{\text{odd numbers}\}\}$.

2. There is one equivalence class for each letter grade awarded. Each equivalence class contains all the students who obtain a particular letter-grade. If we call the equivalence classes $A^+, A, A^-, B^+, \ldots, F$, where, say, $B = \{\text{students obtaining a B-grade}\}$, then

$$\{\text{Students}\}/_\sim = \{A^+, A, A^-, B^+, \ldots, F\}.$$

3. The equivalence classes for this example are a little tricky. First observe that

$$x \equiv y \pmod 5 \implies x^2 \equiv y^2 \pmod 5,$$

so that there are at most five equivalence classes; those of 0, 1, 2, 3 and 4. Are they distinct? If we square each of these modulo 5, we obtain

$x \pmod 5$	0	1	2	3	4
$x^2 \pmod 5$	0	1	4	4	1

Notice that $1 \sim 4$, so they share an equivalence class. Similarly $2 \sim 3$. Indeed the distinct equivalence classes are

$$[0] = \{x \in \mathbb{Z} : x \equiv 0 \pmod 5\}$$
$$[1] = \{x \in \mathbb{Z} : x \equiv 1, 4 \pmod 5\}$$
$$[2] = \{x \in \mathbb{Z} : x \equiv 2, 3 \pmod 5\}$$

In this case the quotient is the set

$$\mathbb{Z}/_\sim = \Big\{[0], [1], [2]\Big\}.$$

Here is one further example of an equivalence relation, this time on $\mathbb{R}^2$. Be careful with the notation: $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is already a Cartesian product, so a relation on $\mathbb{R}^2$ is a subset of $\mathbb{R}^2 \times \mathbb{R}^2$!

Example. Let $\sim$ be the relation on $\mathbb{R}^2$ defined by $(x, y) \sim (v, w) \iff x^2 + y^2 = v^2 + w^2$. We claim that this is an equivalence relation.

Reflexivity $\forall (x, y) \in \mathbb{R}^2,\ x^2 + y^2 = x^2 + y^2$.

Symmetry $\forall (x, y), (v, w) \in \mathbb{R}^2,\ (x, y) \sim (v, w)$
$$\implies x^2 + y^2 = v^2 + w^2$$
$$\implies v^2 + w^2 = x^2 + y^2$$
$$\implies (v, w) \sim (x, y)$$

Transitivity $\forall (x, y), (v, w), (p, q) \in \mathbb{R}^2$, if $(x, y) \sim (v, w)$ and $(v, w) \sim (p, q)$, then $x^2 + y^2 = v^2 + w^2$ and $v^2 + w^2 = p^2 + q^2$. But then $x^2 + y^2 = p^2 + q^2$ and so $(x, y) \sim (p, q)$.

$\sim$ is therefore an equivalence relation. But what are the equivalence classes? By definition,

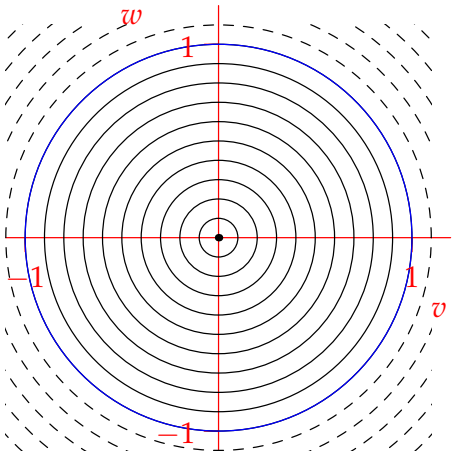
$$[(x, y)] = \{(v, w) \in \mathbb{R}^2 : v^2 + w^2 = x^2 + y^2\}.$$

This isn't particularly helpful. Indeed it is easier to think of each of these sets as

$$\Big\{(v,w) \in \mathbb{R}^2 : v^2 + w^2 \text{ is } \textit{constant}\Big\}.$$

Each equivalence class is therefore a *circle* centered at the origin! Some of the equivalence classes are drawn in the picture: the class $[(1,0)]$ is highlighted. Moreover, the quotient set is

$$\mathbb{R}^2/_{\sim} = \{\text{circles centered at the origin}\}.$$



Exercises

7.3.1 A relation R is *antisymmetric* if $((x,y) \in R) \wedge ((y,x) \in R) \implies x = y$. Give examples of relations R on $A = \{1,2,3\}$ having the stated property.

(a) R is both symmetric and antisymmetric.

(b) R is neither symmetric nor antisymmetric.

(c) R is transitive but $R \cup R^{-1}$ is not transitive.

7.3.2 Let $S = \{(x,y) \in \mathbb{R}^2 : \sin^2 x + \cos^2 y = 1\}$.

(a) Give an example of two real numbers x, y such that $x \sim y$.

(b) Is S reflexive? Symmetric? Transitive? Justify your answers.

7.3.3 Each of the following relations $\sim$ is an equivalence relation on $\mathbb{R}^2$. Identify the equivalence classes and draw several of them.

(a) $(a,b) \sim (c,d) \iff ab = cd$.

(b) $(v,w) \sim (x,y) \iff v^2 w = x^2 y$.

7.3.4 (a) Let $\sim$ be the relation defined on $\mathbb{Z}$ by $a \sim b \iff a + b$ is even. Show that $\sim$ is an equivalence relation and determine the distinct equivalence classes.

(b) Suppose that 'even' is replaced by 'odd' in part (a). Which of the properties reflexive, symmetric, transitive does $\sim$ possess?

7.3.5 For each of the following relations R on $\mathbb{Z}$, decide whether R is reflexive, symmetric, or transitive, and whether R is an equivalence relation.

(a) $a \,R\, b \iff a \equiv b \pmod 3$ **or** $a \equiv b \pmod 4$.

(b) $a \,R\, b \iff a \equiv b \pmod 3$ **and** $a \equiv b \pmod 4$.

7.3.6 We call a real number x *small* if $|x| \leq 1$. Let R be the relation on the set of real numbers defined by

$$x \,R\, y \iff x - y \text{ is small.}$$

Prove or disprove: R is an equivalence relation on $\mathbb{R}$.

7.3.7 Let $A = \{1, 2, 3, 4, 5, 6\}$. The distinct equivalence classes resulting from an equivalence relation R on A are $\{1, 4, 5\}$, $\{2, 6\}$, and $\{3\}$. What is R? Give your answer as a subset of $A \times A$.

7.3.8 $\subseteq$ is a relation on any set of sets. Is $\subseteq$ reflexive, symmetric, transitive? Prove your assertions.

7.3.9 Let S be the set of all polynomials of degree at most 3. An element $s \in S$ can then be expressed as

$$s(x) = ax^3 + bx^2 + cx + d, \qquad \text{where } a, b, c, d \in \mathbb{R}.$$

A relation R on S is defined by

$$p \, R \, q \iff p \text{ and } q \text{ have a common root.}$$

For example $p(x) = (x-1)^2$ and $q(x) = x^2 - 1$ have the root 1 in common so that $p \, R \, q$. Determine which of the properties reflexive, symmetric and transitive are possessed by R.

7.3.10 Let $A = \{2^m : m \in \mathbb{Z}\}$. A relation $\sim$ is defined on the set $\mathbb{Q}^+$ of positive rational numbers by

$$a \sim b \iff \frac{a}{b} \in A$$

(a) Show that $\sim$ is an equivalence relation.

(b) Describe the elements in the equivalence class $[3]$.

7.3.11 A relation is defined on the set $A = \{a + b\sqrt{2} : a, b \in \mathbb{Q},\ a + b\sqrt{2} \neq 0\}$ by $x \sim y \iff \frac{x}{y} \in \mathbb{Q}$. Show that $\sim$ is an equivalence relation and determine the distinct equivalence classes.

7.3.12 The *reflexive, symmetric* and *transitive closures* of a relation R are defined respectively as the smallest relations containing R which also exhibit the given property. Find each of the three closures of $R = \{(1, 2), (2, 3), (3, 3)\} \subseteq \mathbb{Z} \times \mathbb{Z}$.

7.3.13 Recall the description of the real projective line (page 110): if A_m is the line through the origin with gradient m, then

$$\mathbb{P}(\mathbb{R}^2) = \{A_m : m \in \mathbb{R} \cup \{\infty\}\}.$$

Define a relation on $\mathbb{R}^2_* = \mathbb{R}^2 \setminus \{(0, 0)\}$ by $(a, b) \sim (c, d) \iff ad = bc$.

(a) Prove that $\sim$ is an equivalence relation.

(b) Find the equivalence classes of $\sim$. How do the equivalence classes differ from the lines A_m?

7.3.14 Suppose that R, S are relations on some set X. Define the *composition* $R \circ S$ to be the relation

$$(a, c) \in R \circ S \iff \exists b \in X \text{ such that } (a, b) \in R \text{ and } (b, c) \in S.$$

(a) If $R = \{(1, 1), (1, 2), (2, 3), (3, 1), (3, 3)\}$ and $S = \{(1, 2), (1, 3), (2, 1), (3, 3)\}$, find $R \circ S$.

(b) Suppose that R and S are reflexive. Prove that $R \circ S$ is reflexive.

(c) Suppose that P and Q are symmetric. Prove that $(x, y) \in P \circ Q \iff (y, x) \in Q \circ P$.

(d) Give an example of symmetric relations P, Q such that $P \circ Q$ is *not* symmetric. Conclude that if P, Q are equivalence relations, then $P \circ Q$ need not be an equivalence relation.

7.3.15 (*Only for those who have studied Linear Algebra*) Let $\sim$ be the relation on the set of 2×2 real matrices given by $A \sim B \iff \exists M$ such that $B = MAM^{-1}$.

(a) Prove that $\sim$ is an equivalence relation.

(b) What is the equivalence class of the identity matrix?

(c) Show that $\left(\begin{smallmatrix} -11 & 15 \\ -5 & 9 \end{smallmatrix}\right) \sim \left(\begin{smallmatrix} 4 & 10 \\ 0 & -6 \end{smallmatrix}\right)$ (*Hint: think about diagonalizing*)

(d) (Hard) Suppose that $L : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map and $\mathcal{U}, \mathcal{V}$ are bases of $\mathbb{R}^2$. Suppose that $A = [L]_{\mathcal{U}}$ and $B = [L]_{\mathcal{V}}$ are the matrix representations of L with respect to the two bases. Prove that $A \sim B$.

(e) (Hard) Suppose that A, B have the same, but distinct, eigenvalues $\lambda_1 \neq \lambda_2$. Prove that $A \sim B$. *Again use diagonalization, the challenge here is to make your proof work even when the eigenvalues are complex numbers.*

7.4 Partitions

Recall the important observation about our equivalence relation examples: every element of the original set of objects ends up in *exactly one equivalence class.* For instance, every integer is either even or odd but not both. The equivalence classes *partition* the original set in the same way that cutting a cake partitions the crumbs: each crumb ends up in exactly one slice. We shall prove in a moment that equivalence relations *always* do this. Before doing so we reverse the discussion.

Definition 7.10. Let X be a set and $\mathcal{A} = \{A_n : n \in I\}$ be a collection of non-empty subsets $A_n \subseteq X$. We say that X is *partitioned by* $\mathcal{A}$ if

1. $X = \bigcup_{n \in I} A_n$. (the A_n together make up X)
2. If $A_m \neq A_n$, then $A_m \cap A_n = \emptyset$. (*distinct* A_n are pairwise disjoint[a])

We describe the collection $\mathcal{A}$ as a *partition* of X.

[a] Recall that two sets A, B are *disjoint* if $A \cap B = \emptyset$: see Definition 4.6. In this definition we *don't* require the sets A_n to all be different, some could be identical to each other.

Example. Partition the set $X = \{1,2,3,4,5\}$ into subsets $A_1 = \{1,3\}$, $A_2 = \{2,4\}$ and $A_3 = \{5\}$. Now consider the relation R on X, defined by

$$R = \{(1,1),(1,3),(3,1),(3,3),(2,2),(2,4),(4,2),(4,4),(5,5)\}.$$

What does R have to do with the partition? R was constructed by insisting that

$$x \, R \, y \iff x \text{ and } y \text{ are in the } \textit{same} \text{ subset } A_n.$$

Run through your mental checklist: reflexive? symmetric? transitive? Indeed R is an equivalence relation! Moreover, the equivalence classes of R are exactly the sets A_1, A_2, A_3. For example, because 1 belongs to A_1, the element 1 should be related to every other element in A_1. Therefore, the pairs $(1,1)$ and $(1,3)$ should be in R.

The example suggests that partitioning a set actually defines an equivalence relation. Combining this with our previous observation you should be starting to belive that *partitions and equivalence relations are essentially the same thing.*

Examples.

1. The integers can be partitioned according to their remainder modulo 3: define

   $$A_m = \{z \in \mathbb{Z} : z \equiv m \pmod 3\},$$

   then $\mathbb{Z} = A_0 \cup A_1 \cup A_2$. This is certainly a partition:

   - Every integer z has remainder of 0, 1 or 2 after division by 3, and so every integer is in some set A_m.
   - No integer has two distinct remainders modulo 3, so the sets A_0, A_1, A_2 are disjoint.

2. More generally, if $n \in \mathbb{N}$, then the set of integers $\mathbb{Z}$ is partitioned into n sets $A_0, \ldots, A_{n-1}$ where $A_m = \{z \in \mathbb{Z} : z \equiv m \pmod n\}$ is the set of integers with remainder m upon dividing by n.

3. $\mathbb{R}$ is partitioned by the sets of rational and irrational numbers: $\mathbb{R} = \mathbb{Q} \cup (R \setminus \mathbb{Q})$.

Finally, here is an example of a relation which doesn't produce a partition.

Example. Let $R = \{(1,3),(1,4),(2,2),(2,3),(3,1),(3,2),(4,3),(4,4)\}$ be a relation on $X = \{1,2,3,4\}$ and define the sets

$$A_n = \{x \in X : (n,x) \in R\}.$$

Thus A_n is the set of all elements of X which are related to n. We quickly see that

$$A_1 = \{3,4\}, \quad A_2 = \{2,3\}, \quad A_3 = \{1,2\}, \quad A_4 = \{3,4\}.$$

The collection of sets A_n is as follows:

$$\{A_n\}_{n\in X} = \{A_1, A_2, A_3, A_4\} = \{\{3,4\},\{2,3\},\{1,2\}\},$$

where we only have *three* sets in the collection since $A_4 = A_1$. This collection is not a partition because, for instance, $2 \in \{2,3\} \cap \{1,2\}$. In the language of the definition, $\{2,3\} \neq \{1,2\}$ but $\{2,3\} \cap \{1,2\} \neq \emptyset$. More importantly, you should convince yourself that R is *not* an equivalence relation.

Before we present the fundamental result of the chapter, we prove a lemma.

Lemma 7.11. *Suppose that $\sim$ is an equivalence relation. Then $x \sim y \iff [x] = [y]$.*

Proof. ($\Leftarrow$) If $[x] = [y]$, then $x \in [y]$, whence $x \sim y$.
($\Rightarrow$) Suppose that $x \sim y$. We begin by showing the inclusion $[x] \subseteq [y]$. Let $z \in [x]$, then

$$z \sim x \text{ and } x \sim y \implies z \sim y \implies z \in [y]. \qquad \text{(Transitivity)}$$

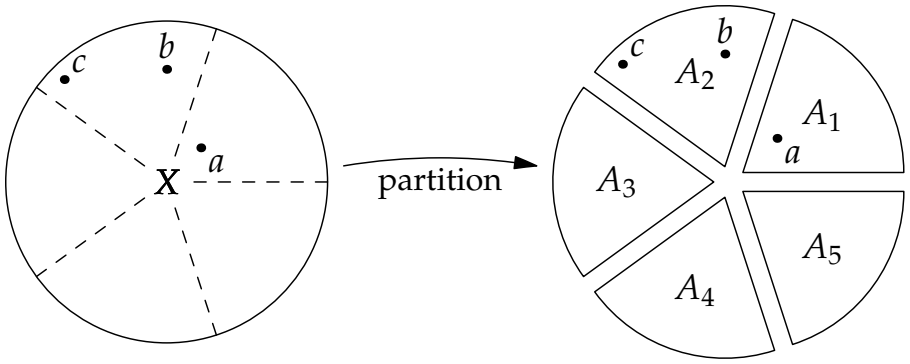
Therefore $[x] \subseteq [y]$. The argument is symmetric in x and y, so we also have $[y] \subseteq [x]$, and thus $[x] = [y]$. ∎

Theorem 7.12. *Let X be a set.*

1. *If $\sim$ is an equivalence relation on X, then X is partitioned by the equivalence classes of $\sim$.*
2. *If $\{A_n\}_{n\in I}$ is a partition of X, then the relation $\sim$ on X defined by*

   $$x \sim y \iff \exists n \in I \text{ such that } x \in A_n \text{ and } y \in A_n$$

   is an equivalence relation.



Each element of X ends up in exactly one subset. In the language of the Theorem, we have

$$A_1 = [a], \quad A_2 = [b] = [c], \quad b \sim c, \quad a \nsim b, \quad a \nsim c.$$

Some things to consider while reading the proof:

- Keep your eyes on the picture: it's where your intuition comes from, and it's how you should remember the result. The algebra merely confirms that the picture is telling a legitimate story.
- In part 1. of the proof, look for where the reflexive, symmetric and transitive assumptions about $\sim$ are used. Why do we need $\sim$ to be an equivalence relation?
- Similarly, in part 2., look for where we use both parts of the defintion of partition. Why are both assumptions required?

Proof. 1. Assume that $\sim$ is an equivalence relation on X. To prove that the equivalence classes of $\sim$ partition X, we must show two things:

(a) That every element of X is in some equivalence class.

(b) That the distinct equivalence classes are pairwise disjoint: if $[x] \neq [y]$, then $[x] \cap [y] = \emptyset$.

For (a), we only need reflexivity: $\forall x \in X$ we have $x \sim x$. Otherwise said, $x \in [x]$, whence every element of X is in the equivalence class defined by itself.

For (b), we prove by the contrapositive method and show that $[x] \cap [y] \neq \emptyset \implies [x] = [y]$. Assume that $[x] \cap [y] \neq \emptyset$. Then $\exists z \in [x] \cap [y]$. This gives

$$\begin{aligned} z \sim x \text{ and } z \sim y &\implies x \sim z \text{ and } z \sim y && \text{(Symmetry)} \\ &\implies x \sim y && \text{(Transitivity)} \\ &\implies [x] = [y] && \text{(Lemma 7.11)} \end{aligned}$$

We have proved (b) and therefore part 1. of the theorem.

2. Now suppose that $\{A_n\}_{n \in I}$ is a partition of X and define $\sim$ by

$$x \sim y \iff \exists n \in I \text{ such that } x \in A_n \text{ and } y \in A_n.$$

We must prove the reflexivity, symmetry and transitivity of $\sim$.

Reflexivity Every $x \in X$ is in some A_n. Thus $x \sim x$ for all $x \in X$.

Symmetry If $x \sim y$, then $\exists n \in I$ such that $x, y \in A_n$. But then $y, x \in A_n$ and so $y \sim x$.

Transitivity Let $x \sim y$ and $y \sim z$. Then $\exists p, q \in I$ such that $x, y \in A_p$ and $y, z \in A_q$. Since $\{A_n\}_{n \in I}$ is a partition and $y \in A_p \cap A_q$, we necessarily have $p = q$. Thus $x, z \in A_p$ and so $x \sim z$.

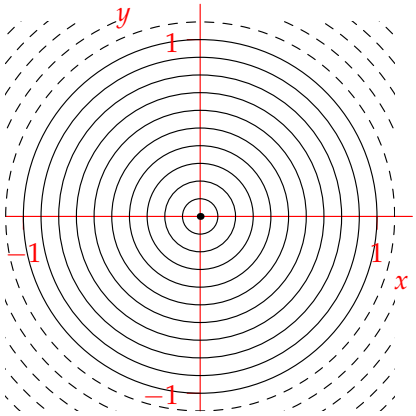
Thus $\sim$ is an equivalence relation. ■

Reading the proof carefully, you should see that reflexivity comes from the fact that $X = \bigcup_{n \in I} A_n$, while transitivity is due to the pairwise disjointness of the parts of the partition. Symmetry is essentially free because the definition of $\sim$ is symmetric in x and y.

Examples of partitions are especially easy to see with curves in the plane. Here we return to the example on page 127 and describe things in our new language.

Example. For each real number $r \geq 0$, define the set

$$A_r = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = r^2\}.$$



This is simply the circle of radius r centered at the origin. We check that $\{A_r\}_{r \in \mathbb{R}_0^+}$ is a partition of $\mathbb{R}^2$.

- Every point of the plane lies on some circle. Precisely, $(x, y) \in A_{\sqrt{x^2+y^2}}$ since $\sqrt{x^2 + y^2}$ is the distance of (x, y) from the origin. Thus $\mathbb{R}^2 = \bigcup_{r \in \mathbb{R}_0^+} A_r$.
- If $r_1 \neq r_2$, then the concentric circles A_{r_1} and A_{r_2} do not intersect. Thus $A_{r_1} \cap A_{r_2} = \emptyset$.

Now define a relation $\sim$ on $\mathbb{R}^2$ via

$$(x, y) \sim (v, w) \iff \exists r \geq 0 \text{ such that } (x, y), (v, w) \text{ both lie on the circle } A_r.$$

By Theorem 7.12 this is an equivalence relation. We can also check explicitly: dropping any mention of the radius r, we see that

$$(x, y) \sim (v, w) \iff x^2 + y^2 = v^2 + w^2.$$

This is exactly the equivalence relation described on page 127. The equivalence classes are precisely the sets A_r. Indeed

$$[(v, w)] = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = v^2 + w^2\} = A_{\sqrt{v^2+w^2}}$$

is just the circle of radius $\sqrt{v^2 + w^2}$.

Geometric Examples

The language of equivalence relations and partitions is used heavily in geometry and topology to describe complex shapes. Here are a couple of examples.

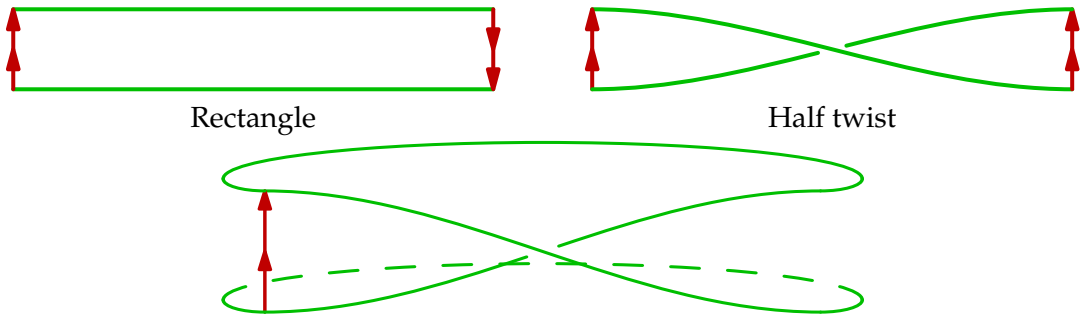
The Möbius Strip Take a rectangle $X = [0,6] \times [0,1]$ and partition into the following subsets.

- If a point does not lie on the left or right edge of the rectangle, place it in a subset by itself: $\{(x,y)\}$ for $x \neq 0, 6$,
- If a point does lie on the left or right edge of the rectangle, place it in a subset with one point from the other edge: $\{(0,y),(6,1-y)\}$ for any y.

The rectangle is drawn below, where the points on the left and right edges are colored red. The arrows indicate how the edges are paired up. For example the point $(0, 0.8)$ (high on the left near the tip of the arrow) is paired with $(6, 0.2)$ (low on the right edge of the rectangle).

These subsets clearly partition the rectangle X. The partitions define an equivalence relation $\sim$ on X in accordance with Theorem 7.12. Note that there are infinitely many equivalence classes. How can we interpret the quotient set $X/\sim$?

This is easier to visualize than you might think. Since each point on the left edge of the rectangle is in an equivalence class with a point on the right edge, we imagine gluing the two edges together in such a way that the correspoinding points are touching. In the picture, we imagine holding X like a strip of paper, giving one side a twist, and then gluing the edges together. This is the classic construction of a Möbius strip.



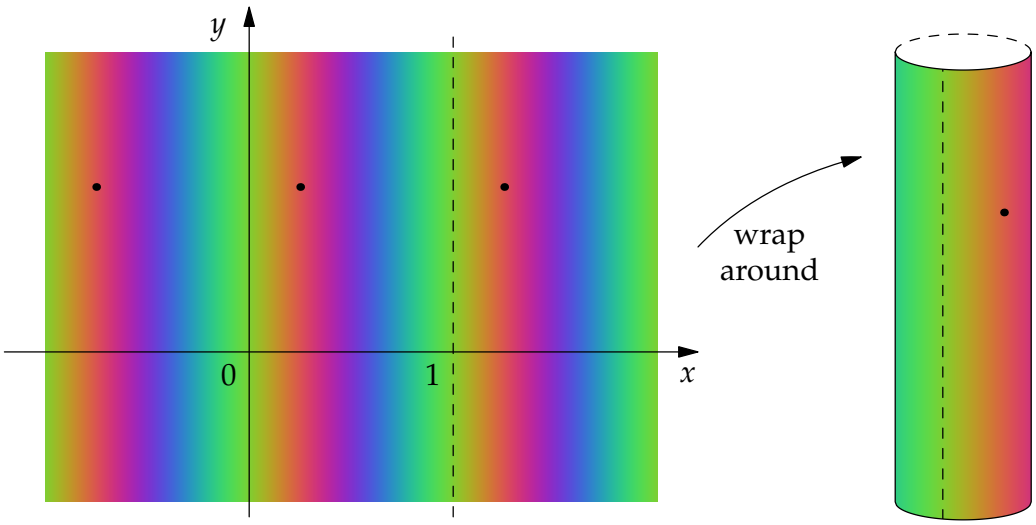
The Cylinder One could construct a cylinder similarly to the Möbius strip, by identifying edges of the rectangle but *without* applying the half-twist. Instead we do something a little different.

Let $X = \mathbb{R}^2$ with equivalence relation $\sim$ defined by

$$(a,b) \sim (c,d) \iff a - c \in \mathbb{Z} \quad \textit{and} \quad b = d.$$

The equivalence classes are horizontal strings of points with the same y co-ordinate. If we imagine wrapping $\mathbb{R}^2$ repeatedly around a cylinder of circumference 1, all of the points in a given equivalence class will now line up. The set of equivalence classes $\mathbb{R}^2/\sim$ can therefore be vizualized as the cylinder.

Alternatively, you may imagine piercing a roll of toilet paper and unrolling it. The single puncture now becomes a row of (almost![27]) equally spaced holes. In the picture, the left hand side is (part of) the plane $\mathbb{R}^2$, displayed so that points in each equivalence class have the same color. The three horizontal dots are all in the same equivalence class. We roll up the plane into a cyclinder so that all the points with the same color end up at the same place.



More complex shapes can be created by other partitions/relations. If you want a challenge in visualization, consider why the equivalence relation

$$(a,b) \sim (c,d) \iff a-c \in \mathbb{Z} \quad \textit{and} \quad b-d \in \mathbb{Z}$$

on $\mathbb{R}^2$ defines a torus (the surface of a ring-doughnut).

Exercises

7.4.1 For each of the collections $\{A_n\}_{n\in\mathbb{R}}$, determine whether the collections partition $\mathbb{R}^2$. Justify your answers, and sketch several of the sets A_n.

(a) $A_n = \{(x,y) \in \mathbb{R}^2 : y = 2x + n\}$.
(b) $A_n = \{(x,y) \in \mathbb{R}^2 : y = (x-n)^2\}$.
(c) $A_n = \{(x,y) \in \mathbb{R}^2 : xy = n\}$.
(d) $A_n = \{(x,y) \in \mathbb{R}^2 : y^4 - y^2 = x - n\}$.

7.4.2 Let X be the set of all humans. If $x \in X$, we define the set

$$A_x = \{\text{people who had the same breakfast } \textit{or} \text{ lunch as } x\}.$$

(a) Does the collection $\{A_x\}_{x\in X}$ partition X? Explain.
(b) Is your answer different if the *or* in the definition of A_x is changed to *and*?

[27] Unfortunately for the analogy, toilet paper has purposeful thickness!

If Jane and Tom had both had the same breakfast and lunch, then $A_{Jane} = A_{Tom}$ so there are likely many fewer distinct *sets A_x than there are humans!*

7.4.3 Let $X = \{1,2,3\}$. Define the relation $R = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,3)\}$ on X.

(a) Which of the properties reflexive, symmetric, transitive does R satisfy?

(b) Compute the sets A_1, A_2, A_3 where $A_n = \{x \in X : x\,R\,n\}$. Show that $\{A_1, A_2, A_3\}$ do not form a partition of X.

(c) Repeat parts (a) and (b) for the relations S and T on X, where

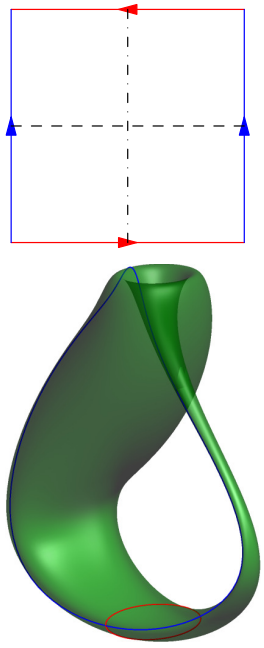
$$S = \{(1,1),(1,3),(3,1),(3,3)\}$$
$$T = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,3)\}$$

Some of the sets A_1, A_2, A_3 might be the same in each of your examples. If, for example, $A_1 = A_3$, then the collection $\{A_1, A_2, A_3\}$ only contains two sets: $\{A_1, A_2\}$. Is this a partition? Compare with the example on page 132.

7.4.4 Using the equivalence relation description of the Möbius strip, prove that you may cut a Möbius strip round the middle and yet still end up with a single loop.
Where would you cut the defining rectangle and how can you tell that you still have one piece?

7.4.5 (Hard!) A *Klein bottle* can be visualized as follows. Define an equivalence relation $\sim$ on the unit square $X = [0,1] \times [0,1]$ so that:

- $(0,y) \sim (1,y)$ for $0 \le y \le 1$.
- $(x,0) \sim (1-x,1)$ for $0 \le x \le 1$.



The result is the picture: the blue edges are identified in the same direction and the red in the opposite. Attempting to visualize this in 3D requires a willingness to stretch and distort the square, but results in the green bottle. The original red an blue arrows have become curves on the bottle. If you are using Acrobat Reader, click on the bottle and move it around.

(a) Suppose you cut the Klein bottle along the horizontal dashed line of the defining square. What is the resulting object?

(b) Now cut the bottle along the vertical dashed line. What do you get this time?

Can you visualize where the two dashed lines are on the green bottle?

7.5 Well-definition, Rings and Congruence

We return to our discussion of congruence (recall Section 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 7.13. *For each $n \in \mathbb{N}$, define $x \sim_n y \iff x \equiv y \pmod{n}$. Then $\sim_n$ is an equivalence relation on $\mathbb{Z}$.*

The theorem is a restatement of Example 2 on page 131, in conjunction with Theorem 7.12. You should prove this yourself, as practice in using the definition of equivalence relation.

The equivalence classes are precisely those integers which are congruent modulo n: the integers which share the same remainder.

$$\begin{aligned}[a] &= \{x \in \mathbb{Z} : x \equiv a \pmod{n}\} \\ &= \{x \in \mathbb{Z} : x \text{ has the same remainder as } a \text{ when divided by } n\} \\ &= \{x \in \mathbb{Z} : x - a \text{ is divisible by } n\}\end{aligned}$$

In this language, we may restate what it means for two equivalence classes to be identical.

Theorem 7.14. $[a] = [b] \iff a \equiv b \pmod{n} \iff \exists k \in \mathbb{Z}$ *such that* $b = a + kn$.

If the meaning of *any* of the above is unclear, re-read the previous two sections: they are critically important!

The equivalence classes of $\sim_n$ partition the integers $\mathbb{Z}$. According to Theorem 7.14, there are exactly n equivalence classes, whence we may describe the quotient set as

$$\mathbb{Z}/_{\sim_n} = \{[0], [1], \ldots, [n-1]\}.$$

We use this set to define an extremely important object.

Definition 7.15. Define two operations $+_n$ and $\cdot_n$ on the set $\mathbb{Z}/_{\sim_n}$ as follows:

$$[x] +_n [y] := [x + y], \qquad [x] \cdot_n [y] := [x \cdot y].$$

The *ring* $\mathbb{Z}_n$ is the set $\mathbb{Z}/_{\sim_n}$ together with the operations $+_n$ and $\cdot_n$.

The operation $+_n$ (similarly $\cdot_n$) is telling us how to add *equivalence classes,* that is, how to produce a new equivalence class from two old ones. $+_n$ is not the same operation as $+$: we are *defining* $+_n$ using $+$. The former combines equivalence classes, while the latter sums integers.

The challenge here is that you have to think of each equivalence class as a single object. When we write

$$[3] +_8 [6] = [3 + 6] = [9] = [1],$$

we are thinking about the sets $[3]$ and $[6]$ as individual objects rather than as collections of elements: remember that $[3] = \{\ldots, -5, 3, 11, 19, \ldots\}$ is an infinite set! There is, moreover, a matter of choice: since, for example, $[3] = [11]$ and $[6] = [22]$ we should be able to observe that

$$[3] +_8 [6] = [11] +_8 [22].$$

Is this true? If not, then the operation $+_8$ would not be particularly useful. Thankfully this is not a problem: according to the defintion of $+_8$, we have

$$[11] +_8 [22] = [11 + 22] = [33] = [1],$$

exactly as we would wish.

Let us think a little more abstractly. Suppose we are given equivalence classes X and Y, how do we compute $X +_n Y$? Here is the process.

1. *Choose* elements $x \in X$ and $y \in Y$.
2. Add x and y to get a new element $x + y \in \mathbb{Z}$.
3. Then $X +_n Y$ is the equivalence class $[x + y]$.

The issue is that there are *infinitely many choices* for the elements $x \in X$ and $y \in Y$. If $+_n$ is to make sense, we must obtain the *same* equivalence class $[x + y]$ **regardless of our choices of** $x \in X$ **and** $y \in Y$**.**

Definition 7.16. A concept is *well-defined* if it is *independent of all choices used in the definition.*

Theorem 7.17. *The operations $+_n$ and $\cdot_n$ are well-defined.*

The choices made in the definitions of $+_n$ and $\cdot_n$ were of representative elements x and y of the equivalence classes $[x]$ and $[y]$. All representatives of these classes have the form

$$x + kn \in [x] \quad \text{and} \quad y + ln \in [y]$$

for some integers k, l. It therefore suffices to prove that

$$\forall k, l \in \mathbb{Z}, \quad [x + kn] +_n [y + ln] = [x] +_n [y] \quad \text{and} \quad [x + kn] \cdot_n [y + ln] = [x] \cdot_n [y].$$

Proof. We prove that $+_n$ is well-defined.

$$\begin{aligned}
[x+kn] +_n [y+ln] &= [(x+kn)+(y+ln)] && \text{(by definition of } +_n\text{)}\\
&= [x+y+(k+l)n] \\
&= [x+y] && \text{(by Theorem 7.14)}\\
&= [x] +_n [y] && \text{(by definition of } +_n\text{)}
\end{aligned}$$

The argument for $\cdot_n$ is similar. ∎

You should now re-read Theorem 3.8 until you are comfortable that we are doing the same thing!

Aside: Ugly notation

Given the usefulness of $\mathbb{Z}_n$, and the cumbersome nature of the above notation, it is customary to drop the square brackets and subscripts and simply write

$$\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}, \qquad x + y := x + y \pmod{n}, \qquad x \cdot y := xy \pmod{n}.$$

When using this description of $\mathbb{Z}_n$, you should realize that we are working with equivalence classes, not numbers. In this context, $-3 \in \mathbb{Z}_8$ makes perfect sense, for it really means $[-3] \in \mathbb{Z}_8$. This is percectly fine, since $[-3] = [5]$ as equivalence classes, and so it is legitimate to write $-3 = 5$ in $\mathbb{Z}_8$. Until you are 100% sure that you know when 3 represents an equivalence class and when it represents a number, you should keep the brackets in place!

Exercises

7.5.1 Give an explicit proof of Theorem 7.13.

7.5.2 (a) Prove the second half of Theorem 7.17, that $\cdot_n$ is well-defined.

(b) Prove by induction that the operation of raising to the power $m \in \mathbb{N}$ is well-defined in $\mathbb{Z}_n$. I.e., prove that

$$\forall m \in \mathbb{N},\ \forall [x] \in {}^{\mathbb{Z}}\!/_{\sim_n} \text{ we have } [x^m] = [x]^m.$$

Be careful! n is fixed, your induction variable is m. What base case(s) do you need?

7.5.3 Consider the relation $\sim$ defined on $\mathbb{Z} \times \mathbb{N} = \{(x, y) : x \in \mathbb{Z}, \text{ and } y \in \mathbb{N}\}$ by

$$(a, b) \sim (c, d) \iff ad = bc.$$

(a) Prove that $\sim$ is an equivalence relation.

(b) List several elements of the equivalence class of $(2, 3)$. Repeat for the equivalence class of $(-3, 7)$. What do the equivalence classes have to do with the set of rational numbers $\mathbb{Q}$?

(c) Define operations $\oplus$ and $\otimes$ on ${}^{\mathbb{Z} \times \mathbb{N}}\!/_{\sim}$ by

$$[(a, b)] \oplus [(c, d)] = [(ad + bc, bd)], \qquad [(a, b)] \otimes [(c, d)] = [(ac, bd)].$$

Prove that $\oplus$ and $\otimes$ are well-defined.

Try to do this question without *using division! We will return to this example in the next section.*

7.6 Functions and Partitions

To complete our discussion of partitions and equivalence relations, we consider how to define functions whose domain is a set of equivalence classes. Take congruence as our motivating example.

Suppose we want to define a function $f : \mathbb{Z}_4 \to \mathbb{Z}_6$. Say $f(x) = 3x \pmod 6$. This certainly looks like a function, but is it? Remember that 'x' and '$3x$' are really equivalence classes, so we should say[28]

$$f\big([x]_4\big) = [3x]_6, \qquad \text{where} \quad [x]_4 \in \mathbb{Z}_4 \quad \text{and} \quad [3x]_6 \in \mathbb{Z}_6.$$

Is *this* a function? To make sure, we need to check that *any* representative $a \in [x]_4$ gives the same result. That is, we need to prove that

$$a \equiv b \pmod 4 \implies 3a \equiv 3b \pmod 6.$$

This is not so hard:

$$\begin{aligned} a \equiv b \pmod 4 &\implies \exists n \in \mathbb{Z} \text{ such that } a = b + 4n \\ &\implies 3a = 3b + 12n \implies 3a \equiv 3b \pmod 6. \end{aligned}$$

It might look like a small difference, but attempting to define $g : \mathbb{Z}_4 \to \mathbb{Z}_6$ by $g(x) = 2x \pmod 6$ does *not* result in a function. If it were, then we should have

$$a \equiv b \pmod 4 \implies 2a \equiv 2b \pmod 6.$$

But this is simply not true: for example $4 \equiv 0 \pmod 4$, but $8 \not\equiv 0 \pmod 6$. It might look like g is a function, but it is not well-defined because $[4] = [0]$ in $\mathbb{Z}_4$ and $g\big([4]\big) \neq g\big([0]\big)$ in $\mathbb{Z}_6$.

Just as in Definition 7.16, the process of verifying that a rule really is a function is called checking *well-definition.* In general, if we are defining a function

$$f : {}^{X}\!/_{\sim} \to A \tag{$*$}$$

whose domain is a quotient set, then it is usually necessary to construct f by saying what happens to a *representative* x of an equivalence class $[x]$:

$$f\big([x]\big) = \text{ 'do something to } x\text{'}.$$

We need to make sure that the 'something' is *independent of the choice of element* x.

Definition 7.18. Suppose that $f : {}^{X}\!/_{\sim} \to A$ is a rule of the form $(*)$. We say that f is a *well-defined* function if

$$[x] = [y] \implies f\big([x]\big) = f\big([y]\big).$$

If you think carefully, this is nothing more than condition 2 of Definition 7.4.

Examples. 1. Show that $f : \mathbb{Z}_n \to \mathbb{Z}_n$ defined by $f(x) = x^2 + 4 \pmod n$ is well-defined.

We must check that $x \equiv y \pmod n \implies x^2 + 4 \equiv y^2 + 4 \pmod n$. But this is trivial!

[28] The notation $[x]_4$ is helpful for reminding us which equivalence relation is being applied. When dealing with functions between different quotient sets, it is easy to become confused.

2. For which integers k is the rule $f : \mathbb{Z}_4 \to \mathbb{Z}_6$ defined by $f(x) = kx \pmod 6$ a well-defined function?
We require $x \equiv y \pmod 4 \implies kx \equiv ky \pmod 6$. Now

$$\begin{aligned} x \equiv y \pmod 4 &\implies \exists n \in \mathbb{Z} \text{ such that } x - y = 4n \\ &\implies kx - ky = 4kn. \end{aligned}$$

For f to be well-defined, we need $kx - ky = 4kn$ to be a multiple of 6 *independently* of x and y. Thus f is well-defined if and only if $6 \mid 4kn$ *for all* $n \in \mathbb{Z}$. This can only be the case if $6 \mid 4k$. Otherwise said,

$$f \text{ is well-defined} \iff 6 \mid 4k \iff 3 \mid 2k \iff 3 \mid k.$$

Given that $kx \in \mathbb{Z}_6$, we need only consider $k \in \{0, 1, 2, 3, 4, 5\}$: equivalent values of k modulo 6 won't change the definition of f. It follows that there are only *two* well-defined functions $f : \mathbb{Z}_4 \to \mathbb{Z}_6 : x \mapsto kx$, namely $f_0(x) = 0$ and $f_3(x) = 3x$. Here they are in tabular form.

x	0	1	2	3
$f_0(x)$	0	0	0	0

x	0	1	2	3
$f_3(x)$	0	3	0	3

It is instructive to play with another value of k, say $k = 5$, and attempt to construct a table:

x	0	1	2	3	4	5	$\cdots$
$f_5(x)$	0	5	4	3	2	1	$\cdots$

The problem is that $4 \equiv 0 \pmod 4$, yet $f_5(4) \not\equiv f_5(0) \pmod 6$. In order to be a function, the second row must repeat with period four. You should compare this with the examples on page 67 and with Exercise 4.4.11.

Functions on the Cylinder and Torus

Recall our construction on page 135, where we viewed the cylinder as the set $\mathbb{R}^2/_\sim$ with respect to the equivalence relation

$$(a, b) \sim (c, d) \iff a - c \in \mathbb{Z} \quad \text{and} \quad b = d.$$

We wish to define a function $f : \mathbb{R}^2/_\sim \to A$ whose domain is the cylinder.[29] *Well-definition* requires that f satisfy

$$(a, b) \sim (c, d) \implies f\Big(\big[(a, b)\big]\Big) = f\Big(\big[(c, d)\big]\Big).$$

Since $(a, b) \sim (a + 1, b)$, we require $f\Big(\big[(a, b)\big]\Big) = f\Big(\big[(a + 1, b)\big]\Big)$, for all $a, b \in \mathbb{R}$. Otherwise said, $f\Big(\big[(x, y)\big]\Big)$ must be periodic in x with period 1. It is easy to see that

$$f\Big(\big[(x, y)\big]\Big) = y^2 \sin(2\pi x)$$

is a suitable choice of function $f : \mathbb{R}^2/_\sim \to \mathbb{R}$.

[29] A is any target set you like. We will choose an example with $A = \mathbb{R}$ in a moment.

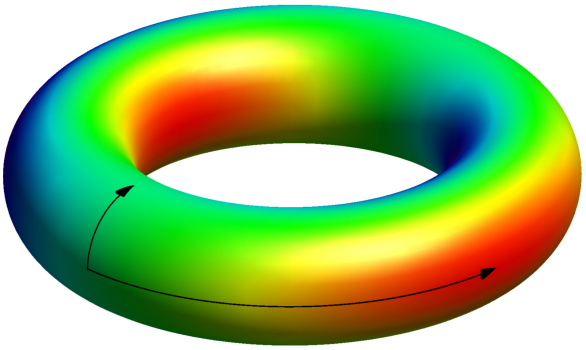
More generally, to define a function whose domain is the torus

$$T^2 = {}^{\mathbb{R}^2}\!/_{\sim} \quad \text{where} \quad (a,b) \sim (c,d) \iff a-c \in \mathbb{Z} \quad \text{and} \quad b-d \in \mathbb{Z},$$

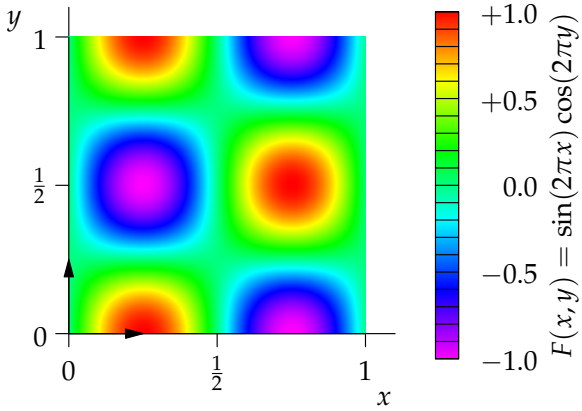
requires a function which has period 1 in *both* x and y. The function $f\Big([(x,y)]\Big) = \sin(2\pi x)\cos(2\pi y)$ is plotted below, with the color on the torus indicating the value of f. It is easier for us to simply consider the function

$$F : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto \sin(2\pi x)\cos(2\pi y).$$

This is also plotted, with the same color for each value.



The function f: domain T^2
The arrows in the two pictures correspond



The function F restricted to $[0,1) \times [0,1)$

Aside: The Canonical Map

To do this justice, and to give you a taste for the details which are necessary in pure mathematics, here is the important definition.

Definition 7.19. Suppose that $\sim$ is an equivalence relation on a set X. The function $\gamma : X \to {}^{X}\!/_{\sim}$ defined by $\gamma(x) = [x]$ is the *canonical map.*[a]

[a] Canonical, in mathematics, just means natural or obvious.

The canonical map has only one purpose; to allow us to construct functions $f : {}^{X}\!/_{\sim} \to A$.

Theorem 7.20. *Suppose that $\sim$ is an equivalence relation on X.*

1. *If $f : {}^{X}\!/_{\sim} \to A$ is a function, then $F : X \to A$ defined by $F = f \circ \gamma$ satisfies $x \sim y \implies F(x) = F(y)$.*
2. *If $F : X \to A$ satisfies $x \sim y \implies F(x) = F(y)$, then there is a unique function $f : {}^{X}\!/_{\sim} \to A$ satisfying $F = f \circ \gamma$.*

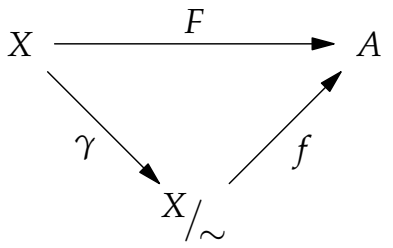
Proof. 1. This is trivial: $x \sim y \implies [x] = [y] \implies \gamma(x) = \gamma(y)$
$\implies f(\gamma(x)) = f(\gamma(y)) \implies F(x) = F(y)$.

2. $f : {}^{X}\!/_{\sim} \to A$ can only be the function *defined* by $f([x]) = F(x)$. We show that this is well-defined:

$$[x] = [y] \implies x \sim y \implies F(x) = F(y) \implies f([x]) = f([y]).$$

■

The proof, like much of mathematics, is a masterpiece in concision that seems to be doing nothing at all. The point is that functions of the form $f : {}^{X}\!/_{\sim} \to A$ are *difficult* to work with. The Theorem says that we never need to explicitly use such functions, and can instead work with *simpler* functions of the form $F : X \to A$. The only condition is that $x \sim y \implies F(x) = F(y)$. Essentially, F is f in disguise!



This result will be resurrected when you study Groups Rings & Fields as part of the famous *First Isomorphism Theorem.*

Exercises

7.6.1 Prove or disprove: $f : \mathbb{Z}_3 \to \mathbb{Z}_5 : x \mapsto x^3 \pmod 5$ is well-defined.

7.6.2 (a) Compute $(x + 4n)^2$.

(b) Suppose that $\forall n \in \mathbb{Z}$, we have $(x + 4n)^2 \equiv x^2 \pmod m$. Find all the integers m for which this is a true statement.

(c) For what $m \in \mathbb{N}_{\geq 2}$ is the function $f : \mathbb{Z}_4 \to \mathbb{Z}_m : x \mapsto x^2 \pmod m$ well-defined.

7.6.3 A rule $f : {}^{X}\!/_{\sim} \to A$ is well-defined if $[x] = [y] \implies f([x]) = f([y])$.

(a) State what it means for $f : {}^{X}\!/_{\sim} \to A$ to be *injective.* What do you observe?

(b) Prove that $f : \mathbb{Z}_7 \to \mathbb{Z}_{35} : x \mapsto 15x$ is a well-defined, injective function.

(c) Repeat part (b) for the function $f : \mathbb{Z}_{100} \to \mathbb{Z}_{300} : x \mapsto 9x$. Compare your arguments for well-definition and injectivity.
This forces you to write your argument abstractly, rather than using a table! You may find it useful that $9 \cdot (-11) \equiv 1 \pmod{100}$.

7.6.4 Define a partition of the sphere $S^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$ into subsets of the form

$$\{(x, y, z), (-x, -y, -z)\}.$$

Each subset consists of two points directly opposite each other on the sphere (antipodal points). Let $\sim$ be the equivalence relation whose equivalence classes are the above subsets.

(a) $f : {}^{S^2}\!/_{\sim} \to \mathbb{R} : [(x, y, z)] \mapsto xyz$ is not well-defined. Explain why.

(b) Prove that $f : {}^{S^2}\!/_{\sim} \to \mathbb{R}^3 : [(x,y,z)] \to (yz, xz, xy)$ is a well-defined function.
The image of this function is Steiner's famous Roman Surface, another example, like the Klein Bottle, of a generalization of the Möbius Strip.

7.6.5 Recall Exercise 7.5.3, where we defined an equivalence relation $\sim$ on $\mathbb{Z} \times \mathbb{N}$.

(a) Prove that the function $f : {}^{\mathbb{Z}\times\mathbb{N}}\!/_{\sim} \to \mathbb{Q}$ defined by $f\Big([(x,y)]\Big) = \frac{x}{y}$ is a *well-defined bijection.*

(b) Prove that f transforms the operations $\oplus$ and $\otimes$ into the usual addition and multiplication of rational numbers. That is:

$$f\Big([(a,b)] \oplus [(c,d)]\Big) = f\Big([(a,b)]\Big) + f\Big([(c,d)]\Big)$$
$$f\Big([(a,b)] \otimes [(c,d)]\Big) = f\Big([(a,b)]\Big) \cdot f\Big([(c,d)]\Big)$$

The technical term for this is that $f : \Big({}^{\mathbb{Z}\times\mathbb{N}}\!/_{\sim}, \oplus, \otimes\Big) \to (\mathbb{Q}, +, \cdot)$ *is an isomorphism of rings.*

8 Cardinalities of Infinite Sets

8.1 Cantor's Notion of Cardinality

During the late 1800's a German mathematician named Georg Cantor almost single-handedly overturned the foundations of mathematics. Prior to Cantor, mathematicians had understood a set to be nothing more than a collection of objects. Via the consideration of certain infinite sets,[30] Cantor demonstrated that this naïve idea is woefully inadequate. Cantor met great resistance from many famous mathematicians, philosophers, and even religious scholars, who felt his ideas were unnatural and risked undermining the divine. Despite strong initial antipathy, Cantor's notion of cardinality is now universally accepted by mathematicians. More importantly, it led to the creation of *axiomatic set theory* and the, still somewhat controversial, modern conception of *set*. Cantor's legacy is arguably the modern axiomatic nature of pure mathematics, where rigor dominates and mathematicians are obliged to follow logic wherever it might lead, regardless of the bizarre paradoxes which might appear.

In this chapter we consider the basics of Cantor's contribution, essentially his extension of the concept of *cardinality* to infinite sets.

Recall that if A is a *finite* set, then $|A|$, the cardinality of A, is simply the number of elements in A. This definition obviously does not extend to infinite sets. However, we can provide an alternative interpretation of cardinality as a tool to *compare* sizes of sets. This interpretation turns out to apply to infinite sets. For example, suppose that

$$A = \{\text{fish}, \text{dog}\}, \quad \text{and} \quad B = \{\alpha, \beta, \gamma\}.$$

Even though the elements of the sets A and B are completely different, we may use cardinality to compare the sizes of A and B: since $|A| = 2$ and $|B| = 3$, we may write $|A| \leq |B|$ to indicate that B has at least as many elements as A. By Theorem 4.12, this condition is equivalent to the existence of an injective (one-to-one) map from A to B. For instance, we can choose the function $f : A \to B$ defined by

$$\text{fish} \longmapsto \alpha, \qquad \text{dog} \longmapsto \beta.$$

In a sense, Theorem 4.12 tells us how to compare cardinalities of finite sets *without* counting elements. Cantor's seemingly innocuous idea was to turn this *theorem* for finite sets into a *definition* of cardinality for infinite sets.

[30] In particular his middle third set.

Definition 8.1. The *cardinalities* of two sets A, B are denoted $|A|$ and $|B|$. We compare cardinalities as follows:

- $|A| \leq |B| \iff \exists f : A \to B$ injective.
- $|A| = |B| \iff \exists f : A \to B$ bijective.

We write $|A| < |B| \iff |A| \leq |B|$ and $|A| \neq |B|$. That is $\exists f : A \to B$ injective but $\nexists g : A \to B$ bijective.

Cardinality is defined as an abstract *property* whereby two sets can be *compared*. To define the cardinality $|A|$ as an object, we need the following theorem.

Theorem 8.2. *On any collection of sets, the relation $A \sim B \iff |A| = |B|$ is an equivalence relation.*

The cardinality of a set A is precisely the equivalence class of A with respect to this relation: $|A| := [A]$. It is now clear that cardinality partitions any collection of sets: every set has a cardinality, and no set has more than one cardinality. To get further it is useful to introduce a symbol for the cardinality of the simplest infinite set.

Countably Infinite Sets

Definition 8.3. The cardinality of the set of natural numbers $\mathbb{N}$ is denoted $\aleph_0$, read *aleph-nought* or *aleph-null*. We say that a set A is *countably infinite*, or *denumerable*[a] if $|A| = \aleph_0$.

[a] Sometimes this is shortened to *countable*, although some authors use countable to mean 'finite or denumerable,' i.e. any A for which $|A| \leq \aleph_0$. Use *countably infinite* or *denumerable* to avoid confusion. $\aleph$ is the first letter of the Hebrew alphabet.

We will discuss in a moment why we need a new symbol; why ∞ doesn't suffice. First we consider an example of Definition 8.1 at work.

Example. Let $2\mathbb{N} = \{2, 4, 6, 8, 10, \ldots\}$ be the set of positive even integers. The function

$$f : \mathbb{N} \to 2\mathbb{N} : n \mapsto 2n$$

is a bijection. It follows that $|2\mathbb{N}| = |\mathbb{N}| = \aleph_0$ and we would say that $2\mathbb{N}$ is denumerable.

This example shows one of the first strange properties of infinite sets: $2\mathbb{N}$ is a *proper subset* of $\mathbb{N}$, and yet the two sets are in bijective correspondence with one another! You should feel like you want to say two contradictory things simultaneously:

- $\mathbb{N}$ has the same 'number of elements' as $2\mathbb{N}$.
- $\mathbb{N}$ has twice the 'number of elements' as $2\mathbb{N}$.

If this doesn't make you feel uncomfortable, then read it again! The remedy to your discomfort is to appreciate that *cardinality* and *number of elements* are different concepts. Replacing 'number of elements' with 'cardinality' in the two statements makes both true! Indeed it is completely legitimate to write $2\aleph_0 = \aleph_0$.

Here is another example of the same phenomenum; $\mathbb{N}$ has one more element than $\mathbb{N}_{\geq 2}$ and yet they have the same cardinality: $\aleph_0 + 1 = \aleph_0$.

Example. The function $g : \mathbb{N} \to \mathbb{N}_{\geq 2} : n \mapsto n + 1$ is a bijection, whence $\mathbb{N}_{\geq 2} = \{2, 3, 4, 5, \ldots\}$ is denumerable.

As practice in using the definition of cardinality, we prove the following.

Theorem 8.4. *Suppose that A is a finite set. Then $|A| < \aleph_0$.*

Proof. The $n = 0$ case is left to the Exercises. Suppose that $|A| = n \geq 1$ so that we may list the elements of A as $\{a_1, \ldots, a_n\}$. We must prove two things:

1. $|A| \leq \aleph_0$. That is, $\exists f : A \to \mathbb{N}$ which is *injective.*
2. $|A| \neq \aleph_0$. That is, $\nexists g : A \to \mathbb{N}$ which is *bijective.* By symmetry this is equivalent to showing that there is no bijective function $h : \mathbb{N} \to A$.[a]

For part 1., simply define f by $f(a_k) = k$ for each $k \in \{1, 2, 3, \ldots, n\}$. This is injective since the distinct elements a_k of A map to distinct integers.
For part 2., suppose that $h : \mathbb{N} \to A$ is bijective. Consider the set

$$h\big(\{1, \ldots, n+1\}\big) = \big\{h(1), \ldots, h(n+1)\big\} \subseteq A.$$

Since A has n elements, by Dirichlet's box principle, at least two of the values $h(1), \ldots, h(n+1)$ must be equal. Therefore h is not injective and consequently not bijective. A contradiction. ∎

[a] If $g : A \to \mathbb{N}$ is a bijection, then $g^{-1} : \mathbb{N} \to A$ is also a bijection.

Aside: $\aleph_0$ versus ∞: what's the difference?

It can be difficult to grasp why $\aleph_0$ and ∞ are not the same thing. The problem is compounded by references to an 'infinite number' of objects any time that the cardinality of a set is not finite. This loose phrase is commonly used, but risks conflating the concepts of 'infinite set' and 'infinity.'
So what is the difference between $\aleph_0$ and ∞? If there aren't an 'infinite number' of natural numbers, how many are there? Theorem 8.4 says that $\aleph_0$ is 'larger than any natural number.' Is this not what we mean by infinity? The reason we need a new symbol $\aleph_0$, and why it and ∞ are different, is twofold:

1. As we shall see shortly, there are infinite sets with greater cardinality than $\aleph_0$: in a naïve sense, there are multiple infinities. The single symbol ∞ is insufficient to distinguish sets with different cardinalities.

2. More philosophically, $\aleph_0$ is an *object* in its own right; an object to which the cardinality of some set may be equal. Indeed, by Theorem 8.2, $\aleph_0$ is an equivalence class.

   By contrast, ∞ is not a object. Think back to where you've seen ∞ before. It is mostly used in *interval notation* (e.g., $[1, \infty)$) and when talking about *limits:* for example $\lim_{x\to 3} \frac{1}{(x-3)^2} = \infty$ is shorthand for the notion that the function $f(x) = \frac{1}{(x-3)^2}$ gets unboundedly larger as x approaches 3. The danger with this notation is that you mistakenly think of ∞ as a number: it isn't! An elementary calculus student might be tempted to write $f(3) = \frac{1}{(3-3)^2} = \infty$, but this makes absolutely no sense.

   Similarly, it is easy to mistake the appearance of ∞ in interval notation for a number: e.g. $(2, \infty)$ merely means 'all numbers greater than 2.' To say 'greater than 2 and *less than infinity*' would be an error.

The challenge of Cantor's notion of cardinality is to appreciate that the question, 'How many natural numbers are there?,' is meaningless!

We conclude this section with two important examples of denumerable sets.

Theorem 8.5. *The integers $\mathbb{Z}$ are denumerable.*

Proof. We must construct a bijective function $f : \mathbb{N} \to \mathbb{Z}$. By experimenting, you may feel it is enough simply to write down the first few terms of a suitable function:

n	1	2	3	4	5	6	7	8	9	10	$\cdots$
$f(n)$	0	1	-1	2	-2	3	-3	4	-4	5	$\cdots$

With a bit of thinking, it should be obvious what the function is doing, and that it is bijective. For a bit more formality, we can write

$$f(n) = \begin{cases} \frac{1}{2}n & \text{if } n \text{ even,} \\ -\frac{1}{2}(n-1) & \text{if } n \text{ odd.} \end{cases}$$

Now we check that this is bijective:

(*Injectivity*) Let $m, n \in \mathbb{N}$, and suppose that $f(m) = f(n)$. Without loss of generality, there are three cases to consider.

(m, n *both even*) $f(m) = f(n) \implies \frac{m}{2} = \frac{n}{2} \implies m = n.$

(m, n *both odd*) $f(m) = f(n) \implies -\frac{1}{2}(m-1) = -\frac{1}{2}(n-1) \implies m = n.$

(m *even,* n *odd*) $f(m) = f(n) \implies \frac{m}{2} = -\frac{1}{2}(n-1) \implies m + n = 1$. But $m, n \in \mathbb{N}$, so $m + n \geq 2$, which is a contradiction.

Therefore f is injective.

(*Surjectivity*) With a little calculation, you should be able to see that, for any $z \in \mathbb{Z}$, there exists a

positive integer n such that $f(n) = z$, namely:

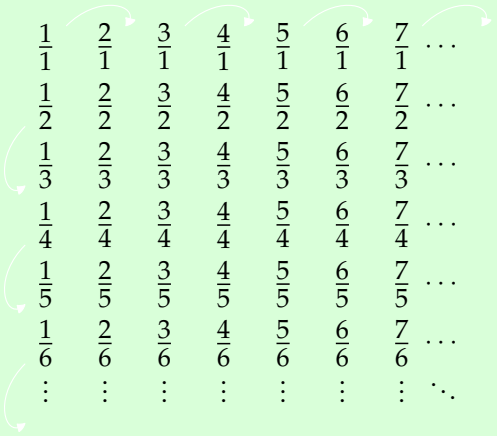
$$z = \begin{cases} f(2z) & \text{if } z > 0, \\ f(1-2z) & \text{if } z \leq 0. \end{cases}$$

Hence f is surjective. ∎

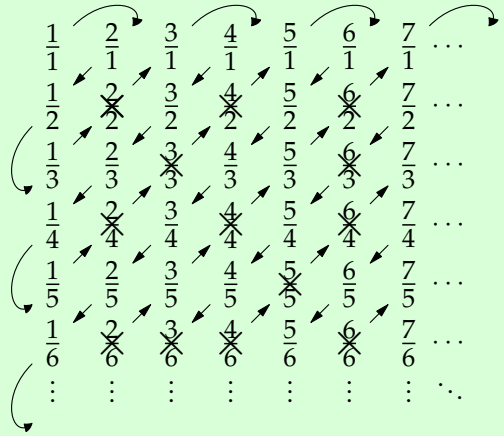
As you build up examples, you no longer have to compare denumerable sets directly with $\mathbb{N}$. A set A is denumerable if and only if $\exists f : A \to B$ bijective where B is *any other* denumerable set. This holds because the composition of bijective function is also bijective (Theorem 4.15).

Theorem 8.6. *The rational numbers $\mathbb{Q}$ are denumerable.*

Proof. We do this in stages. First we construct a bijection between the positive rational numbers $\mathbb{Q}^+$ and the natural numbers $\mathbb{N}$.
For each $a, b, \in \mathbb{N}$, place the fraction $\frac{a}{b}$ in the ath row and bth column of the infinite square as shown below. Now list the elements by tracing the diagonals as shown, deleting any number that has already appeared in the list ($\frac{2}{2} = \frac{1}{1}$, $\frac{6}{4} = \frac{3}{2}$, etc.).



The infinite square



Trace diagonals and delete repeats

We obtain the *ordered set*

$$A = \{a_1, a_2, a_3, a_4, \ldots\} = \left\{ \frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{1}{3}, \frac{3}{1}, \frac{4}{1}, \frac{3}{2}, \frac{2}{3}, \frac{1}{4}, \frac{1}{5}, \cdots \right\}.$$

Now define the function $f : \mathbb{N} \to \mathbb{Q}^+$ by $f(n) = a_n$. We claim that this is a bijection.

(*Injectivity*) Let $m = n \in \mathbb{N}$, and suppose that $f(n) = f(m)$. Then $a_m = a_n$. But in the construction of A we deleted any number which had already appeared in the list. Thus a_m can only equal a_n if $m = n$.

(*Surjectivity*) A positive rational number $\frac{a}{b}$ appears in the ath row and bth column of the square (and in many other places). When constructing A, note that $\frac{a}{b}$ will not be deleted unless it has *already*

appeared elsewhere in A. Therefore every positive fraction $\frac{a}{b}$ is in the set A.

To finish things off, extend the function to all rational numbers by

$$g : \mathbb{Z} \to \mathbb{Q} : n \mapsto \begin{cases} f(n) & \text{if } n > 0, \\ 0 & \text{if } n = 0, \\ -f(-n) & \text{if } n < 0. \end{cases}$$

Now $g : \mathbb{Z} \to \mathbb{Q}$ is a bijection, from which we deduce that $|\mathbb{Q}| = |\mathbb{Z}| = \aleph_0$. ■

This result should surprise you! Any sensible person should feel that there are far, far more rational numbers than integers and yet the two sets have the same cardinality. Bizarre.

There are other denumerable sets that appear to be even larger. For example, we can show that $\mathbb{N} \times \mathbb{N}$ is denumerable (using almost the same proof as for $\mathbb{Q}^+$ except that there are no repeats to delete). For a much larger-seeming denumerable set, consider the set of *algebraic numbers:*

$$\{x \in \mathbb{R} : \exists \text{ a polynomial } p \text{ with integer coefficients such that } p(x) = 0\}.$$

Algebraic numbers are the zeros of polynomials with integer coefficients. Clearly every rational number $\frac{a}{b}$ is algebraic, since it satisfies $p(x) = 0$ for $p(x) = bx - a$. There are many more algebraic numbers than rational numbers: e.g. $\sqrt[5]{2} - 3$ is algebraic since it is a root of the polynomial $p(x) = (x+3)^5 - 2 = 0$. Not all real numbers are algebraic however: those which aren't, such as π and e, are termed *transcendental.*

Exercises

8.1.1 Refresh your proof skills by proving that the following functions are bijections:

(a) $f : \mathbb{N} \to 2\mathbb{N} : n \mapsto 2n$.

(b) $g : \mathbb{N} \to \mathbb{N}_{\geq 2} : n \mapsto n + 1$.

8.1.2 Construct a function $f : \mathbb{N} \to \mathbb{Z}_{\geq -3} = \{-3, -2, -1, 0, 1, 2, 3, 4, \ldots\}$ which proves that the latter set is denumerable: you must show that your function is a bijection.

8.1.3 Prove that the set $3\mathbb{Z} + 2 = \{3n + 2 : n \in \mathbb{Z}\}$ is denumerable.

8.1.4 Show that the set of all triples of the form $(n^2, 5, n + 2)$ with $n \in 3\mathbb{Z}$ is denumerable by explicitly providing a bijection with a denumerable set A. (*You must check that the set A is denumerable, and that your map is indeed a bijection.*)

8.1.5 Imagine a hotel with an infinite number of rooms: Room 1, Room 2, Room 3, Room 4, etc.. Show that, even if the hotel is full, the guests may be re-accommodated so that there is always a room free for one additional guest.
Hint: consider the function $f : \mathbb{N} \to \mathbb{N} : n \mapsto n + 1$.

8.1.6 Prove that $A \subseteq B \implies |A| \leq |B|$. (*You need an injective function $f : A \to B$*)

8.1.7 Prove Theorem 8.2. (*You need little more than Theorem 4.15 on the composition of bijective functions.*)

8.1.8 Prove that the set $\mathbb{N} \times \mathbb{N}$ is denumerable. You should base your proof on Theorem 8.6.

8.1.9 We know that $\mathbb{Q}$ is denumerable, and we saw (Theorem 8.6) that there most exist a bijective function $f : \mathbb{N} \to \mathbb{Q}$. Show that $g : \mathbb{N} \times \mathbb{N} \to \mathbb{Q} \times \mathbb{Q}$ defined by $g(m, n) = (f(m), f(n))$ is a bijection. Appeal to the previous question to show that $\mathbb{Q} \times \mathbb{Q}$ is denumerable.

8.1.10 Here we consider the $n = 0$ case of Theorem 8.4. Recall the definition of function in Section 7.2.

(a) If $|A| = 0$, then $A = \emptyset$. Suppose that $f : \emptyset \to \mathbb{N}$ is a function. Use Definition 7.4 to prove that $f = \emptyset$.

(b) State what it means, in the language of Definition 7.4, for a function $f : A \to \mathbb{N}$ to be injective. Show that $f = \emptyset$ is an injective function.

(c) Suppose that B is a set with $|B| \geq 1$. Prove by contradiction that there are no functions $h : B \to \emptyset$. Conclude that $0 < \aleph_0$.

8.1.11 Suppose that the set A_n is denumerable for each $n \in \mathbb{N}$. We may then list the elements of each set: $A_n = \{a_{n1}, a_{n2}, a_{n3}, a_{n4}, \ldots\}$. Now list the elements of the sets $A_1, A_2, A_3, \ldots$ as follows:

$$\begin{aligned} A_1 &= \{a_{11}, a_{12}, a_{13}, a_{14}, \ldots\} \\ A_2 &= \{a_{21}, a_{22}, a_{23}, a_{24}, \ldots\} \\ A_3 &= \{a_{31}, a_{32}, a_{33}, a_{34}, \ldots\} \\ &\vdots \end{aligned}$$

Use this construction to prove that $\bigcup_{n \in \mathbb{N}} A_n$ is a denumerable set.
This result is often stated, 'A countable union of countable sets is countable.'

8.1.12 (Hard!) In this question we prove the converse of Theorem 8.4: if $|A| < \aleph_0$, then A is a finite set. Otherwise said, $\aleph_0$ is the smallest infinite cardinal.
We prove by contradiction. Suppose that A is an infinite set such that $|A| < \aleph_0$. Then there exists an injective function $f : A \to \mathbb{N}$. List the elements of the image of f in increasing order:

$$\operatorname{Im} f = \{n_1, n_2, n_3, \ldots\}.$$

(a) Prove that $\operatorname{Im} f$ is an infinite set.

(b) Show that for all $k \in \mathbb{N}$, there exists a unique $a_k \in A$ satisfying $f(a_k) = n_k$.

(c) Define $g : \mathbb{N} \to A$ by $g(k) = a_k$. Prove that g is a bijection.

(d) Why do we obtain a contradiction?

8.2 Uncountable Sets

You might think, since $\mathbb{Q}$ seems so large, that there can't be any sets with strictly larger cardinality. But we haven't yet thought about the set of real numbers.

Definition 8.7. A set A is *uncountable* if $|A| > \aleph_0$, that is if there exists an injection $f : \mathbb{N} \to A$ but no bijection $g : \mathbb{N} \to A$.

Theorem 8.8. *The interval* $[0, 1]$ *of real numbers is uncountable.*

We denote the cardinality of the interval $[0, 1]$ by the symbol $\mathfrak{c}$ for *continuum.* The theorem may therefore be written $\mathfrak{c} > \aleph_0$.

Proof. First we require an injective function $f : \mathbb{N} \to [0, 1]$. The function defined by $f(n) = \frac{1}{n}$ clearly fits the bill, for

$$f(n) = f(m) \implies \frac{1}{n} = \frac{1}{m} \implies n = m.$$

Therefore $\aleph_0 \leq \mathfrak{c}$.
Next, we prove that there exists no bijection from $\mathbb{N}$ to $[0, 1]$, arguing by contradiction. Suppose that $g : \mathbb{N} \to [0, 1]$ is a bijection and consider the sequence of values $g(1), g(2), g(3), \ldots$ These are real numbers between 0 and 1, hence they may all be expressed as decimals of the form $0.a_1a_2a_3a_4a_5 \cdots$, where each $a_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.[a] We can write:

$$\begin{aligned}
g(1) &= 0.b_{11}b_{12}b_{13}b_{14}b_{15}b_{16} \cdots \\
g(2) &= 0.b_{21}b_{22}b_{23}b_{24}b_{25}b_{26} \cdots \\
g(3) &= 0.b_{31}b_{32}b_{33}b_{34}b_{35}b_{36} \cdots \\
g(4) &= 0.b_{41}b_{42}b_{43}b_{44}b_{45}b_{46} \cdots \\
g(5) &= 0.b_{51}b_{52}b_{53}b_{54}b_{55}b_{56} \cdots \\
&\vdots
\end{aligned}$$

By assumption, g is bijective, so it is certainly surjective. It follows that all of the numbers in $[0, 1]$ appear in the above list of decimals. Since g is injective, there are no repeats in the list. Now define a new decimal

$$c = 0.c_1c_2c_3c_4c_5 \cdots \qquad \text{where} \quad c_n = \begin{cases} 1 & \text{if } b_{nn} \neq 1, \\ 2 & \text{if } b_{nn} = 1. \end{cases}$$

c is a non-terminating decimal whose digits are only 1's and 2's: it therefore has no other decimal representation. Since c disagrees with $g(n)$ at the nth decimal place, we have $c \neq g(n)$, $\forall n \geq 1$. Hence c is *not* in the above list. However $c \in [0, 1]$ and g is surjective with $\operatorname{Im} g = [0, 1]$, so we have a contradiction. We conclude that $\mathfrak{c} \neq \aleph_0$.

Putting this together with the first part of the proof, we see that $\mathfrak{c} > \aleph_0$. ∎

[a] Certain numbers, like $0.\overline{12} = 0.12121212\cdots$ have a unique decimal representation. Others, like $0.317 = 0.3169999\cdots$ have both a finite decimal representation and an infinite representation that ultimately becomes an infinite sequence of 9's. For the purposes of this proof it does not matter which representation is chosen when there is a choice. We are forced, however, to take $1 = 0.999999\cdots$, due to our insistence that all elements are written with zero units.

The interval $[0,1]$ has a strictly larger cardinality than the set of integers. Since $[0,1] \subseteq \mathbb{R}$, it follows immediately that the real numbers are also uncountable. Indeed we shall see in a moment that the real numbers have cardinality $\mathfrak{c}$, as does any interval (of positive width). More amazingly, the Cantor middle-third set (page 111) also has cardinality $\mathfrak{c}$, despite seeming vashishingly small.

More advanced ideas

Our countable and uncountable examples are merely scratching the foothills of a truly weird subject. Here are a couple more ideas.

The following theorem is very useful for being able to compare cardinalities. It allows us to prove that two sets have the same cardinality *without* explicitly constructing bijective functions. Injective functions are usually much easier to build.

Theorem 8.9 (Cantor–Schröder–Bernstein)**.** *If* $|A| \leq |B|$ *and* $|B| \leq |A|$*, then* $|A| = |B|$.

The theorem seems like it should be obvious, but pause for a moment: it is *not* a result about *numbers! A* and *B* are *sets,* and so the theorem must be understood in the context of Definition 8.1. In this language the theorem becomes:

Suppose that there exist *injective* functions $f : A \to B$ and $g : B \to A$.
Then there exists a *bijective* function $h : A \to B$.

The proof is beautiful, though a little long to reproduce here. If you are interested it can be found in any text on set theory. The applications of the theorem are more important to our purposes.

Theorem 8.10. *The interval* $(0,1)$ *has cardinality* $\mathfrak{c}$.

It is possible to define a bijection $h : (0,1) \to [0,1]$, though it is extremely messy. Instead we construct two injections.

Proof. $f : (0,1) \to [0,1] : x \mapsto x$ is clearly an injection, whence $|(0,1)| \leq |[0,1]| = \mathfrak{c}$. Now define

$$g : [0,1] \to (0,1) : x \mapsto \frac{1}{2}x + \frac{1}{4}.$$

g is certainly injective (g isn't surjective, since $\mathrm{Im}(g) = [\frac{1}{4}, \frac{3}{4}] \neq (0,1)$), and so $\mathfrak{c} \leq |(0,1)|$.
By the Cantor–Schröder–Bernstein Theorem, the sets $(0,1)$ and $[0,1]$ have the same cardinality $\mathfrak{c}$. ∎

By a similar trick, covered in the Exercises, one can see that $\mathbb{R}$ also has cardinality $\mathfrak{c}$.

For a final idea, we prove Cantor's Theorem, which says that the power set of a set always has a strictly larger cardinality than the original set. In Theorem 6.6 we saw that, if A is finite, then $|\mathcal{P}(A)| = 2^{|A|}$ for finite sets. We therefore already believe that Cantor's Theorem is true for finite sets. The proof we shall give also works for infinite sets.

The main implication of this is that *there is no largest set!* We can always make a larger set simply by taking the power set of what we already have: now rinse and repeat! For example, $\mathcal{P}(\mathbb{R})$ has larger cardinality than $\mathbb{R}$. If you want a set with larger cardinality, why not take $\mathcal{P}(\mathcal{P}(\mathbb{R}))$? Or $\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{R})))$. There is no limit to the cardinality of sets.

Theorem 8.11 (Cantor)**.** *If A is any set, then $|A| \lneq |\mathcal{P}(A)|$.*

Proof. We must show two things:

- $\exists f : A \to \mathcal{P}(A)$ which is injective.
- $\nexists g : A \to \mathcal{P}(A)$ which is bijective.

For the first, note that $f : a \mapsto \{a\}$ is a suitable injective function.[a]
Now suppose for a contradiction that $\exists g : A \to \mathcal{P}(A)$ which is bijective. For every $a \in A$, $g(a)$ is a subset of A. Consider the set

$$X = \{a \in A : a \notin g(a)\}.$$

[a] This even works if $A = \emptyset$, for then f is itself the 'empty' function! If this sort of thinking disturbs you, don't worry. We have already proved Cantor's Theorem for all *finite* sets, so we only need the proof to work for infinite sets.

This is a difficult set to think about. Before proceeding, let us consider an example. Suppose that $g : \{1,2\} \to \mathcal{P}(\{1,2\})$ is defined by

$$g(1) = \{1,2\}, \qquad g(2) = \{1\}.$$

Then $1 \in g(1)$ and $2 \notin g(2)$, whence the above set is $X = \{2\}$. Since we are trying to show that a bijection g as in the proof does not exist, it is important to note that the function g in our example is *not bijective!*

Proof Continued. By assumption, g is bijective, hence it is certainly surjective. Because $\operatorname{Im} g = \mathcal{P}(A)$, the set X *is in the image of* g. Otherwise said, there exists $\hat{a} \in A$ such that $g(\hat{a}) = X$. We ask whether $\hat{a}$ is an element of X. Think carefully about the definition of X, and observe that

$$\begin{aligned} \hat{a} \in X &\iff \hat{a} \notin g(\hat{a}) && \text{(by the definition of } X\text{)} \\ &\iff \hat{a} \notin X && \text{(since } X = g(\hat{a})\text{)} \end{aligned}$$

Look at what we have: $\hat{a} \in X \iff \hat{a} \notin X$. This is clearly 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 large part in pushing set theory towards axiomatization. Here is a conundrum motivated by the theorem: If a 'set' is just a collection of objects, then we may consider the 'set of all sets.' Call this A. Now consider the power set of A. Since $\mathcal{P}(A)$ is a set of sets, it must be a subset of A, whence $|\mathcal{P}(A)| \leq |A|$. However, by Cantor's Theorem, we have $|A| \lneq |\mathcal{P}(A))|$. The conclusion is the palpable contradiction

$$|\mathcal{P}(A)| \lneq |\mathcal{P}(A)|!$$

The remedy is a thorough definition of 'set' which prevents the collection of all sets from being a set. This is where *axiomatic set theory,* and a completely new approach, begins.

Exercises

8.2.1 You may assume that $[0,1]$ has cardinality $\mathfrak{c}$.

(a) Construct an explicit bijection $f : [0,1] \to [3,8]$ which proves that the interval $[3,8]$ also has cardinality $\mathfrak{c}$. *Try a linear function mapping the endpoints of* $[0,1]$ *to the endpoints of* $[3,8]$.

(b) Let $a, b \in \mathbb{R}$ with $a < b$. Generalizing the previous example, construct a bijection which proves that the closed interval $[a,b]$ has cardinality $\mathfrak{c}$.

8.2.2 (a) Suppose that $g : \{1,2,3,4\} \to \mathcal{P}(\{1,2,3,4\})$ is defined by

$$g(1) = \{1,2,3\}, \qquad g(2) = \{1,4\}, \qquad g(3) = \emptyset, \qquad g(4) = \{2,4\}.$$

Compute the set $X = \big\{a \in \{1,2,3,4\} : a \notin g(a)\big\}$.

(b) Repeat part (a) for $g : \mathbb{N} \to \mathcal{P}(\mathbb{N}) : n \mapsto \{x \in 2\mathbb{N} : x \leq n\}$.

8.2.3 The proof of Cantor's Theorem makes use of a construction similar to *Russell's Paradox.* Let X be the set of all sets which are not members of themselves: explicitly

$$X = \{A : A \notin A\}.$$

(a) Assume that X is a set, and use it to deduce a contradiction: ask yourself if X is a member of itself.

(b) Russell's paradox (and indeed the proof of Cantor's Theorem) is one avatar of an ancient logical paradox which appears in many guises. For example, suppose that a town has one hairdresser, and suppose that the hairdresser is the person who cuts the hair of all the people, and only those people, who do not cut their own hair. Who cuts the hairdresser's hair? Can you explain the connection with Russell's paradox/Cantor's Theorem?

The point of Russell's paradox is that we need a definition of 'set' which prevents objects like X from being sets.

8.2.4 Recall the Cantor set as described in the notes, where we proved that $\mathcal{C}$ is the set of all numbers in $[0,1]$ possessing a ternary expansion consisting only of zeros and twos. Modeling your answer on the proof that the interval $[0,1]$ is uncountable, prove that $\mathcal{C}$ is uncountable.

8.2.5 (a) Show that $|(0,1)| \leq |\mathbb{R} \setminus \mathbb{N}| \leq |\mathbb{R}|$.

(b) Construct a bijection $f : (0,1) \to (-\frac{\pi}{2}, \frac{\pi}{2})$. (*Try a linear function*)

(c) Show that $g : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R} : x \mapsto \tan x$ is a bijection.

(d) Use the Cantor–Schröder–Bernstein Theorem to conclude that $|\mathbb{R} \setminus \mathbb{N}| = |\mathbb{R}| = \mathfrak{c}$.