

Explorations in Elementary Mathematics

Explorations in Elementary Mathematics

Julia Kameryn Williams
Cape Cod Community College

June 16, 2026

Website: juliakw.net/books/eem/

©2026–2026 Julia Kameryn Williams

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit CreativeCommons.org

Contents

Preface: To the student	1
0 Critical Thinking	2
0.1 Reading mathematics	2
0.2 Problem solving	5
0.3 Exercises	8
1 The Natural Numbers	9
1.1 What even is counting?.	9
1.2 Sets.	10
1.3 The meaning of the natural numbers.	11
1.4 Operations on sets	12
1.5 Number classes.	14
1.6 The arithmetic of natural numbers	14
1.7 Division	17
1.8 Order of Operations	18
1.9 Exercises	19
2 Numeration	22
2.1 Roman Numerals	22
2.2 The Hindu–Arabic numeral system	23
2.3 Algorithms for adding, subtracting, and multiplying	25
2.4 Other bases	31
2.5 Exercises	33
3 Number Theory	35
3.1 Prime numbers.	35
3.2 Divisibility Rules	37
3.3 GCD and LCM	39
3.4 Exercises	42
4 The Integers	44
4.1 Some motivation	44
4.2 Defining the integers.	45
4.3 The geometric meaning of integers.	46

4.4	Arithmetic with integers	48
4.5	Clocks and modular arithmetic	49
4.6	Exercises	51
5	The Rational Numbers	53
5.1	Motivating the rational numbers	53
5.2	The definition of the rational numbers	53
5.3	Arithmetic of fractions	55
5.4	Percents	59
5.5	Exercises	60
6	The Real Numbers	62
6.1	Writing fractions in decimal	62
6.2	The arithmetic of Hindu–Arabic numerals	64
6.3	Approximations	66
6.4	Defining the real numbers	70
6.5	Characterizing rational and irrational numbers	74
6.6	Exercises	75

Back Matter

Preface: To the student

This book is designed for students pursuing a career in early childhood or elementary education. Its aim is for you to understand the mathematics you will teach to your students.

When teaching, you should understand the topic more deeply than your students. This is for many reasons. One is that later learning builds on earlier education, and you need to know where students are going to better prepare them. This is especially the case in mathematics education, due to the cumulative nature of the discipline. Another reason is that you will occasionally get advanced students who quickly pick up the material and want more. You want to know enough to be able to guide them to further self-study. Yet another reason is that methods and fads in pedagogy change over time. A deeper understanding of the material enables you to better adapt to these changes and see how they are new looks at the same concepts.

For these reasons, this book is not organized the same as the elementary mathematics curriculum. Instead, chapters are broken up by topic, with each chapter covering material that is spread across multiple years in an elementary school. The goal is for you to understand the unifying mathematical concepts and see how topics across multiple years are aspects of a larger pattern.

[Chapter 0](#) is an introduction covering how to read mathematics and how to think about problem solving. The next chapters are organized around different number systems. [Chapter 1](#) introduces the **natural numbers**, those numbers used for counting. [Chapter 2](#) then covers how to write natural numbers and how to do arithmetical calculations with them, and [Chapter 3](#) goes into a different perspective on the natural numbers. In [Chapter 4](#) expand the natural numbers to build the **integers**, then the **rational numbers** in [Chapter 5](#) and the **real numbers** in [Chapter 6](#).

Chapter 0

Critical Thinking

Mathematics is the science of patterns. Special attention is paid to patterns involving quantities and shapes. Part of the job of an elementary educator is to introduce young learners to this discipline. This is a process that takes many years; patterns build upon patterns up to as high a level of abstraction as you can take.

It is important to emphasize, however, that this is grounded in concrete problem solving. There are some deviants who love the abstraction for its own sake—we call such people mathematicians. But the reason every culture has developed mathematics and the reason every schoolchild is required to learn it is because understanding these patterns is useful for practical reasoning about the world.

In this chapter we talk about how to apply mathematical reasoning to solve real world problems. But first we take a detour to discuss how to read mathematics, including this very textbook.

0.1 Reading mathematics

Mathematics is a specialized genre. Reading a mathematics textbook like this one as such requires knowing some conventions for how mathematicians write. The key thing underlying all these conventions is, mathematics is a discipline that requires precision.

Probably the best known quirk of mathematicians is the tendency to use letters to refer to numbers. While this quirk is a hurdle for many when they first see it, its purpose is simple. It's convenient to have names to refer to things. You might, as with the word problems in [Section 0.2](#), want to determine the value of some quantity. To do this, you can write an equation which describes the quantity and use rules of algebra to solve for it. For this repeated symbol pushing it's nice to have a short name. You could call it many different things—Fred, question mark, smiley face, whatever—but convention in mathematics is to use a single letter.

Indeed, it can be useful to use names even if you know the value of a quantity. For example, we know the circle constant is approximately 3.14159, and rather than writing all that out it's easier to just call it π .

Mathematicians call these names **variables**. While numbers are where students first see variables, they're also used for any kind of mathematical object—from shapes and sets to complex manifolds. There's a few different phrases that are used to give names to mathematical objects. What they all have in common is, they have to say what the variable is, that is what letter is

being used, and they have to say what object is being named. Here are a few examples.

- “Let x be...” This can be quite specific, like “let x be the height of the flag pole” or it could say as little as “let x be a number”, depending on what’s appropriate for the context.
- “Consider a positive distance d .” Here you’re told the name (d), a mathematical fact about it (it’s positive, > 0), and what it represents (a distance), all in a single compact phrase.
- “Call t the time when the projectile reaches its highest point.” This example doesn’t explicitly say what type of mathematical object t is, but you can infer that it’s a real number (i.e. a quantity you can express in decimal) based on what it’s representing.

Another quirk of mathematical writing comes in its use of logical language. Some of these should be highlighted. The first is that “or” always allows the possibility of both options. A sentence like “ n is a multiple of 2 or a multiple of 3” could mean that n is a multiple of both 2 and 3. (For example, $n = 6$.)

Don’t go out to lunch with a mathematician. The waiter will ask “Would you like the soup or salad with that?” and she will think it’s funny to answer “Yes”.

Mathematics is often interested in logical consequences; if I know this then I can also conclude that. There are multiple ways “if X is true then Y is true” gets phrased. Here are a few.

- “ X implies Y .”
- “Whenever X , also Y .”
- “ Y only if X .”
- “ X is a sufficient condition for Y .”
- “ Y is a necessary condition for X .”

Example 0.1.1 “If two triangles have the same angles then they can be rescaled to have the same side lengths.” The way you can use a fact like this is, if you know an instance of the “if” part—if you have two triangles and their angles are the same—then you get to for free also know the “then” part—you know one can be rescaled to be the same size as the other. For example, this means that the ratio of the side lengths, such as the ratio of the side adjacent to an angle A and the hypotenuse, are the same across both triangles. This if-then statement is one of the key facts underlying the area of mathematics known as **trigonometry**. ▲

Also of interest is when two things are equivalent. This can be expressed as “ X if and only if Y ” or, if you’re stuffy, as “ X is a necessary and sufficient condition for Y ”. Either way, this amounts to asserting both if X then Y and if Y then X .

Example 0.1.2 “A triangle has all three angles the same if and only if all its sides are the same length.” This fact tells you these two statements are equivalent. If you know one, you automatically know the other. ▲

Ordinary nonfiction writing is broken up into paragraphs, with information being unfolded across them. Mathematicians break from this convention by pulling out certain statements from the rest of the text. One place this is done is for important definitions.

Definition 0.1.3 Prime numbers. A positive whole number p is **prime** if it has exactly two factors, itself and 1. ◆

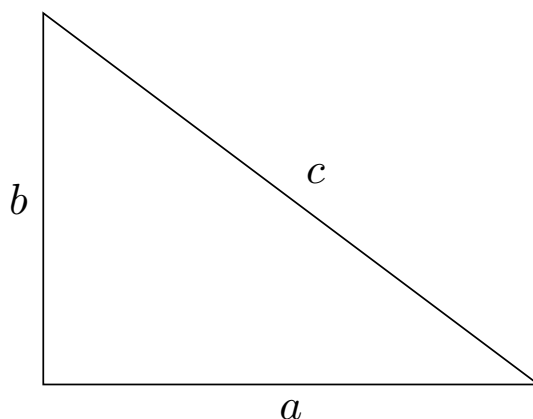
The italicized term, in this case **prime**, is what is being defined. The rest of the paragraph (just a single sentence in this case, though more complicated definitions can take many sentences) describes what objects the definition applies to and how it is defined. In this case, it only makes sense to ask about positive whole numbers being prime; so you wouldn't ask whether, say, $\frac{1}{2}$ or π is prime.

Sometimes definitions don't appear in a special block. For instance, I did this above with **variable**. The same conventions apply, but when reading you should take note of anything the author considers important enough to emphasize by separating it from the main text.

The other main thing you see pulled from the main text are theorems. A **theorem** is a mathematical fact, usually one that we think is cool and useful and informative.

Theorem 0.1.4 The Pythagorean theorem. *If a and b are the lengths of the legs of a right triangle and c is the length of its hypotenuse, then*

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



Mathematicians aren't just interested in what's true, but in why it's true. We call an explanation of why a theorem is true a **proof**. In this book, proofs are mostly omitted, as they are too long or too difficult. I do include a few, however, where they are not too bad. You should read these. We tell students to show their work on math problems because the process is more important than the final result. The same is true with theorems; understanding why something is true is more important than just the theorem itself. If you don't fully understand a proof on one read, that is normal. Come back to it later or ask about it at office hours. No one perfectly understands mathematics immediately.

Theorem 0.1.5 *If a number is even so are any of its multiples.*

Proof. Let n be an even number and consider one of its multiples, call it an . Being even means being a multiple of 2, and so there is some number k so that $n = 2k$. Substituting this, we get

$$an = a \cdot 2k = 2 \cdot ak.$$

Thus we see that an is a multiple of 2. This shows it is even. ■

Note that this theorem takes the form of an if-then statement. Many theorems take this form. The explanation for why it's true follows from this form.

“If X then Y ” means that if we know X then we can conclude Y . So to show why it’s true, we assume that X is true and we work to show that Y must also be true.

Mathematics books will include material they ask the reader to explain. While it may seem this comes from sadism, it’s actually because you learn mathematics by doing it. Passively reading or watching lecture is not enough. In this book, these take two forms. One is exercises collected at the end of each chapter. These are problems to practice your understanding of the material. The other are checkpoints inserted at various points in the text. These are smaller exercises to do as part of your reading; as the name says, they serve as checkpoints to test that you understand what is going on.

Checkpoint 0.1.6 Try out the definition-theorem style of writing mathematics yourself. Write out a definition of a mathematical concept and a mathematical fact using the style you’ve seen above. If you’re feeling extra adventurous, write a proof for your theorem.

0.2 Problem solving

I promised that mathematics is useful for real world problems. Instead of giving examples of this, we will talk about contrived fake problems commonly called **word problems**. Let me give an example to demonstrate why we use word problems instead of real problems.

Checkpoint 0.2.1 You are trying to determine your budget for the month. Your only expenses are rent and food. If you have 2000 dollars and your share of rent is 1400 dollars, how much do you have left over for food?

Solution. Give a name to the unknown quantity, namely how much money you have for food. The name doesn’t really matter, but I will pick F —for “food”. Now write an equation that describes how the unknown quantity relates to the known quantities.

$$1400 + F = 2000$$

Next use algebra to solve this equation for F . In this case, solving the equation is the single step of moving the 1400 to the other side by subtracting it from both sides.

$$F = 2000 - 1400 = 600$$

This is the answer to the question. You have 600 dollars to spend on food this month.

In the real world, household budgeting is a task many have to do. It is much more complicated, however. You have more than two expenses to worry about—energy bill, cell phone bill, internet bill, car expenses, school supplies, child care, entertainment to distract you from your economic precarity, and so on. The numbers aren’t going to be nice, round multiples of 100, and they’re going to change from month to month. Mathematics can still help you to determine your budget. But the simplified problem is clearer for demonstrating problem solving techniques. If you understand it, you can apply that same understanding to the more complicated problem.

If a future student ever asks you why they have to solve problems about someone buying forty watermelons, maybe something like the above will be helpful to them. Or maybe an analogy is more their style. A hockey coach has her players drill by skating in between cones while carrying the puck. She doesn’t do this because that’s what a match of hockey consists of. Rather, it is a drill that trains the skills used in a real match.

The above example also illustrates a basic principle of mathematical problem solving: identify the known and unknown quantities, and determine how they relate. Usually this takes the form of writing an equation expressing the relationship between the quantities. This transfers your problem into the mathematical realm, where you can use patterns to extract more information. In the previous case, those patterns took the form of algebra rules for manipulating equations.

Try a more complicated problem. Attempt it yourself before you check versus my solution.

Checkpoint 0.2.2 You are packing a suitcase for a vacation, and you want to take as many paperback books as possible. The airline enforces a 50 pound limit for checked baggage, and your suitcase weighs 4 pounds. If each book weighs 0.3 pounds, what is the largest number of books you can fit?

Solution. Call by B the number of books you take. The total weight is then $4 + 0.3B$, and we want to know how big B can be and this still be ≤ 50 . We will figure this out by solving $4 + 0.3B = 50$, rounding down if necessary. (You can't pack a third a book!) Solving this equation takes a couple steps.

$$\begin{aligned} 0.3B &= 50 - 4 = 46 \\ B &= \frac{46}{0.3} \approx 153.33 \end{aligned}$$

We round down and conclude you can pack 153 books. Probably you will take fewer—some must be sacrificed for clothes, etc.—but that is the maximum before going over weight.

Many applications of mathematics require specialized formulas, especially in the sciences and engineering. That is, the process requires a lot of work to determine which mathematical concept gives a good model of a real world phenomenon. While this kind of application is the most common and most important use of mathematics, it makes for bad examples. Either a formula must be given without explanation, or a long detour must be taken to explain another subject.

Let's see an example to illustrate this.

Checkpoint 0.2.3 A ball is launched upward from the ground with an initial speed of 30 meters per second. Determine its height after 2 seconds.

Hint. If an object is launched straight upward from the ground with an initial speed of v , then its height y (in meters) after t seconds is given by the formula

$$y = vt - 10t^2.$$

Solution. Use the formula from the hint, plugging in 2 for t and 30 for v .

$$y = 30 \cdot 2 - 10 \cdot 2^2 = 60 - 40 = 20$$

We conclude the object is 20 meters above the ground.

This exercise was probably unsatisfying. Without the hint, you had no way to proceed. And the formula in the hint wasn't explained in any way. Where does that come from?

It gets even worse. This formula is inaccurate. Even if we had gone through a detour of calculus and physics to explain the formula, we'd still get a bad answer. The culprit is friction. This formula works if there is negligible air resistance, but that probably isn't the case. (You can try this yourself. Throw a ball and it makes a nice arc, but toss a piece of paper and it floats about.)

If we want a more accurate formula we'd have to take an even longer, more difficult detour to arrive at an even more complicated formula.

If you are an engineer, then you want to take these extra detours. You need more accurate calculations. But if you just want to see an example of how to apply math, they're no good. So instead we go with contrived, simplistic examples.

On that note, let's see a few more contrived problems.

Checkpoint 0.2.4 You have two rods. You measure that the longer one is 6 inches longer and when you lay them end-to-end the total length is 32 inches. How long are the rods?

Solution. Give names to the lengths. Call the longer length x and the shorter length y , both measured in inches. We know that

$$\begin{aligned}x - y &= 6 \\x + y &= 32.\end{aligned}$$

Adding the two equations together, we get $2x = 38$, and so $x = 19$. Plugging this into the second equation we get $19 + y = 32$ and so $y = 13$.

An alternative way you might have solved this is, the longer rod being 6 inches longer can be written in an equation as $x = y + 6$. This can be substituted into the second equation to get $2y + 6 = 32$. From this you can solve for y , and then use that to find x .

Another way you might have gone about solving this is **guessing and checking**. Knowing that the longer rod is 6 inches longer, you might have guessed they are 12 and 18 inches long. This makes the total length 30 inches, which is too short. So you revise your guess to get the extra two inches by upping them each by one inch, arriving at the correct answer of 13 and 19.

Guessing and checking can be a slow process, since it may take many rounds of guessing to settle on the correct answer. The virtue of this method is that it works even if you don't see how to write an equation (or, in this case, two equations) which describe the unknown quantities. It's also useful if you have an equation you don't know how to solve. As such it can be a useful problem solving technique to keep in your toolbox.

Checkpoint 0.2.5 A rectangular flower bed has an area of 12 square feet, and the longer side is 4 feet longer than the shorter side. What are the dimensions of the flower bed?

Solution. Let x denote the length of the shorter side. Then the sides have lengths x and $x + 4$, so the area is expressed by the equation

$$x(x + 4) = 12.$$

This is a **quadratic equation**, which can be solved by factoring or the quadratic formula. If you don't remember either of those methods, you can use guessing and checking. After a few tries you should settle on $x = 2$, giving you side lengths of 2 and 6 as being the guess that works.

0.3 Exercises

Exercises

1. A college drama club is selling tickets to a production they are putting on. Early bird tickets are sold for 8 dollars each, while tickets at the door cost 10 dollars. If the club sold 40 early bird tickets, how many tickets at the door do they need to sell to make their target revenue of 600 dollars? What if they want to make 800 dollars? (Round your answer to the nearest whole number.)
2. You are designing a conical pour-over coffee brewer. If the radius of the base of the cone must be 3 centimeters and you want the cone to be able hold 200 milliliters, how tall must the cone be? Round up to the nearest centimeter.

Hint. One milliliter is the same as one cubic centimeter. And the formula for the volume of a cone is

$$V = \frac{1}{3}\pi r^2 h,$$

where V is the volume, r is the radius of the base, and h is the height. Of course π is the familiar circle constant $\pi \approx 3.14159$.

3. Use guessing and checking to find the whole numbers closest to the solutions to these equations.

$$2^x = x + 4, \quad 3^{x-2} = x + 5, \quad \frac{6}{x} = x^2 - 2x$$

4. Write your own word problem about arithmetic. Then solve it.
5. Write your own word problem about geometry. Then solve it.

Chapter 1

The Natural Numbers

This chapter introduces the natural numbers. Also known as the counting numbers, these are the numbers used for, you might expect, counting. Put differently, these numbers are the answers to “how many?” questions. They measure the quantity of discrete collections of objects. As such, to understand these numbers we will also need to understand discrete collections, which mathematicians call sets.

1.1 What even is counting?

This may seem like a silly question. Everyone knows what counting is! But when teaching children, you should resist the urge to dismiss a question. It is your job to teach those things that “everyone knows”, and children ask silly questions in order to learn.

Mathematicians are a lot like children. Much progress in mathematics is made by thinking deeply about an ‘obvious’ question. This understanding can then be the foundation for building new concepts. So let’s think about this question.

Checkpoint 1.1.1 Answer this question. What do you think counting is?

Perhaps your answer went something like this. When you count, you point at a collection of things one by one and list off the numbers in order: “one, two, three” and so on. When you’re pointing at the last thing, the number you said is the count of the collection.

This is a fine enough answer, but let’s dig deeper. What is a collection? Books on a shelf, social media accounts you follow, and solutions to an algebraic equation are examples of collections you can count. What about the tea in a mug, is that a collection? You might very well count the mug, but you can’t count the tea. With counting, there’s always a next thing in line to point at. That isn’t so with tea. You might have 200 mL of tea, but there’s no next quantity. You could have 201 mL, but you could also have 200.1 mL.

What we’re seeing is, we can only count **discrete** collections, where discrete means the collection can be separated into individual members. This is in contrast to a **continuous** collection, like the tea in a mug. To measure continuous collections we will need a different kind of number, which is the topic of [Chapter 6](#).

Another thing we need to count a collection is for it to be **finite**, meaning that no matter what order we count it there’s a start and an end. Collections of physical objects in ordinary life are all finite. But if you’re talking about

collections of mathematical objects, they can be **infinite**, meaning not finite. For example, you cannot count the collection of natural numbers. No matter how you order them, there's no last one. There's always a bigger number.

In summary, counting is about finite, discrete collections. This is the topic we now turn to.

1.2 Sets

Definition 1.2.1 A **set** is a discrete collection. A set could be a collection of physical things, such as the set of students in a classroom. It could also be a collection of abstract things, such as the set of counting numbers. The objects in a set are called its **elements** or **members**. ♦

Many of the sets of interest to mathematicians are infinite. Our focus, however, will be on finite sets, as those are the ones counting is about. We will talk about a few important infinite sets later, however.

First, however, let's talk about how we write sets.

Definition 1.2.2 Set notation.

- We name sets with capital letters, such as A or X . To say that an object is an element of a set we use the symbol \in , and we write \notin to say something is not an element. For example, if A is the set of cats the author of this book loves then $\text{Cleo} \in A$ but $\text{Willow} \notin A$.
- One way to define a set is by listing its elements. We write this by putting the elements between curly braces. For example, $P = \{2, 3, 5, 7\}$ is the set of prime numbers < 10 .
- For large sets it's impractical to define them by listing all their elements. For these we can use **set-builder notation**. The basic format is

$$\{x : \text{a pattern for which } x\text{'s are included}\}.$$

Here, x is a **variable** or name to refer to an unspecified object. For example, to define the set of even numbers below 100 you could write

$$E = \{n : n \text{ is an even number and } n < 100\}.$$

- Sometimes when using set-builder notation you want to only look at objects in a set you're already talking about. We do this by writing $x \in A$ on the left side of the colon. For example, with our set E from earlier you might want to restrict to those numbers which are also a multiple of 5. To do that, you'd write

$$\{n \in E : n \text{ is a multiple of } 5\}.$$

This set is called a **subset** of E , because it is contained inside E . ♦

Warning 1.2.3 When defining a set by listing its elements, stay away from dotdotdots! To you, it might be obvious what pattern is intended with $\{1, 2, \dots, 32\}$, but others won't find it obvious. (The pattern, by the way, was the powers of 2 up to 2^5 and so the set was $\{1, 2, 4, 8, 16, 32\}$. Is that what you thought the pattern was?) Instead, use set-builder notation to explicitly

say what the pattern is. In this case, it'd be better to write the set as

$$\{n : n = 2^k \text{ where } 0 \leq k \leq 5\}.$$

Checkpoint 1.2.4 Consider the set

$$C = \{\text{red, orange, yellow, green, blue, violet}\}$$

of colors.

1. List the elements of $\{c \in C : c \text{ is a primary color}\}$.
2. Use set-builder notation to write the sets $\{\text{orange, green, violet}\}$ and $\{\text{red, orange, yellow}\}$.

Checkpoint 1.2.5 Consider the set

$$N = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}.$$

Use set-builder notation to define two different subsets of N , and list the elements of those subsets.

Remark 1.2.6 A set can be formed from any objects, even if they're wildly unrelated. For example, $\{7, \text{the color blue, Nelson Mandela}\}$ is a perfectly fine definition for a set. In practice, however, the elements of a set are usually related. It's not useful to consider a set with three random elements.

Definition 1.2.7 Special sets. A few sets are worth highlighting, to understand the edge cases of the concept.

- The **empty set** is the set containing zero elements. We write it with the symbol \emptyset . That is, for any object x we have $x \notin \emptyset$.
- A set containing only one element is called a **singleton**. You would write it like, for instance, $\{a\}$. That is, an object x is an element of $\{a\}$ if and only if $x = a$.

◆

Checkpoint 1.2.8 Is a singleton set the same thing as its element? For example, is the number 2 the same thing as the set $\{2\}$? Explain your answer.

1.3 The meaning of the natural numbers

Definition 1.3.1 The natural numbers. The **natural numbers** are numbers which measure the quantities of finite sets. In mathematical jargon, we say that the **cardinality** of a set is the number of elements in it, and so the natural numbers are the cardinalities of finite sets. We write \mathbb{N} to refer to the set of all natural numbers. And we write $|A|$ for the cardinality of the set A .

When we need names for natural numbers, it is common practice to use variables from either the middle of the alphabet, such as i, k, n , or from the beginning of the alphabet, such as a, b, c . ◆

Speaking informally, it is common to talk about the size of a set when referring to its cardinality.

Unfolding jargon into ordinary language, the natural numbers measure quantities of finite, discrete collections. They are the answer to “how many?” questions. How many students failed the math exam? The answer could be seven, three, or, best of all, zero. On the other hand “how much?” questions

probably don't natural numbers as their answer. "How much water did you drink today?" most likely has a non-whole number as an answer. To measure **continuous** quantities like these we need a different number system, which we study in [Chapter 6](#).

We can give names to some natural numbers. Let's start with the very smallest a set can be, namely the empty set. We call the cardinality of the empty set **zero** and write 0. The next smallest a set can be is if we add a single element to it, call that element a . Starting from the empty set this gives a singleton $\{a\}$. We call the cardinality of a singleton **one** and write 1.

We can keep going with this process. We always get the next smallest cardinality by adding a single element. In this way we get **two** (2), **three** (3), **four** (4), **five** (5), and so on. We've so far named six natural numbers, and there's infinitely many still to go. It's apparent that at some point we need a more systematic naming scheme; that will be a topic in [Chapter 2](#).

What we will do in this chapter, however, is talk about operations on natural numbers. After all, you don't just want numbers to count, you also want them to calculate. Since natural numbers are so closely tied to sets it makes sense to first talk about the corresponding operations on sets. Once we understand those then we can see how they give rise to arithmetic operations.

1.4 Operations on sets

First let's talk about how to compare sets.

Definition 1.4.1 Subsets and set equality. A set A is a **subset** of a set B , written $A \subseteq B$, if every element of A is an element of B . That is, if $x \in A$ then $x \in B$. Similar to how we write \neq for not equals, we write $A \not\subseteq B$ to say that A is not a subset of B .

Two sets A and B are equal if they have the same elements. That is, for any object x , we have $x \in A$ if and only if $x \in B$. Note that this says $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$. \blacklozenge

Checkpoint 1.4.2 Explain why the "note that" from the definition is true.

The following summarizes some important properties of the subset relation \subseteq . Each property is given its formal mathematical name.

Theorem 1.4.3 Properties of \subseteq .

Reflexivity	$A \subseteq A$
Antisymmetry	If $A \subseteq B$ and $B \subseteq A$ then $A = B$.
Transitivity	If $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$.

These symbols can be unpacked into ordinary language. Reflexivity says that any set is contained in itself. Antisymmetry says that if two sets are contained in each other they must be the same. Transitivity says that if a set is contained inside a bigger set contained inside a yet bigger set, then the smallest set is contained inside the biggest.

Checkpoint 1.4.4 When comparing numbers, either they are the same or one is bigger than the other. This isn't the case with sets. Given two sets it is possible neither one is a subset of the other. Demonstrate this by finding two sets A and B so that $A \not\subseteq B$ and $B \not\subseteq A$.

In addition to comparing sets, we also want to combine them to make new sets. There are multiple ways we can do this

Definition 1.4.5 Intersection. The **intersection** of two sets is the set of their common elements. We use \cap as the intersection symbol. Using set-builder notation:

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$



Example 1.4.6 Consider $X = \{1, 3, 5, 7, 9\}$ and $Y = \{2, 3, 5, 7\}$. Then

$$X \cap Y = \{3, 5, 7\}.$$



Consider two **disjoint** sets, meaning they have no elements in common. For instance, consider $A = \{1, 3, 5\}$ and $B = \{2, 4, 6\}$. Then their intersection will have no elements, because nothing is in both A and B , and be the empty set. In symbols, $A \cap B = \emptyset$.

Definition 1.4.7 Union. The **union** of two sets is the set containing all elements from each of the sets. We use \cup as the union symbol. Using set-builder notation:

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$



Example 1.4.8 Consider $X = \{1, 3, 5, 7, 9\}$ and $Y = \{2, 3, 5, 7\}$. Then

$$X \cup Y = \{1, 2, 3, 5, 7, 9\}.$$



Definition 1.4.9 Set difference. The **set difference** of B from A , written $A \setminus B$, is the set you obtain by throwing everything from B out of A . Using set-builder notation:

$$A \setminus B = \{x \in A : x \notin B\}.$$



Example 1.4.10 Consider $X = \{1, 3, 5, 7, 9\}$ and $Y = \{2, 3, 5, 7\}$. Then

$$X \setminus Y = \{1, 9\}$$

and

$$Y \setminus X = \{2\}.$$



Often you want to express the idea of “first this thing, then that thing”. For example, when you buy a ticket at a movie theater you might have to pick the row and then the seat in that row. In mathematics, we abstract this kind of choice in the object of an **ordered pair**. Namely, we write (x, y) for the ordered pair whose first object is x and whose second object is y .

Definition 1.4.11 Cartesian product. Let A and B be sets. The **Cartesian product** of A and B , written $A \times B$, is the set of all ordered pairs (x, y) where $x \in A$ and $y \in B$. In set-builder notation:

$$A \times B = \{(x, y) : x \in A \text{ and } y \in B\}.$$



Checkpoint 1.4.12 What is the Cartesian product of $A = \{1, 2\}$ and $B = \{3, 4\}$?

Answer. $A \times B = \{(1, 3), (1, 4), (2, 3), (2, 4)\}$

You can visualize a Cartesian product as a rectangle. The vertical side represents the set A with all of its elements and the horizontal side represents the set B with all of its elements. To specify a position inside the rectangle you need to give both the vertical and horizontal position, which amounts to specifying first an element from A and second an element from B . Thus, the positions inside correspond to all the possible ordered pairs from A followed by B .

1.5 Number classes

Before we see how these set operations correspond to arithmetic operations with natural numbers, let's take a detour to look at some special sets much beloved by mathematicians. Namely, mathematicians work a lot with different kinds of numbers, and so it is convenient to have names for the sets of those numbers. These form a hierarchy of larger and larger classes of numbers. We use a special font, called “blackboard bold” to write these number classes. This font choice makes it clear that we are talking about a number class and not using a letter as a variable for some other kind of thing.

Definition 1.5.1 Number classes.

- The **natural numbers** \mathbb{N} are the counting numbers used to measure discrete collections. (\mathbb{N} for Natural.)
- The **integers** \mathbb{Z} are the positive and negative whole numbers. (\mathbb{Z} for Zahlen, the German word for integers.)
- The **rational numbers** \mathbb{Q} are the numbers formed from fractions of integers. (\mathbb{Q} for Quotient.)
- The **real numbers** \mathbb{R} are the numbers which measure continuous quantities. (\mathbb{R} for, you guessed it, Real.)
- The **complex numbers** \mathbb{C} are an extension of the reals used for advanced applications in, for instance, physics and engineering.

These number classes are larger and larger as you go along. In set-theoretic notation:

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$



In this chapter we are looking at the natural numbers. The next two chapters will give a deeper look at them. Then in [Chapter 4](#) we look at the integers, followed by the rational numbers in [Chapter 5](#) and the real numbers in [Chapter 6](#). We won't look at the complex numbers in this book.

1.6 The arithmetic of natural numbers

Now that we've seen operations for combining sets we can talk about their counterpart operations on numbers.

If you have four mangos and three pineapples, how many fruits do you have in total? If M is the set of your mangos and P is the set of your pineapples, what's being asked is, what's the cardinality of $M \cup P$?

Definition 1.6.1 Addition of natural numbers. Suppose m and n are natural numbers. Let A and B be disjoint sets with $|A| = m$ and $|B| = n$.

Then

$$m + n = |A \cup B|.$$

When we need to refer to the different parts of addition, we call $m + n$ the **sum**, while m and n are the **summands**. ♦

You can phrase this as being about counting choices. If you have your choice either from one of m options or from one of n options, then you have $m + n$ many choices. For example, if a restaurant offers 3 options of soup or 2 options of salad, that is $3 + 2$ many options total.

Once you know about addition you can give a clear definition of what it means for one natural number to be bigger than another. On the level of sets, the idea is, a finite set A is larger than another B if you can add things to B to make it the same size as A . On the level of numbers, m is larger than n if you can add to n to get m .

Definition 1.6.2 Order on natural numbers. Suppose m and n are natural numbers. Then $m \geq n$ if there is a natural number k so that $m = n + k$. If $m \geq n$ but $m \neq n$ we say $m > n$. And we reverse the order of the symbols to mean the opposite direction: $m \leq n$ means $n \geq m$ and $m < n$ means $n > m$. ♦

Next we define subtraction. If you have nine fruits and I take away your five guavas, how many fruits do you have in total? If F is the set of your fruits and G is the set of your guavas, what's being asked is, what's the cardinality of $F \setminus G$?

Definition 1.6.3 Subtraction of natural numbers. Suppose $m \geq n$ are natural numbers. Let A be a set with cardinality m and let $B \subseteq A$ be a subset of cardinality n . Then

$$m - n = |A \setminus B|.$$

When we need names for the different parts, $m - n$ is the **difference**, the number m being subtracted from is the **minuend**, and the number n being subtracted is the **subtrahend**. ♦

An alternative way you can think about subtraction is as the opposite of addition. Namely, $m - n = k$ is true if and only if $n + k = m$. For example, $17 - 4 = 13$ because $4 + 13 = 17$. In words, to solve $17 - 4$ you ask, what do I add to 4 to get a sum of 17?

This perspective on subtraction is fruitful as a foundation to build further understanding, as we will see in [Chapter 4](#). In general in mathematics, having multiple perspectives on the same idea is great. It gives us more tools and more approaches to understand things and to solve problems.

Note that, if we are only dealing with natural numbers, subtraction only makes sense when the subtrahend is \leq the minuend. If you have 5 fruits you can't take 7 away! In [Chapter 4](#) we will see how to expand to a larger number system which allows you to subtract any two numbers without worrying about which one is bigger.

Multiplication is a step trickier to define. To illustrate, suppose you have four boxes, each of which contains six papayas. How many papayas do you have in total? One way to think of this is, how many ways are there for you to pick out a single papaya? It's a two-step process: first pick a box, then choose a papaya from that box. Mathematically, we represent a two-step process as an ordered pair. If we number the boxes and number the papayas in each box, each choice can be represented as the pair (b, p) of the box number b followed by the papaya number p .

Definition 1.6.4 Multiplication of natural numbers. Suppose m and n are natural numbers. Let A be a set of cardinality m and B be a set of cardinality n . Then

$$m \cdot n = |A \times B|.$$

When we need a name for the parts of multiplication, $m \cdot n$ is the **product** while m and n are the **multiplicands**.

When used with variables, multiplication is often written just by putting them next to each other, e.g. mn instead of $m \cdot n$. One also sees this when one of the two multiplicands is a number, such as $3x$. More rarely, one sees multiplication written as $m \times n$. ♦

Like with addition, you can phrase this in terms of counting choices. If you make one choice from m options then you independently make a second choice from n options, there are mn different ways of making the two-step choice. For example, if a fast food restaurant offers 6 sandwiches and 4 sides, then there are $6 \cdot 4$ many combos you could order.

Now that we've established the meaning of the basic operations, let's make clear the basic rules they satisfy.

Theorem 1.6.5 Arithmetic properties of natural numbers. *All of the following are true for any integers x, y, z .*

Additive identity	$x + 0 = x$
Associativity of addition	$x + (y + z) = (x + y) + z$
Commutativity of addition	$x + y = y + x$
Multiplicative identity	$x \cdot 1 = x$
Multiplicative annihilation	$x \cdot 0 = 0$
Associativity of multiplication	$x \cdot (y \cdot z) = (x \cdot y) \cdot z$
Commutativity of multiplication	$x \cdot y = y \cdot x$
Distributivity	$x \cdot (y + z) = xy + xz$

Mathematicians call these properties the **commutative semi-ring axioms**, but that term will not be helpful to us.

We will tackle exponentiation in a different way. With addition, subtraction, and multiplication, we defined these in terms of cardinalities of sets based on certain operations. That approach is possible for exponentiation—indeed, the exercises in this chapter guide you through that approach—but it is helpful to take another perspective.

Namely, we will consider exponentiation as iterated multiplication. In plain terms, exponentiation means repeatedly multiplying a number by itself.

Definition 1.6.6 Exponentiation of natural numbers. Let a and n be natural numbers. Then,

$$a^n = \underbrace{a \cdot a \cdot \dots \cdot a}_{n \text{ many}}$$

We call a the **base** and n the **exponent**. ♦

This rule doesn't tell us what to do with an exponent of 0. We will do the only sensible thing, and say that $a^0 = 1$, no matter what a is.

Like with the other operations, we can write down rules that describe how exponentiation behaves.

Definition 1.6.7 Rules for exponentiation. All of the following are true for any integers a, b, m, n .

- $a^n \cdot a^m = a^{n+m}$
- $(a^n)^m = a^{nm}$
- $a^n \cdot b^n = (ab)^n$

◆

1.7 Division

Division, denoted by the symbol $/$, is the most difficult—one might say the most divisive—operation to handle, so let's address it in its own section.

When talking about the other operations, we saw that one way to think about subtraction is as the opposite of addition. We can do the same for division. Namely, define that $n/d = q$ if $dq = n$. We call n the **dividend**, d the **divisor**, and q the **quotient**. For example, $6/3 = 2$ because $3 \cdot 2 = 6$.

You can phrase division as being about dividing up a collection. That is, $n/d = q$ if you can divide n many objects up into d many groups each with q things. For example, $6/3 = 2$ because you can divide six kiwis into three groups of two kiwis each.

A difficulty with division is that if you divide two random natural numbers almost certainly the quotient won't be a whole number. For example, living in \mathbb{N} we can't do the division $3/2$ because there is no whole number which gives 3 when multiplied by 2. If you have three students and you are trying to split them into two equally sized groups, there is no way to do that.

On the other hand, if you are trying to split three oranges between two students, you can do that. Each student gets one and a half oranges. What this example mathematically illustrates is, we can make sense of divisions like $3/2$ but it requires going beyond the whole numbers. This will be the subject of [Chapter 5](#).

If we want to divide in the whole numbers, instead of **exact division** we have to do **division with remainder**.

Theorem 1.7.1 Division with remainder. *Let n and d be natural numbers, with $d \neq 0$. Then there are unique numbers q and $0 \leq r < d$, the **quotient** and **remainder**, so that*

$$n = dq + r.$$

It is perhaps easier to see what is going on if you phrase this equation as an equivalent one about fractions:

$$\frac{n}{d} = q + \frac{r}{d}.$$

In words, if you want to divide n objects into d groups, you can split them into groups of size q with r many left over.

For example, if we want exact division then $14/3$ cannot be done in the whole numbers. But we can do division with remainder to get a quotient of 4 and a remainder of 2. That is, $14 = 3 \cdot 4 + 2$. You can think of it this way: if

you want to divide 14 students into groups of 3 you will have 4 groups with 2 students left over.

Another way to think about division with remainder is, if you tried to get an exact answer to $14/3$ you wouldn't get a whole number—in this case, you get about 4.67. The quotient for division with remainder is the natural number you get if you round down to the nearest whole number, and the remainder represents the fractional part many 3's. In this case, the remainder 2 is about 0.67 3's.

In [Chapter 2](#) we will talk about how you compute division with remainder. In [Chapter 3](#) we will talk more about what further math you can do with this idea.

Checkpoint 1.7.2 Explain why division by zero is never defined. That is, explain why there is never a unique number q you can pick so that $n/0 = q$ makes sense.

Hint 1. Remember that division is the opposite of multiplication. Turn $n/0 = q$ into an equivalent equation about multiplication.

Hint 2. Treat the case where the dividend $n = 0$ separately. It needs a different explanation.

1.8 Order of Operations

When doing calculations, you often have to combine different operations. The order you combine them matters. For example, consider

$$4 + 3 \times 2.$$

Which operation is done first, addition or multiplication? If you do addition first, the value of the expression is $7 \times 2 = 14$. If you do multiplication first, the value of the expression is $4 + 6 = 10$.

It's convenient to have a convention for which operation gets performed first.

Definition 1.8.1 Order of operations. When computing arithmetic expressions, do them in the following order.

1. (Parentheses) Parentheses (and) are used to indicate priority. Do operations inside parentheses first.
2. (Exponentiation) Next do exponentiation.
3. (Multiplication/Division) Next do multiplication and division. The order doesn't matter here; you get the same value whether you multiply or divide first.
4. (Addition/Subtraction) Last do addition and subtraction. The order doesn't matter.

Additionally, when doing multiplication or addition you can freely write three or more terms without using parentheses to indicate priority. For instance, $3 + 4 + 5$ is the same value whether you do left sum or the right sum first, so you don't need to specify. On the other hand, take care with subtraction or division with three or more terms. Convention is to go left to right, so $6 - 4 - 1$ means $(6 - 4) - 1$ (that is, do the left subtraction then the right subtraction). But this is a different value than $6 - (4 - 1)$. \blacklozenge

Example 1.8.2

$$2 \cdot (3 \cdot 4 - 5)$$

The parentheses mean you must do what's inside before you can do the multiplication on the outside. Inside the parentheses, multiplication must be done before subtraction. So the order of operations is:

1. Multiply 3 and 4.
2. Then subtract 5 from that.
3. Then multiply that by 2.



Checkpoint 1.8.3 Consider the following arithmetic expression.

$$(1 + 2)(3 - 4 \cdot (5 + 6))$$

Draw a tree diagram which illustrates the order of operations for this expression.

Some rules of mathematics can't be bent (or at least, not without breaking a lot of other things). For example the rules in [Theorem 1.6.5](#) are firm. Anything times 0 is 0, $a + b$ is the same as $b + a$, and so on. The order of operations, however, isn't like that. It's merely a convention in how we write expressions. We could choose a different order, and math would work just as well. The reason for this specific choice is that it makes many common expressions fast to write. For example, a **polynomial** like

$$2x^3 - 3x^2 + 4x - 1$$

is efficiently written using this order of operations.

Other conventions for order of operations get used in some calculators and programming languages. One example is **prefix notation**, also called **Polish notation**. (The latter name is in honor of Jan Łukasiewicz, the Polish logician who invented the notation.) In this notation, the operations are written first and the numbers/inputs afterward. For example,

$$+ 3 2$$

reads "add the next two things, namely 3 and 2". More complicated expressions can also be made.

$$+ \times 3 2 1$$

reads "add the next two things, namely the product of 3 and 2, and 1". On the other hand,

$$\times + 3 2 1$$

reads "multiply the next two things, namely the sum of 3 and 2, and 1". As shown by these two examples, prefix notation doesn't need parentheses to give priority, which makes it appealing to logicians and computers. But it's not how humans usually write math, and we won't discuss it any further in this book.

1.9 Exercises

Exercises

For these counting exercises, write your answer as a formula, not as a number. For example, write $(2 + 3) \cdot 4$ instead of 20.

1. A date night special at a restaurant lets you pick either a soup or salad as a starter, then an entree, and finally a dessert. If there are three soups, two salads, five entrees, and four desserts, how many different meals can you choose? If you and your date independently choose your own meals, how many combinations are there for how the two of you can order?

2. In a race with 6 racers, how many different outcomes are there for who gets first through sixth place?

Hint. Think of choosing an outcome as first picking who comes in first, then picking a different person for second, and so on to sixth place. How many options do you have for first? Then how many for second? And so on.

3. How many ways are there to pick a 4 digit PIN? A 6 digit PIN. A n digit PIN for any positive whole number n ?

Hint. Think of picking each digit as an independent choice. How many options for one digit? Now put it together for multiple digits.

In more advanced uses of sets, it's helpful to talk about mappings or correspondances between them. How do you assign each element of the input set an element of the output set?

Definition 1.9.1 Functions. Let A and B be sets. A **function** f from A to B , which we write as $f: A \rightarrow B$ is a way of assigning each input $x \in A$ to an output $f(x) \in B$. \blacklozenge

4. Consider $A = \{1, 2, 3\}$ and $B = \{1, 2, 3, 4, 5\}$. Give an example of a function $f: A \rightarrow B$ by specifying the three outputs $f(1), f(2), f(3)$.

We can use functions to give an alternative definition of exponentiation. Namely, a^n counts the number of functions $f: N \rightarrow A$ from a set N of size n to a set A of size a .

5. Confirm the two definitions of exponentiation are equivalent for a specific case by counting how many functions there are from a set N with 4 elements to a set A with 2 elements.

Hint. Think of choosing a function as choosing an output for the first element of N , then the next one, and so on until you've picked an output for every function.

6. Confirm that this is in general equivalent to the definition by iterated multiplication by explaining why there are

$$\underbrace{a \cdot a \cdots a}_{n \text{ many}}$$

functions from a set N of size n to a set A of size a .

Hint. Same hint as before, but now instead of having a specific value for n and a you have to think a step more abstractly.

A function $f: A \rightarrow B$ is called **one-to-one** if different inputs get sent to different outputs. It is called **onto** if every element of B is the output $f(x)$ for some input $x \in A$. If f is both one-to-one and onto it is called **bijective**.

7. Let $A = \{1, 2, 3\}$ and $B = \{a, b, c\}$. Come up with a function $f: A \rightarrow B$ which is one-to-one by specifying the outputs $f(x)$ for every input $x \in A$. Also come up with a function $g: A \rightarrow B$ which is onto. Can you come up with a function $h: A \rightarrow B$ which is onto but not one-to-one? Either demonstrate such a function or explain why you think it's impossible.

8. Consider $f: \mathbb{N} \rightarrow \mathbb{N}$ defined as $f(n) = n + 1$. Explain why f is one-to-one but not onto. Also consider $g: \mathbb{N} \rightarrow \mathbb{N}$ defined as $g(n)$ is the whole number

you get rounding down from $\frac{n}{2}$. Explain why g is onto but not one-to-one.

One use of functions is to give a way to compare sizes of sets. If A and B are sets, then $|A| \leq |B|$ if there is a one-to-one function $f: A \rightarrow B$ and $|A| = |B|$ if there is a bijective function $f: A \rightarrow B$.

9. Demonstrate that your left hand and right hand have the same number of fingers by describing a bijection from the set of fingers on your left hand to the set of fingers on your right hand.
10. Demonstrate that any two singletons $\{a\}$ and $\{b\}$ have the same cardinality by giving a bijective function between them.

Chapter 2

Numeration

In the previous chapter we learned about the natural numbers, those numbers that are used for counting. We also learned the basic arithmetic operations and how they are related to counting. While this is important theoretical knowledge for working with numbers, it doesn't tell us how to work with them in practice. This chapter is about two closely related topics. We will learn about numeral systems, how you write or name numbers, and how these facilitate calculations.

2.1 Roman Numerals

It is easy to look at ideas that are the products of millennia of work and think they are natural or obvious, and thereby take them for granted. To forestall this mismeasure, let's look at a clumsy numeral system before an efficient one. We will look at one which still sees limited use, e.g. sometimes for dates. Namely we will look at **Roman numerals**.

The kernel of the idea for Roman numerals can be seen in tally marks. Here, a number is represented by the corresponding number of tallies. For example, $1 = |$ and $3 = |||$. Beyond very small numbers this hard to read, so there are devices such as grouping tally marks by fives by writing the fifth tally mark diagonally across the group. Roman numerals iterate on this idea by having different letters stand for different quantities. Rather than writing three groups of ten tally marks to mean 30 you would just write XXX, with each X standing for ten.

Definition 2.1.1 Roman numerals. From smallest to largest, these are the quantities letters represent in Roman numerals.

- I represents 1.
- V represents 5.
- X represents 10.
- L represents 50.
- C represents 100.
- D represents 500.
- M represents 1000.

These are combined according to the following rules.

- Up to three I's, X's, C's, or M's may be written in a row to represent that many 1s, 10s, and so on. V's, L's, and D's can only be written in groups of one, possibly followed by the next lowest denomination. For example, VII represents $5 + 2 = 7$.
- To represent four or nine copies you use subtractive notation: IV for 4, IX for 9, XL for 40, XC for 90, CD for 400, and CM for 900.
- Write the letters representing different denominations from largest to smallest. For example, MCXI represents the number 1111 and MMCDVI represents the number 2406.



Example 2.1.2 Here are a few examples of Roman numerals.

- 48 is $XL + VIII = XLVIII$.
- 2026 is $MM + XX + VI = MMXXVI$.
- 3999 is $MMM + CM + XC + IX = MMMCMXCIX$, the largest number that can be represented.
- 2003 is $MM + III = MMIII$.
- 409 is $CD + IX = CDIX$.



There are many downsides to Roman numerals. Just writing and reading them can be awkward, as you maybe got from these examples. The rules are finicky, and the subtractive notation means that a symbol can play two different roles; depending on location, X might mean add ten, or it might mean subtract ten.

Another downside is that Roman numerals can only represent numbers up to 3999. While that suffices for many everyday uses, such as dates, it quickly becomes a problem. You can address the problem by expanding the notation—introduce new letters for 5000, 10,000, and so on. But that just pushed the problem upward a little. What do you do when you run out of letters?

The biggest problem, however, is that Roman numerals don't admit easy calculations. How do you quickly compute the sum $LXXIV + XLVIII$? Indeed, even when Roman numerals were in common use to write numbers, actual calculations were done by representing numbers with abacuses. For that reason, we won't even try to come up with algorithms for doing calculations with Roman numerals. Instead, we will look at a better system.

2.2 The Hindu–Arabic numeral system

A key insight of this numeral system is that it is a **positional** system. This means that the magnitude a symbol represents is determined in part by its location. This enables a greater flexibility, that both allows arbitrarily large numbers to be written and facilitates easier calculations. In this section we describe how to write numbers in this system, and in the following section we will see how to do calculations with them.

First, a little history to explain the name. The Hindu–Arabic numeral system originates in India and moved west to the Arabic world. An important milestone here is the work of the Persian mathematician Al-Khwarizmi (whose name the word “algorithm” derives from). He wrote books on calculations

with these numerals and on algebra calculation. Indeed, the word “algebra” comes from the title of one of his works. From the Arabic world these numerals eventually made their way to Italy and the rest of Europe. Eventually, they overtook their competitors due to their better facilitation of calculation, until they became the numeral system we teach to schoolchildren today.

An important discovery to enable the positional system was the number zero. For example, consider the numbers 19, 109, and 1009. In each of these, 9 is in the units place, counting the number of ones in the quantity. The 1, however, varies. In these numbers it counts, in order, the number of tens, hundreds, or thousands. If we didn’t have a digit for zero how would we distinguish these? If you write them all as 19 you can’t tell them apart. Leaving space is hardly any better and is prone to error. How many spaces does 1 9 have?

One solution to this problem, such as in a numeral system used by the ancient Greeks, is to use a different symbol for 100 versus 10. Then there is no ambiguity which is meant. This has similar shortcomings as the Roman numeral system. You also need a new symbol for 1000, 10,000, and so on. Useable if you’re only dealing with small numbers, but unwieldy when things start to get large.

With 0 you can use the single symbol 1 to refer to all of these, with the position—the number of other digits to the right—determining whether it represents one ten, one hundred, one thousand, or so on.

Example 2.2.1 Consider the number 372. This represents three hundreds plus seven tens plus two ones. (It is common to say “units” instead of “ones”.) That is, this represents the number given by

$$300 + 70 + 2.$$

Or consider the number 30720. This represents

$$30000 + 700 + 20.$$

Or consider the number 300000072. This represents

$$3 \times 10^9 + 70 + 2.$$

Note that in each case, each digit represents some number of a power of ten. ▲

As these examples illustrate, this numeral system can flexibly represent numbers of any magnitude using just ten symbols. Sometimes people use a comma to separate out blocks of digits to make it easier to read. Standard convention in America is to separate out blocks of three. For example, rather than write 3000000072 you might write

$$3,000,000,072.$$

This way you can easily count blocks and see that this number is about three billion.

Definition 2.2.2 The Hindu–Arabic numeral system. Numbers in the **Hindu–Arabic numeral system** are written as finite sequences of **digits**. Digits are the numerals

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9$$

and represent the first ten natural numbers, starting at zero. Each digit in a numeral represents the corresponding number of a power of ten. The rightmost place represents the number of units = 10^0 s, the next place represents the

number of tens = 10^1 s, the next represents the number of hundreds = 10^2 s, and so on.

In symbols, we write a number as

$$d_n d_{n-1} \cdots d_1 d_0$$

where each digit is a numeral 0 through 9. The i th digit d_i represents the number of 10^i s. So this represents the number given by

$$d_n \cdot 10^n + d_{n-1} \cdot 10^{n-1} + \cdots + d_1 \cdot 10^1 + d_0 \cdot 10^0.$$

This numeral system is also called a **decimal** or **base ten** numeral system because it uses ten different digits and each place represents a power of ten. \blacklozenge

This numeral system lets us represent numbers efficiently. If you have seen logarithms before, to represent the number n we need about $\log n$ many digits. If you haven't seen logarithms, the point is that this number of digits is much smaller than n itself. For example, to represent one million you need only seven digits. Seven is much smaller than one million! Compare to a tally mark system. To represent n you need n many tallies. That is horribly inefficient.

While this system is nice for writing and reading numbers, its main virtue is in how it enables efficient algorithms for arithmetic. This is the topic we turn to in the next section, and will make up the bulk of this chapter.

2.3 Algorithms for adding, subtracting, and multiplying

Now that we know how to write any natural number, we want to be able to do arithmetic with them. In this section we will talk about how to use Hindu-Arabic numerals to do addition, subtraction, and multiplication.

A key thing a positional numeral system does is allow us to break down a large calculation into small pieces, which can be individually tackled. All of our arithmetic algorithms will be carried out digit by digit. Rather than having to memorize what $342 + 596$ is you can break it down into three single digit problems.

Before seeing the general algorithm, let's see an example to illustrate what's going on.

Example 2.3.1 Consider the sum $342 + 596$. We will work from the units digit to the left. For the units digit, we compute the sum $2 + 6 = 8$. We know that will be the units digit of the sum because the rest of the digits of the summands are too large to affect the units digit. Next we handle the tens digit. We compute $4 + 9 = 13$. That is, the tens from the two summands add up to 13 tens in the sum. We will keep 3 as the tens digit of the sum, carrying the 1 over to be part of the hundreds digit. Finally, for the hundreds digit we do $3 + 5 + 1 = 9$. Altogether, the sum is 938. \blacktriangle

A few points. First, this example was intentionally wordy to emphasize the process. The algorithm you remember from elementary school is the same idea, but in a compact and efficient form. Second, note that the number of steps in the process was the number of digits in the summands. Each extra digit adds one more step. In other words, you can solve a problem ten times as large with just one extra step. That is very efficient. Third, for each step we had to do a sum of one-digit numbers. For this to be fast we relied on having the one-digit sums memorized. You can do this process without having them

memorized, but it will be a slower process.

This explains the reason for making children learn times tables and similar for addition/subtraction. Certainly one can critique the implementation—for example, timed quizzes creating counterproductive stress for students. But larger calculations consist of multiple single digit calculations done in succession. For example, multiplying two three digit numbers requires nine one-digit multiplications. Besides the speed advantage of automatic recall of these one-digit results, it reduces the cognitive load on the student. This allows them to focus their attention on mastering the new concepts.

This is a pattern that recurs in mathematics education. A student who needs to spend a lot of mental energy on arithmetic will have a harder time picking up algebra. A student who needs to spend a lot of mental energy on algebra will have a harder time picking up calculus. This is why elementary educators—people like you!—play such an important role in math education. What you teach is the foundation for the rest of their education.

This example done, let's move on to the general algorithm. First I'll verbally describe it, then I'll show a compact way to visually organize the calculation.

Definition 2.3.2 Algorithm for addition. To add two numbers written in the Hindu–Arabic numeral system, use the following procedure, starting at the units digit and moving to the left.

- Add the two digits together to get the total for the corresponding digit of the sum. If the total is < 10 , that is the digit for the sum. If the total is ≥ 10 , the units digit becomes the digit for the sum and the 1 in the tens digit gets carried left to be added to the next digit.
- If one number has more digits than the other, treat the empty digits as 0s. That is, the digit for the sum is just the lone digit copied, possibly with +1 carried from the previous digit.
- Stop when you have exhausted all digits for both numbers. If there is a +1 being carried, put it in the next digit for the sum.

◆

Now let's see a visual representation of a sum.

$$\begin{array}{r}
 1 \\
 342 \\
 + 596 \\
 \hline
 938
 \end{array}$$

Figure 2.3.3 The sum $342 + 596$. This is the same example as before, but with a compact representation rather than a short essay.

This algorithm is flexible and can be modified to handle adding more than two numbers. To do so, rather than add two digits per column, you add more than that. This has the extra complication that carrying is more common and it's possible to have to carry +2 or more. (If you add a truly absurd volume

of numbers you might have to carry forward two digits, but there's a simple solution to that: don't add so many numbers at once. If you must, break it up into smaller sums and combine them.) Here's an example.

$$\begin{array}{r}
 + \quad 123 \\
 + \quad 369 \\
 + \quad 789 \\
 \hline
 1281
 \end{array}$$

Figure 2.3.4 The sum $123 + 369 + 789$.

Let's move on now to subtraction. Like with addition, we proceed digit by digit, starting at the units digit moving to the left. An example before the general algorithm.

Example 2.3.5 We compute the difference $437 - 182$. For the units digit we do $7 - 2 = 5$ to get the units digit of the difference. For the tens digit, we first try $3 - 8$, but that's a problem because $3 < 8$. To make this work, we have to borrow a 1 from the hundreds column. Then the subtraction is $13 - 8 = 5$, giving us the tens digit of the difference. Finally, for the hundreds digit we borrowed a 1 so we have to reduce our minuend by one: $4 - 1 = 3$. We then do $3 - 1 = 2$ to get the hundreds of the difference. Altogether, this gives 255 as the difference. ▲

With addition, we sometimes had extra value spill over and be carried to the next column. Subtraction, being the opposite of addition, has the opposite phenomenon—we have to borrow extra value from the next column. You might worry that we borrow so much that eventually there's not enough left. That's what would happen if you tried to subtract a larger number from a smaller. But as long as we keep where subtraction is defined on the natural numbers—subtracting a smaller number from a larger—this will not be a problem.

Remember the terminology. It's minuend $-$ subtrahend $=$ difference!

Definition 2.3.6 Algorithm for subtraction. To subtract a subtrahend from a larger minuend, both written in the Hindu–Arabic numeral system, use the following procedure, starting at the units digit and moving to the left.

- Subtract the two digits together to get the corresponding digit of the difference. If you are subtracting a smaller digit, there is no complication. If you are subtracting a larger digit, you need to borrow a 1 from the next digit so that you are subtracting a smaller number.
- If the previous digit had to borrow a 1, subtract 1 from the digit of the minuend before subtracting digits. This might require you to borrow from the next digit.
- If the minuend has more digits than the subtrahend, treat the empty digits as 0s. That is, the digit for the difference is just the lone digit copied, possibly with 1 borrowed by the previous digit.
- Stop when you have exhausted all digits for both numbers.



$$\begin{array}{r} 437 \\ - 182 \\ \hline 255 \end{array}$$

Figure 2.3.7 The difference $437 - 182$. This is the same example as before, but with a compact representation rather than a short essay.

Now let's discuss multiplication. Algorithms from multiplication are built on an algebra fact, namely the **distributive** property of multiplication:

$$x \cdot (y + z) = xy + xz.$$

If x is itself a sum, you can distribute twice:

$$\begin{aligned} (a + b) \cdot (y + z) &= (a + b)y + (a + b)z \\ &= ay + by + az + bz. \end{aligned}$$

You can visualize this double distributivity by drawing it as a 2×2 box. The two columns represent a and b while the two rows represent y and z . Each square in the box represents the product of the two, and the overall product is the sum of all four squares.

Table 2.3.8 The multiplication box for $(a + b) \cdot (y + z)$.

	a	b
y	ay	by
z	az	bz

If each of a, b, y, z represents a single digit, then to carry out the multiplications inside the squares only requires you to know the one-digit products, similar to how we reduced the algorithms for addition and subtraction to knowing the one-digit sums and differences. For example consider the multiplication

$$13 \cdot 27.$$

We think of this as $(10 + 3) \cdot (20 + 7)$, and we can use a multiplication box.

Table 2.3.9 The multiplication box for $13 \cdot 27$.

	10	3
20	200	60
7	70	21

To get the overall product we add up these four numbers:

$$200 + 60 + 70 + 21 = 351.$$

And thus we have calculated $13 \cdot 27$.

Checkpoint 2.3.10 Use the box method to calculate the product $52 \cdot 29$.

This box method can be used for any two digit multiplication. If you do enough of them, you'll start to notice that the upper left number always has two 0s, the upper right and lower left squares have one 0, and the bottom right square has none. (Caveat: you may have extra zeros from the one-digit

multiplication, for example if its $4 \cdot 5 = 20$.) You may find it convenient to not write the extra zeroes and only write the value of the one-digit product, knowing where to put zeroes when adding them up. Or you may prefer to always write out the zeroes.

You can adjust the box method to work with more two digits of multiplication. For example, to multiply two three-digit numbers you would need a 3×3 box with 9 squares in it. You would do the multiplication for each square, then add up all nine numbers. In general, to multiply an m -digit number by an n -digit number you need a $m \times n$ box with mn many squares to add up.

Larger boxes can get unwieldy, so let's look at a different presentation of the same idea. This looks more like the presentation of the algorithms for addition and subtraction. It essentially amounts to collapsing the box into a single column. Accordingly, as a stepping stone we will first look at multiplying by a one-digit number.

Example 2.3.11 Let's do the multiplication 37×3 . We think of this as $30 \times 3 + 7 \times 3$. Do each one-digit multiplication separately and then add the answers. We can organize this into a box.

Table 2.3.12 The multiplication box for $37 \cdot 3$.

	30	7	
3	90	21	

Adding up the values we get that the product is $90 + 21 = 111$.

You could also write this in the column format similar to the algorithms for addition and multiplication. First compute 7×3 to get 21, which means 1 in the units place and carry the 2 to the tens digit. Then in the tens digit we do 3×3 and add in the carried 2 to get 11. That means 1 in the tens column and carry the other 1 to the hundreds place.

$$\begin{array}{r}
 37 \\
 \times 3 \\
 \hline
 111
 \end{array}$$

Figure 2.3.13 The product 37×3 .

▲
This process lets us do a multiplication by a one-digit number as a single row. To do a multiplication by a multi-digit number we do this process for each digit, then add everything together. When we do this, we offset the multiplication by a number of tens to account for the position shift given by the ten. Let's see an example.

Example 2.3.14 Let's do the multiplication 28×52 . We think of this as $28 \times 2 + 28 \times 50$. Each of these multiplications requires computing two one-digit products. For the first, $8 \times 2 = 16$, and we must remember to carry the one to the next column, and $2 \times 2 = 4$, so the overall product is 56. For the second, $8 \times 50 = 400$ (carry the four) and $2 \times 5 = 10$ (carry the one again). So

the overall product is 1400. Note the extra zero, which came from the zero in 50. We then add up these two numbers to get an overall product of 1456.

This description of the algorithm can be organized neatly.

$$\begin{array}{r}
 28 \\
 \times 52 \\
 \hline
 56 \\
 \\
 140 \\
 \hline
 1456
 \end{array}$$

Figure 2.3.15 The product 28×52 .

Checkpoint 2.3.16 Use the box method to compute 28×52 . Compare the two methods. ▲

Both methods require four one-digit multiplications. The box method organizes it into four different squares while the columns method does them two at a time, organizing them into rows. This makes it more compact, which is convenient when doing larger multiplications. For instance, to multiply two three-digit numbers it would have three rows to add up for the overall product, while the box method would have nine squares to add.

Definition 2.3.17 Algorithm for multiplication.. To multiply two multi-digit numbers a and b , do the following.

1. Write a and b aligned so that the units, tens, etc. places are in the same columns.
2. Multiply a by the units digit of b , according to the above method. Write this product aligned with the units digits of the inputs.
3. Multiply a by the next digit of b . Write this product aligned with the next column of the inputs.
4. Repeat this process until all digits of b are used up. Finally, sum up the partial products to get the overall product.

This algorithm works fine if b has more digits than a . But it's nicer to organize it so the second multiplicand is the one with fewer digits. This way there's fewer rows to write down and add up. ◆

Let's see an example. We will multiply 347×72 .

$$\begin{array}{r}
 347 \\
 \times 72 \\
 \hline
 694 \\
 2429 \\
 \hline
 24984
 \end{array}$$

Figure 2.3.18 The product 347×72 .

Checkpoint 2.3.19 Use a version of the box method to calculate 347×72 . (You will need a 3×2 box.) Compare the two methods.

The box method and the columns method are essentially the same algorithm. At base, each reduces the multi-digit multiplication into many one-digit multiplications. And both rely on the same algebraic property to work, namely the distributivity of multiplication over addition. In a sense, it's not quite right to say they are two different methods. Rather, they are two different ways of organizing the same method. A common trick in mathematics education is to present multiple views on the same concept. Different lenses work better for different people, and the flexibility to see the same thing in different ways is valuable for future learning.

For certain products you can optimize the space the column method takes up. Namely, if a digit is 0 then multiplying by it gives 0, so you can just not write down the row corresponding to multiplication by that digit. For example, the multiplication 273×30020 can be done with only two rows.

$$\begin{array}{r}
 20703 \\
 \times 30020 \\
 \hline
 41406 \\
 62109 \\
 \hline
 621504060
 \end{array}
 \qquad
 \begin{array}{r}
 20703 \\
 \times 30020 \\
 \hline
 00000 \\
 41406 \\
 00000 \\
 00000 \\
 62109 \\
 \hline
 621504060
 \end{array}$$

Figure 2.3.20 The product 20703×30020 .

2.4 Other bases

Our number system is base 10 or **decimal**. There are ten digits, and each column we write represents a power of 10. There's nothing special about 10 here. You can write numbers with any base and the algorithms we've discussed still work, with appropriate modification.

The most used base besides 10 is base 2, also called **binary**. In binary, there are only two **bits** (= **binary digits**), namely 0 and 1. Each column

represents a power of 2, with 1 representing that you include that power of 2 while 0 represents excluding that power of 2. This simplicity of each bit being either on or off is part of why binary works so well in computers. The ons versus offs—1s versus 0s—can be implemented in electrical circuits and you can connect wires with **logic gates** to create hardware that can perform calculations such as addition or multiplication.

For example, consider the binary number 10011. This number represents the value

$$2^4 + 2^1 + 2^0 = 16 + 2 + 1 = 19.$$

Or consider these examples.

$$11111 = 2^4 + 2^3 + 2^2 + 2^1 + 2^0 = 16 + 8 + 4 + 2 + 1 = 31$$

$$10101 = 2^4 + 2^2 + 2^0 = 16 + 4 + 1 = 21$$

$$11000 = 2^4 + 2^3 = 16 + 8 = 24$$

$$00101 = 2^2 + 2^0 = 4 + 1 = 5$$

Note that it takes more bits than it did digits to represent the same numbers. In decimal we can write two digits for 31 while in binary it takes five bits 11111.

Another base that gets used with computers is base 16, also called *hexadecimal*. Because $16 = 2^4$, hexadecimal lets you collapse four bits into a single **hexit** (= **hexadecimal digit**). Because we need sixteen hexits, we need new symbols for the ones beyond 9. Standard practice is to use the letters a through f, so that

$$a = 10, b = 11, c = 12, d = 13, e = 14, f = 15.$$

For example, consider the two hexit number af . This number represents

$$10 \cdot 16^1 + 15 \cdot 16^0 = 160 + 15 = 175.$$

You might instead convert it to binary. This is easier to do, because you can simply replace each hexit with the four bits it is equivalent to. In this case, $a = 1010$ and $f = 1111$ so

$$af = 10101111.$$

Back to binary, when doing one-digit sums or products there's only a small number of possible cases. This allows you to produce simple rules. For example, when adding you will carry a 1 just in case both inputs are 1s. And for the value left in that column, it will be 0 if both inputs are the same and it will be 1 if the columns are different. If you're adding three numbers, you carry if at least two of them are 1s and the value in that column is 1 just in case an odd number of inputs are 1s. Rather than having a table of about a hundred one-digit sums memorized, you can simply apply these a few simple rules repeatedly. This comes with the trade-off mentioned earlier; more bits to represent the same number means you have to apply this simpler rule a larger number of times.

Indeed, what I've just described is essentially how the circuit called an **adder** works. The inputs are ons and offs representing two numbers in binary, and these rules are repeatedly applied to quickly calculate the sum. In this way you can make it so a computer's processor can compute sums. Other operations such as multiplication, division, and so on can be implemented using more complicated circuits.

Checkpoint 2.4.1 Compute the following binary sum: $1011 + 1101$.

2.5 Exercises

Exercises

1. An important fact we used over and over in the algorithms for multiplying is that multiplying by a power of ten means adding zeroes to the right of a number. Give an explanation for this fact. That is, explain why if d is a digit then $d \times 10$ can be written as $d0$. More generally, explain why $d \times 100$ can be written as $d00$ and $d \times 1000$ can be written as $d000$. What is the rule for $d \times 10^n$?
2. Do the following calculations using the column algorithms. Show all steps in your work.

$$5437 + 984, \quad 5437 - 984, \quad 145 + 201 + 810$$

3. Calculate the following product using both the box and columns methods. Show all steps in your work.

$$143 \times 219$$

4. When adding two numbers you might have to carry a 1 into the next digit, but you never have to carry a 2 or higher. Explain this phenomenon. If you add three numbers, what are the possible values you might have to carry to the next digit? What if you add four numbers? Explain.
5. When multiplying a number by a one-digit number, you often have to carry a value into the next digit. Unlike with adding two numbers where you only carried 1s, with multiplying you can also carry 2s, 3s, and so on. Is it ever possible that you have to carry so much that you to carry ahead to the next two digits, not just the next one digit? Explain your answer. If it is possible, give an example to demonstrate. If it is not possible, what is the largest amount you might have to carry to the next digit?
6. Describe a cube method for calculating the product of three two-digit numbers. Use your method to calculate $12 \times 23 \times 34$. Reflect on whether this is better than doing a two step multiplication (i.e. first multiply 12×23 and then multiply the result by 34).
Hint. Use a $2 \times 2 \times 2$ cube where each of the 8 spots corresponds to a multiplication of three one-digit numbers.
7. Consider the number 100 written in base 10. Rewrite it in bases 2, 3, and 16.
8. Modify the addition algorithm to calculate the base 3 sum

$$2102 + 1121.$$

Show all your work.

Hint. Instead of each step in the calculation being a one-digit decimal sum, it will be a one-digit base 3 sum.

9. Modify the subtraction algorithm to calculate the base 3 difference

$$2102 - 1121.$$

Show all your work.

Hint. Instead of each step in the calculation being a one-digit decimal difference, it will be a one-digit base 3 difference.

10. Modify the multiplication method—your choice of the box version of the column version—to do the binary multiplication

$$1010 \times 1110.$$

Show all your work.

Chapter 3

Number Theory

Blah blah this chapter continues our study of the natural numbers. Here we take on a new perspective, thinking about natural numbers as being built up by multiplication.

3.1 Prime numbers

If we think of natural numbers as being formed by multiplication, then the building blocks are those numbers we can't split up any further. These are what we call the prime numbers. For example, we can break up 6 as $6 = 3 \cdot 2$, but there's no way to break up 7 into a multiplication of two smaller numbers.

Definition 3.1.1 Divisibility. Let m and n be natural numbers. We say that m **divides** n if n is a multiple of m , that is if $n = ma$ for some natural number a . Symbolically we write this as $m \mid n$. This concept has many synonyms; you also hear m is a divisor of n , m is a factor of n , or n is a multiple of n . ♦

For example, 2 is a factor of 6 because $6 = 2 \cdot 3$.

A useful fact is that the divisibility relation \mid is what mathematicians call **transitive**: if $a \mid b$ and $b \mid c$ then $a \mid c$. For example, if you know that $10 \mid 1450$ then you can automatically conclude that $2 \mid 1450$, because $2 \mid 10$.

Definition 3.1.2 Primes and composites. A natural number is called **prime** if it has exactly two divisors, itself and 1. A natural number is called **composite** if it has more than two divisors. Equivalently, n is composite if it can be written as a product of two smaller numbers: $n = ab$ where $a, b < n$. ♦

A few comments are in order. First, observe that 1 is neither prime nor composite. This is intentional. One practical reason is that allowing 1 to be prime means we'd have to awkwardly exclude it in important results about prime numbers; see [Theorem 3.1.4](#) and the discussion afterward. Lurking behind this is the fact that multiplying by 1 doesn't change a number. If we allowed 1 as a basic building block of forming numbers by multiplication we'd always be able to add more 1 blocks. One way to think about this is, 1 is the value you get before you start doing any multiplying; we want to think of 1 as what you get by multiplying together 0 many numbers. Thus it doesn't make sense to consider 1 as a prime number.

Second, let's see why the two definitions of composite are equivalent. For one direction, think of a number like 6 with more than two divisors. Since it has itself and 1 and divisors, that means it has a divisor between 1 and itself. But if, say, 2 is a divisor of 6 that means you can split 6 into a product of

two smaller numbers: $6 = 2 \cdot 3$. For the other direction, if you can write, for example, $10 = 2 \cdot 5$ then that gives more divisors on top of 1 and 10, so 10 has more than two divisors.

Checkpoint 3.1.3 List the first ten prime numbers.

A **prime factorization** of a number is a complete factorization of it down to primes. For example, $30 = 6 \cdot 5$ is not a prime factorization, because 6 is not prime. But $30 = 2 \cdot 3 \cdot 5$ is a prime factorization, since you can't factor any further. An important result is every number has a prime factorization.

Theorem 3.1.4 The fundamental theorem of arithmetic. *Every positive natural number has a unique prime factorization.*

Note that this theorem says two things. First, it says that for any number there is some way to write a prime factorization of it. Second, it says that there is only one way to do so. To be more clear, you can write a prime factorization in different orders. For instance,

$$60 = 2^2 \cdot 3 \cdot 5 = 5 \cdot 2^2 \cdot 3 = 3 \cdot 5 \cdot 2^2.$$

Rearranging the order of the multiplication doesn't count as a different factorization; as we know, order doesn't matter for multiplication. Also, let me highlight the role of 1. If we had decided that 1 counted as a prime number then we would have many different prime factorizations for any number:

$$6 = 2 \cdot 3 = 1 \cdot 2 \cdot 3 = 1^2 \cdot 2 \cdot 3 = 1^{100} \cdot 2 \cdot 3.$$

So we would've had to phrase the result like, "every positive natural number has a unique prime factorization where we exclude 1s". Thus we see the practical utility in excluding 1 from the prime numbers.

Second, you might wonder whether 1 itself has a prime factorization. It does. Namely, 1 is what you get when you multiply zero primes together. If this bothers you, feel free to imagine the fundamental theorem of algebra saying that every number > 1 has a prime factorization.

To close out this section, let me mention another fundamental result about prime numbers. This theorem dates back at least as far as the ancient Greek mathematician Euclid (circa 300 BCE).

Theorem 3.1.5 The infinitude of primes. *There are infinitely many prime numbers. Phrase differently, for any natural number n there is a prime number $p > n$.*

A more recent result, from 1896, tells us more about where the prime numbers are. By Euclid, no matter how far out you go there's always more prime numbers. The **prime number theorem** says that they get more and more sparse; there's more prime numbers to be found but you will have to look further and further out. (The precise statement of the prime number theorem requires both logarithms and ideas from calculus, so let me leave you with this gloss.)

The fundamental theorem of arithmetic tells us that every number has a prime factorization. Euclid's theorem tells us that these factorizations aren't composed of just a few small prime numbers. In general, the prime factorization of a number can include primes at the same order of magnitude of the number itself. As such, you might not be surprised to hear that the problem of finding the prime factorization for a number is in general very hard. Indeed, it can be so hard that it is used as the basis of the **RSA cryptosystem**. Put crudely, if you've ever wondered how you can access your bank info over the internet without anyone stealing it, that is because it can be really hard to factor numbers.

Nonetheless, for some small numbers there are rules you can learn to determine divisibility. This is the topic we turn to in the next section.

3.2 Divisibility Rules

In this section we learn rules for checking divisibility by a small number. Let's start with the easiest and, for that reason, the least useful.

Theorem 3.2.1 Rule for divisibility by 1. *A natural number n is always divisible by 1.*

Proof. Because $n = 1 \cdot n$. ■

Let's see a more useful rule.

Theorem 3.2.2 Rule for divisibility by 10. *A natural number n is divisible by 10 if the last digit of n is 0.*

Proof. Every other digit of n represents 10 times something, so the only way n could fail to be a multiple of 10 is if the units digit is nonzero. ■

Knowing the rule for divisibility by 10 lets us also know rules for divisibility by 2 and 5, because $10 = 2 \cdot 5$.

Theorem 3.2.3 Rule for divisibility by 2. *A natural number n is divisible by 2 if and only if its last digit is even: 0, 2, 4, 6, or 8.*

Proof. The other digits of n represent a number divisible by 10, and hence also divisible by 2. So all that remains is to check that the quantity represented by the units digit is divisible by 2. Those are the five even numbers which are one digit. ■

Theorem 3.2.4 Rule for divisibility by 5. *A natural number n is divisible by 5 if and only if its last digit is either 0 or 5.*

Proof. Same logic as the rule for 2. The only possible issue is the units digit, so check if the units digit is one of the two one-digit numbers divisible by 5. ■

Before we see the rule for 3, let's look at an example. Consider the number 123. You can break this apart as follows.

$$\begin{aligned} 123 &= 1 \cdot 10^2 + 2 \cdot 10 + 3 \cdot 1 \\ &= 1 \cdot (9 + 1)^2 + 2 \cdot (9 + 1) + 3 \cdot 1 \end{aligned}$$

When you expand out the multiplications, everything with a 9 as a term is a multiple of 3. The remainder are $1 \cdot 1$, $2 \cdot 1$, and $3 \cdot 1$. So 123 is a multiple of 3 if and only if these remainders sum to a multiple of 3. But this is just looking at the sum of the digits.

Theorem 3.2.5 Rule for divisibility by 3. *A natural number n is divisible by 3 if and only if the sum of its digits is a multiple of 3. (To check if the sum of the digits is a multiple of 3, repeat this process until you get to a small enough number you don't need to check.)*

Proof. Do what we did with 123, except you have to write it abstractly to work for any number. ■

The same idea gives us a divisibility rule for 9. What we used in the breakdown for 3 was 9s, and those will also give divisibility by 9.

Theorem 3.2.6 Rule for divisibility by 9. *A natural number n is divisible by 9 if and only if the sum of its digits is a multiple of 9.*

Proof. Same as with 3. ■

We can get a rule for 6 by combining the rules for 2 and 3. This is because $6 = 2 \cdot 3$ and so checking divisibility by 6 is the same as checking divisibility by both 2 and 3.

Theorem 3.2.7 Rule for divisibility by 6. *A natural number n is divisible by 6 if and only if its last digit is even and the sum of its digits is a multiple of 3.*

Proof. Explained above. ■

The rules for 4 and 8 are maybe a little less convenient, but they are the best we can do.

Theorem 3.2.8 Rule for divisibility by 4. *A natural number n is divisible by 4 if and only if its last two digits form a number divisible by 4.*

Proof. The rest of the digits represent a quantity divisible by 100, and 100 is a multiple of 4. So the only thing to check is that the last two digits are divisible by 4. ■

Theorem 3.2.9 Rule for divisibility by 8. *A natural number n is divisible by 8 if and only if its last three digits form a number divisible by 8.*

Proof. Similar to the rule for 4. Namely, the rest of the digits represent a quantity divisible by 1000, and $1000 = 8 \cdot 125$. So the only thing to check is that the last three digits are divisible by 8. ■

There's too many two digit and three digit numbers to memorize which of them are divisible by 4 and 8. But at least checking a two or three digit number is easier than many digit number.

Conspicuously missing from this list is a divisibility rule for 7. This is not because there is no rule for 7, but because any rule for 7 is too complicated to be practically useful. You will discover this in an exercise for this chapter.

Putting these rules together, we have a method to get the prime factorization of a number n , provided its only prime factors are small. Namely, run through the rules. If you determine n is divisible by, say, 5, then divide it by 5. Keep going with this process until all the rules fail. At that point, you may have to manually check if you can divide by small primes like 7 or 11, until you recognize that what's left cannot be factored any further.

Checkpoint 3.2.10 Find the prime factorization of 1512.

Answer. $1512 = 2^3 \cdot 3^3 \cdot 7$.

Solution. Immediately you can see that 1512 is divisible by 2. You might further notice that it's divisible by 4, because $12 = 4 \cdot 3$. Divide out by $4 = 2^2$:

$$1512 = 2^2 \cdot 378.$$

This number is even so we can pull out another 2:

$$1512 = 2^3 \cdot 189.$$

We now see we cannot factor out any more 2s. Next we check divisibility by 3: $1 + 8 + 9 = 18$, and repeating the summing the digits rule we get $1 + 8 = 9$. This is a multiple of 9, so in fact we can pull out two copies of 3:

$$1512 = 2^3 \cdot 3^2 \cdot 21.$$

We repeat the rule for 3 with 21: $2 + 1 = 3$. Thus we can pull out yet another 3:

$$1512 = 2^3 \cdot 3^3 \cdot 7.$$

At this point, all that is left over is 7, which is small enough that we can recognize it as a prime. We can't go any further, so we have the final prime factorization.

Checkpoint 3.2.11 Find the prime factorization of 4455.

3.3 GCD and LCM

Let's talk about some other topics regarding divisibility. Given two integers, they will have some divisors in common. For example, consider 24 and 30. Their common divisors are 1, 2, 3, and 6. It can be useful to know about common divisors. This is used, for instance, in fraction arithmetic; see [Chapter 5](#).

Definition 3.3.1 Greatest common divisor (GCD). Given two numbers a, b , their **greatest common divisor** (GCD) is the largest number which divides both of them. It also makes sense to talk about the GCD of a list of more than two numbers; it is the largest number which divides all of them. We write $\gcd(a, b)$ for the GCD of a and b . \blacklozenge

Note that $\gcd(a, 0)$ is always a . This is because any number divides 0—after all, $0 \cdot n = 0$ for any n —and so the largest common divisor is the largest divisor of a , namely a itself. On the other hand, $\gcd(0, 0)$ is undefined; every number divides both 0 and 0, so there isn't a largest one.

One way to determine the GCD of a group of numbers is to write out all of their prime factorizations, look for the primes common to all, and then multiply them together.

Checkpoint 3.3.2 Find the GCD of 168 and 180.

Solution. First find the prime factorizations, by the methods of the last section.

$$\begin{aligned} 168 &= 2^3 \cdot 3 \cdot 7 \\ 180 &= 2^2 \cdot 3^2 \cdot 5 \end{aligned}$$

Looking at these factorizations, we see that they have in common 2^2 and 3. We multiply those together to get $\gcd(168, 180) = 12$.

This method works, but it ranges from tedious to practically impossible, because it first requires you to find the prime factorizations of all the numbers. This is reasonable for two digit numbers, but it quickly gets awful. Surely, one hopes, there is a better way. If you are like me, then in elementary school you were not taught a better way. This happened because math teachers hate their students and want them to suffer. You can break the cycle of violence, however. You can teach your own students to use the Euclidean algorithm.

Theorem 3.3.3 *If $a > b$ are positive numbers then $\gcd(a, b) = \gcd(b, a - b)$.*

Proof. Suppose d is a divisor of both a and b . That is, $a = md$ and $b = nd$ for some numbers m, n . Then $a - b = md - nd = (m - n)d$, so $d \mid a - b$. A similar argument shows that if d is a divisor of both b and $a - b$, then $d \mid a$. Together, these show that the common divisors of a and b are the same as the common divisors of b and $a - b$. Since they have the same set of common divisors, they have the same largest common divisor. \blacksquare

Definition 3.3.4 The Euclidean algorithm for finding the GCD, first version. To find the GCD of a and b , where we assume $a \geq b$, follow this process.

1. If $a = b$, then we are done. Return that value as the GCD.

2. If $a > b$, then instead compute $\gcd(b, a - b)$, using this same process.

While this algorithm may look circular, at each stage you have decreased one of the two numbers you are looking at. You can't keep decreasing numbers forever—you eventually hit bottom at 1—so this process will eventually stop. ♦

An algorithm like this, which reduces a problem to a smaller or easier case until it eventually bottoms out, is called a **recursive** algorithm.

Checkpoint 3.3.5 Find $\gcd(30, 18)$.

Solution. We recursively apply the Euclidean algorithm (first version).

$$\begin{aligned}\gcd(30, 18) &= \gcd(18, 30 - 18) = \gcd(18, 12) \\ &= \gcd(12, 18 - 12) = \gcd(12, 6) \\ &= \gcd(6, 12 - 6) = \gcd(6, 6) \\ &= 6\end{aligned}$$

This algorithm is fastest when one number is approximately half as big as the other. When one is much bigger than the other or they are close in size it can take a long time for the subtraction process to bottom out. There is a faster version of the Euclidean algorithm using a different operation from subtraction.

Definition 3.3.6 The mod operation. Let a and b be natural numbers. Then $a \bmod b$ is the remainder when a is divided by b . ♦

Mod is short for “modulo”, which is a Latinate word meaning “up to this specified difference”. This use comes from its use in phrases like “7 and 37 are equivalent modulo 10”, meaning that 7 and 37 are equivalent after you divide out by 10 and only look at the remainder. A slightly more informal use would be “3.5 and 8.5 are the same modulo the whole number part”. If you want your friends to think you're a nerd, you can sprinkle this word into your everyday vocabulary.

For an example, $39 \bmod 4 = 3$ because $39 = 4 \cdot 9 + 3$.

When calculating mods by hand, often the quickest way is to repeatedly subtract until you can see the remainder. For example, to compute $13 \bmod 7$ you can just do $13 - 7 = 6$. Or to compute $13 \bmod 4$ you repeatedly subtract 4 until you get a result smaller than 4:

$$\begin{aligned}13 - 4 &= 9 \\ 9 - 4 &= 5 \\ 5 - 4 &= 1,\end{aligned}$$

so $13 \bmod 4 = 1$.

This method is reasonable when there aren't many subtractions to do, but is clunky when calculating $a \bmod b$ for a much larger than b .

If you are using a calculator, a trick to calculate mods is do the division then multiply the fractional part by the divisor. For example, to calculate $13 \bmod 4$ you'd calculate $13/4 = 3.25$ and then you'd multiply $0.25 \times 4 = 1$ to get the remainder. Some calculators and most programming languages will have a mod function to calculate it directly, in which case you can just use that.

Checkpoint 3.3.7 Explain why $a \bmod 10$ is always the units digit of a . How do you quickly see $a \bmod 100$?

Theorem 3.3.8 If $a > b$ are positive numbers then $\gcd(a, b) = \gcd(b, a \bmod b)$.

Proof. The idea is much the same as before. Suppose that d is a divisor of both a and b . That is, $a = md$ and $b = nd$ for some numbers m, n . Doing division with remainder, $a = qb + r$, where q is the quotient and the remainder $r = a \bmod b$. Then,

$$\begin{aligned} r &= a - qb \\ &= md - qnd \\ &= (m - qn)d, \end{aligned}$$

showing that $d \mid r$. A similar argument shows that if d is a divisor of both b and $a \bmod b$ then $d \mid a$.

Like before, this shows that the common divisors of a and b are the same as the common divisors of b and $a \bmod b$, and so they have the same greatest common divisor. ■

Definition 3.3.9 The Euclidean algorithm for finding the GCD, fast version. To find the GCD of a and b , where we assume $a \geq b$, follow this process.

1. If $b = 0$, then we are done. Return $a = \gcd(a, 0)$ as the GCD.
2. If $a > b$, then instead compute $\gcd(b, a \bmod b)$, using this same process.

Like before, this recursive process eventually bottoms out because we're dealing with smaller numbers at each step. ◆

Repeatedly computing $a \bmod b$ instead of $a - b$ will give you a much faster decreasing sequence, and so this version of the algorithm will take fewer steps to bottom out. The trade-off is that computing mod is usually slower than computing subtraction. You can use a calculator to help you.

Checkpoint 3.3.10 Find the GCD of 168 and 180.

Solution. We repeatedly apply the Euclidean algorithm (fast version).

$$\begin{aligned} \gcd(180, 168) &= \gcd(180, 180 \bmod 168) = \gcd(168, 12) \\ &= \gcd(12, 168 \bmod 12) = \gcd(12, 0) \\ &= 12 \end{aligned}$$

Going in the opposite direction, given two integers you can talk about their common multiples. For instance, the common multiples of 4 and 6 are

$$12, 24, 36, \dots$$

Just like two numbers have a largest common divisor, they also have a smallest common multiple.

Definition 3.3.11 Least common multiple (LCM) lcm. Given two numbers a, b their **least common multiple**, written $\text{lcm}(a, b)$ is the smallest number which is a multiple of both of them. In symbols, $\text{lcm}(a, b)$ is the smallest number ℓ so that $a \mid \ell$ and $b \mid \ell$. ◆

Theorem 3.3.12 *The least common multiple can be calculated as follows:*

$$\text{lcm}(a, b) = \frac{ab}{\gcd(a, b)}.$$

Example 3.3.13 The LCM of 8 and 10 can be calculated as

$$\text{lcm}(8, 10) = \frac{8 \cdot 10}{\gcd(8, 10)} = \frac{80}{2} = 40,$$

where you can use any of the previously described methods for finding the GCD. In this example, the numbers are small enough that you might immediately see that their GCD is 2. (And if you didn't immediately see that then you can look at their prime factorizations or use the Euclidean algorithm.) ▲

An alternative way to compute the LCM goes through prime factorizations. The smallest common multiple of two numbers must contain the prime factors of both of them, but without anything else. So to find the LCM list out the numbers' prime factorizations. Then grab all the primes which occur in at least one of the two factorizations, with the largest exponent that occurs.

Example 3.3.14 To calculate the LCM of 84 and 120, first find their prime factorizations.

$$\begin{aligned}120 &= 2^3 \cdot 3 \cdot 5 \\84 &= 2^2 \cdot 3 \cdot 7\end{aligned}$$

We now go through the factorizations, grabbing the largest exponent which occurs on one side or the other. Thus we get

$$\text{lcm}(84, 120) = 2^3 \cdot 3 \cdot 5 \cdot 7 = 840.$$

This method also can be used to explain why the division by the GCD method works. Namely, in the formula ▲

$$\text{lcm}(a, b) = \frac{ab}{\text{gcd}(a, b)}$$

the numerator ab says to take the prime factorization of a and add in the prime factorization of b . But some primes are added in twice, namely the ones that are common to both factorization. So to throw out the excess we divide by the common prime factors, which is the same as dividing by the GCD.

Checkpoint 3.3.15 What is $\text{lcm}(5, 0)$? What about $\text{lcm}(a, 0)$ for any positive number a ?

3.4 Exercises

Exercises

1. Come up with rules for divisibility by 100, 1000, and 10,000. Explain why your rule works. Can you give a general rule for divisibility by 10^n , for any number n of zeroes?
2. Some people think it is odd that 2 is the only even prime. Analogously to the definition of even, say that a number is **threeven** if it is a multiple of three. How many threeven primes are there? What are they? Explain. Do the same for **fiveven** numbers (those which are multiples of 5) and **sevenven** numbers (multiples of 7). Reflect on whether it is odd that 2 is the only even prime.
3. Explain why the number 0 is even.
4. Go to the [wikipedia page for divisibility rules](#) and look at the rules for divisibility by 7. Use one of them (your choice!) to determine whether 679 is a multiple of 7. Do you think this is a practical calculation to do by hand? Reflect.

5. Compute the prime factorization for 18900. It is okay to use a calculator for division, but show your work for each factor.
6. A reason for the use of 360 for the number of degrees in a circle is that 360 is highly divisible with many numbers dividing it. List all the divisors of 360.
Hint. If you know the prime factorization of 360, each divisor corresponds to a choice of primes from the factorization.
7. Find the GCD of 45 and 120 using two different methods. Which method do you find nicer to use? Reflect.
8. You can talk about the GCD of more than two numbers. For example, the GCD of three numbers a , b , and c is the largest number which divides all of them. Use the prime factorization method to find $\gcd(12, 30, 54)$.
Hint. The prime factorization method to find the GCD of two numbers had you look for common factors in their prime factorizations. How do you adapt that to find the GCD of three numbers?
9. Find the LCM of 12 and 15 using two different methods. Which method do you prefer? Reflect.
10. You can talk about the LCM of more than two numbers. For example, the LCM of three numbers a , b , and c is the smallest number which is a multiple of all of them. Use the prime factorization method to find $\text{lcm}(4, 10, 18)$.
Hint. The prime factorization method to find the LCM of two numbers had you look at their prime factorizations. How do you adapt that to find the LCM of three numbers?

Chapter 4

The Integers

We've spent the last three chapters thoroughly understanding the natural numbers. While these numbers are great, they don't do everything we want. There are some mathematical problems you cannot solve if you live only in \mathbb{N} . In this chapter and the next two we will learn about extensions of the natural numbers, adding in new numbers to allow us to solve more problems.

The first new number system we will study is called the integers. The name comes from a Latin adjective meaning "whole" or "complete", which is not helpful for explaining how they differ from the natural numbers.

4.1 Some motivation

When we talked about arithmetic in the natural numbers, subtraction was how you counted a taking away. For example, $7 - 3 = 4$ means that if you have seven oranges and I take three away then you have four remaining. A consequence of this is that subtraction doesn't make sense if the amount you are taking away is larger. The formula $3 - 7$ doesn't make sense in \mathbb{N} because if you have only three oranges you don't have enough for me to take away seven of them.

For many uses in mathematics, this limitation is inconvenient. We want to be able to subtract any two numbers without having to know in advance which one is larger. One place this shows up is in solving equations. Consider the equation

$$3 - x = 1.$$

One way to solve it would be to subtract 3 from each side. But if we only live in \mathbb{N} that isn't allowed, even though this equation's solution is a natural number, namely $x = 2$. Further, there are equations which don't have a solution at all if you only are looking at natural numbers. For instance,

$$3 + x = 1$$

has no solution in \mathbb{N} . (One way to see this: in the natural numbers, adding can never make a value smaller, so nothing you add to 3 can make it smaller to become 1.)

If we want to be able to solve equations like $3 + x = 1$ we need to introduce new numbers. As part of this, we need to explain the meaning of these new numbers and make sure that our familiar rules of arithmetic and algebra continue to work in these new numbers. These are the tasks to which we now turn.

4.2 Defining the integers

No doubt you know the solution already—throw in negative numbers!—but let's pretend that we don't already know about negative numbers. How might we come to invent them?

Our original definition of subtraction in \mathbb{N} , namely counting what happens when you take away from a collection, won't work. So we need a different way of thinking about subtraction to guide us. Another way we've understood subtraction is, it is the opposite of addition. For example, the formula $8 - 3$ can be understood as asking, what do you need to add to 3 to get a sum of 8? This way of thinking can be generalized. A good starting place is subtracting from 0, the additive identity. The equation $x = 0 - 4$ means, x is the value you add to 4 to get 0. This is saying that x is a sort of opposite of 4. It turns out that this one idea is enough to completely solve our subtraction problem.

Definition 4.2.1 Additive inverses. If x is a number then the **additive inverse** of x , written $-x$, is the number so that

$$x + (-x) = 0.$$

◆

Inverse means opposite. So additive inverse means opposite in the sense of addition. The following theorem makes precise the idea that subtraction is the opposite of addition.

Theorem 4.2.2 For any two numbers x and y , subtracting is the same as adding an inverse:

$$x - y = x + (-y).$$

Thus, for any number x we have

$$0 - x = -x.$$

Definition 4.2.3 Positive and negative numbers. A **positive number** is one which is > 0 . A **negative number** is an additive inverse of a positive number. The number 0 is considered neither positive nor negative. The **sign** of a number refers to whether it is positive or negative. ◆

Definition 4.2.4 Absolute value. The **absolute value**, also called **magnitude** and written $|x|$, of a number x is its distance from 0. More precisely, if $x \geq 0$ then $|x| = x$ and if x is negative then $|x| = -x$. ◆

Warning! While we write negative numbers by putting a minus sign in front, such as -8 , if you put a minus sign in front of a variable the result could be either negative or positive. For example, $-x$ will be negative if x is positive but it will be positive if x is negative. In general you won't know which.

Definition 4.2.5 The integers \mathbb{Z} . The **integers** consist of the natural numbers plus including additive inverses for every natural number. That is, the integers consist of both the positive and negative whole numbers. We write \mathbb{Z} for the set of integers. ('Z' stands for the German word *Zahlen*, meaning integers. Mathematics is a cosmopolitan discipline so some symbols come from non-English words.) ◆

Note that while we have now said what the integers are, we haven't yet said how to make sense of arithmetic with integers. It will turn out that all of your favorite arithmetic properties about the natural numbers will continue to be true about the integers, plus the new property that additive inverses always

exist. For convenience they are listed below, and afterward we will talk about how to make sense of arithmetic in \mathbb{Z} .

Theorem 4.2.6 Arithmetic properties of integers. *All of the following are true for any integers x, y, z .*

Additive identity	$x + 0 = x$
Additive inverses	x has an additive inverse.
Associativity of addition	$x + (y + z) = (x + y) + z$
Commutativity of addition	$x + y = y + x$
Multiplicative identity	$x \cdot 1 = x$
Multiplicative annihilation	$x \cdot 0 = 0$
Associativity of multiplication	$x \cdot (y \cdot z) = (x \cdot y) \cdot z$
Commutativity of multiplication	$x \cdot y = y \cdot x$
Distributivity	$x \cdot (y + z) = xy + xz$

Mathematicians call these properties the **commutative ring axioms**. There are many interesting mathematical structures, called **commutative rings**, which have these same properties. We will see one example later in this chapter when we look at clocks.

4.3 The geometric meaning of integers

In the last section we talked about how we build the integers from the natural numbers, namely throwing in additive inverses. Now we talk about a geometric interpretation of \mathbb{Z} , which will give us a way to understand the meaning of addition and multiplication.

Think of \mathbb{Z} as a line of dots, stretching infinitely far in both directions. One dot represents 0. Moving right from it you count up in the positive numbers $1, 2, 3, \dots$ and moving left from it you count down in the negative numbers $-1, -2, -3, \dots$. Call this picture the **integer line**.

In this picture, absolute value represents the number of steps to get to a number starting from 0. For example, 7 and -7 have the same absolute value—namely 7—because they are the same distance from 0.

What this picture tells us is that positive versus negative is about direction. With the natural numbers, you could get bigger and bigger forever, but if you tried going the other way you'd eventually get stuck at 0. With the integers, we add in an opposite direction which you can also travel infinitely.

This direction also tells us how to define order $<$ for the integers.

Definition 4.3.1 Order $<$ on integers. For two integers x and y , we say $x < y$ if x is to the left of y on the number line. Put explicitly:

- Any negative number is less than any positive number.

- If x and y are both ≥ 0 then $x < y$ if and only if $|x| < |y|$.
- If x and y are both ≤ 0 then $x < y$ if and only if $|x| > |y|$.

As we did with natural numbers, we use \leq , $>$ and \geq . ◆

The use of left for the negative direction and right for the positive direction is merely a convention. If your native language is, like English, written left-to-right this probably makes sense. Going forward means going to the right, going backward means going to the left. If your native language is written right-to-left, such as Arabic, then this might seem wrong, and you might ask if you could draw it differently. Everything would work just as well. What is important is that the negative direction and the positive direction are opposites.

This picture and the idea of positive versus negative as directions tells us the meaning of addition. Consider $4 + 7 = 11$. Think of this as meaning you start at 4 and you take 7 steps to the right. (Right because that is the positive direction.) Then you would end at 11. Or consider $4 - 7$. This means starting at 4 and taking 7 steps to the left (the negative direction). Counting them off you see $4 - 7 = -3$.

In general, adding x means taking $|x|$ steps, where you go to the right if x is positive and to the left if x is negative. (Adding 0 is taking no steps, which doesn't change your position.)

You can think about in terms of a single starting and ending point, but it's also useful to think about what addition does to the entire line. No matter your starting point, adding -3 means taking 3 steps to the left. This is the entire line stepping 3 to the left.

Adding, for instance, 2 is the entire line stepping 2 to the right.

Mathematicians liking to have too many names for things, we call these **translations** or **shifts**. In summary, the geometric meaning is translation. Adding by x shifts the entire number line $|x|$ units, with the sign of x telling which direction—left for negative, right for positive.

Now let's think about multiplication. Consider $2 \cdot 3$. This takes the starting position at 2 and moves it 3 times as far away from 0, namely to 6. Or consider $-1 \cdot 3$. Like before, this moves the starting position 3 times as far away from 0, ending at -3 . Before this movement, our two starting points were only 3 steps apart, but at the end they were 9 steps apart. Multiplying stretched the number line so the points become farther apart.

In general, multiplying by a positive number stretches out the number line. If you multiply by the positive number x you stretch the line so points are x times farther apart. More formally, we call this a **scaling**.

How should we understand multiplication by a negative number? Let's focus on the simplest case, namely multiplication by -1 . Multiplying by -1 should be the same as putting a minus sign in front of the number; more formally, multiplying x by -1 should give you the additive inverse of x . What this amounts to is changing which side of the number line you are on. For example, $3 \cdot -1 = -3$, so it gets moved from the right side to the left, while $-3 \cdot -1 = 3$, so it gets moved from the left side to the right.

Altogether, what multiplication by -1 does is flip the number line, swapping the left and right sides. More formally, we call this a **reflection**. You should think of this as swapping the two directions.

Here's an algebraic way to convince yourself that multiplying by -1 should give the additive inverse. Consider the formula $x \cdot (1 + (-1))$. On the one hand, $1 + (-1) = 0$ and so the product will be 0. On the other hand, the distributivity

property lets us rewrite this as $x + (-1 \cdot x)$. Putting those together we get

$$x + (-1 \cdot x) = 0,$$

whence we get that $-1 \cdot x$ is the thing you add to x to get 0—in other words, its additive inverse.

Consider now multiplication by, say, -3 . Think of this multiplication as a two-step process: $-3 = -1 \cdot 3$. Geometrically, we reflect the number line and then scale by a factor of 3.

Multiplication by 0 is a special case. Anything multiplied by 0 becomes 0, so multiplying by 0 collapses the entire number line down to a point. This usually isn't interesting or useful.

To sum up this section, **addition** corresponds to **translation**, where the magnitude tells you how far to translate and the sign tells you what direction. **Multiplication** corresponds to **scaling** and **reflection**, where the magnitude tells you how much to scale by and the sign tells you whether to reflect.

4.4 Arithmetic with integers

Now that we understand the geometric meaning of the addition and multiplication, let's see how that provides us algorithms to compute arithmetic operations. We already know algorithms for adding, subtracting, multiplying, and dividing natural numbers, so our task will be to explain the extra steps added by needing to consider negative numbers.

Definition 4.4.1 Algorithm for adding integers. How you add integers x and y depends on their signs.

- If one or more of the numbers are 0, simply use that adding 0 doesn't change the value.
- If their signs are the same their sum will have the same sign, with the magnitude being the sum of the magnitudes. That is, if x and y are both positive then $x + y$ is positive, and if they are negative then $x + y$ is negative. In either case, $|x + y| = |x| + |y|$, where you can use your favorite way of adding natural numbers to calculate this sum of magnitudes.
- If x and y have opposite signs and equal magnitudes, then $x + y = 0$. This is because in this case x and y are additive inverses.
- If x and y have opposite signs and different magnitudes, then the sign of $x + y$ is the same as the sign of the one with larger magnitude. Let's suppose that x happens to name the one with larger magnitude. The magnitude of the sum is then found by subtraction: $|x + y| = |x| - |y|$.

◆

Definition 4.4.2 Algorithm for subtracting integers. Use that subtraction means addition by the additive inverse, so you can use the algorithm for adding. Subtracting a positive number is the same as adding a negative and subtracting a negative number is the same as adding a positive. ◆

Definition 4.4.3 Algorithm for multiplying integers. When you multiply integers x and y , the magnitude of the product is always the product of the magnitudes: $|xy| = |x| \cdot |y|$. To determine the sign of the product, use the following rule.

- If x and y have the same sign, then xy is positive.
- If x and y have opposite signs, then xy is negative.



The intuition for the rule of signs is, multiplying by -1 represents a reflection. So multiplying by two negatives is two reflections, which cancel out. In general if you are multiplying more than two numbers, to know the sign of the product you count the number of negatives. An even number of negatives being multiplied all cancel out to be positive, while an odd number of negatives being multiplied gives a negative product.

Checkpoint 4.4.4 If you multiply together many numbers, what is the rule for when the product is positive and when it is negative? Explain.

4.5 Clocks and modular arithmetic

Natural numbers let us count up infinitely forever. Integers extend that to let us count infinitely in two directions. But there are many important phenomena in our lives which are cyclic, not stretching forth forever. You might count forward days of the week, and when you reach the end of the week you start over again at the beginning. Or you might count forward hours, resetting back to 1 after you pass 12.

Example 4.5.1 If it is currently 7 what time will the clock show after 8 hours? One way to calculate this is to count up from 7, resetting to 1 after passing 12. This is the same as calculating $7 + 8 = 15$ then subtracting 12. Thus we see the time will be 3.

The counting up method is less practical if you ask what time the clock will show after 73 hours. What you want to do is, calculate $7 + 73 = 80$ and get rid of the extra copies of 12. That is, you want to know what the remainder is if you divide 80 by 12. Doing division with remainder, you can see that $80 = 6 \cdot 12 + 8$, and so the time will be 8. ▲

A minor inconvenience with the remainder method is what happens at 12. For example, if it is 7 then in 5 hours it will be $7 + 5 = 12$. But if we did the remainder method we'd get a remainder of 0 (because 12 is a multiple of itself). That is, using this method you have to know that a remainder of 0 means the time is 12, because we don't allow 0 as a time.

This inconvenience goes away if you allow 0 as a time. This is commonly done with 24 hour clocks. In this system, 0 is midnight and the hours count up throughout the day, until 23 : 59 is the minute before the next midnight. Whereas with 12 hour clocks each morning and evening is 12 hours and we use am versus pm to distinguish them.

The same methods for counting time with 12 hour clocks work with 24 hour clocks.

Checkpoint 4.5.2 Using a 24 hour clock, suppose it is 17. What time will it be after 11 hours? After 71 hours?

Solution. For the first question, we calculate $17 + 11 = 28$ and then calculate the remainder when we divide by 24. In this case, that is mostly easily seen by calculating $28 - 24 = 4$, so it will be 4.

For the second question, we calculate $17 + 71 = 88$ and then calculate the remainder when we divide by 24. A little work tells us $88 = 3 \cdot 24 + 16$, and so it will be 16.

In [Chapter 3](#) we introduced the mod notation to talk about remainders

after division, and this is a place where it is useful. As a reminder, $a \bmod b$ refers to the remainder when a is divided by b . For instance, we could succinctly write our previous calculation as $(17 + 71) \bmod 24 = 16$.

If the same process works for both 12 hour and 24 hour clocks, you might suspect it also works for 13 hour clocks, or 67 hour clocks, or any number of hours. You would be correct. Indeed, you can do this clock arithmetic not only with addition but also with subtraction and multiplication. This gives what mathematicians call **modular arithmetic**.

Definition 4.5.3 Modular arithmetic. Fix a positive integer n . Then **arithmetic modulo n** is the following number system.

- The numbers are the integers $0, 1, \dots, n - 1$. That is, the numbers in arithmetic modulo n are the natural numbers $< n$.
- All arithmetic operations are done $\bmod n$, so that the result is between 0 and $n - 1$. You add and take the sum $\bmod n$, subtract and take the difference $\bmod n$, multiply and take the product $\bmod n$.

To differentiate modular arithmetic from ordinary arithmetic, we write e.g.

$$7 \cdot 4 + 1 \equiv 9 \pmod{10}.$$

The use of \equiv instead of $=$ reminds us that $7 + 4 \cdot 1 = 29$ and 9 aren't equal as integers, but are only the same modulo 10. \blacklozenge

A 12 hour clock is the same as arithmetic modulo 12 (except that you have to keep in mind we say 12 where mathematicians would say 0) and a 24 hour clock is the same as arithmetic modulo 24.

Example 4.5.4 Arithmetic modulo 2 is the arithmetic of **parity**—even versus odd. Even means 0 modulo 2 and odd means 1 modulo 2. You can turn questions about parity into calculations modulo 2. For example, if you want to know the parity of the product of an even and odd number you can compute $0 \cdot 1 \equiv 0 \pmod{2}$ and conclude that the product must be even. \blacktriangle

Checkpoint 4.5.5 If you add the product and sum of two odd numbers, is the result even or odd?

Solution. Do arithmetic mod 2:

$$1 \cdot 1 + (1 + 1) \equiv 3 \equiv 1 \pmod{2},$$

so the result must be odd.

Example 4.5.6 Arithmetic modulo 1 is boring. The only number is 0 and the result of any calculation is that one number 0. \blacktriangle

A nice fact about modular arithmetic is that all the usual algebraic rules for integers also apply. So you can do all the same calculations with all the same rules, just remembering that you have to do a mod at the end.

There is an exception here, namely that order doesn't make sense in modular arithmetic. In the integers, it makes sense to say that $5 < 7$. But this doesn't make sense for the integers modulo n . Think in terms of clocks: 5 is before 7, but it's also after 7 (the 7 from the previous cycle). You can say that 5 this morning is before 7 this morning, but you can't in general say 5 is before 7.

But so long as you just work with the arithmetic operations you can freely use the familiar rules.

Theorem 4.5.7 *The integers modulo n form a commutative ring. That is, they satisfy all the axioms in [Theorem 4.2.6](#).*

One use of modular arithmetic is in **cryptology**, the practice of encrypting data to ensure secure communication without eavesdropping. In brief, the way it works is your data, such as bank info, is turned into a number. Some calculations are done modulo a really big n with this number and a secret key. This produces a **ciphertext**, which is then transmitted over the internet. Knowing the secret key you can reverse the calculations to recover the data. But without that knowledge it's very difficult and slow to reverse the calculations. So someone spying on your internet traffic won't be able to figure out your bank info.

4.6 Exercises

Exercises

1. Do both subtractions $159 - 374$ and $374 - 159$ by hand, showing all work. What do you notice about the results?

2. Draw transformations of the integer number line which represent the three-step process given by the formula $-(2x+3)$. You should have four number lines in total, the starting one and then one for each step of the process.

Hint. That formula represents the following three-step process applied to an input integer x .

- (a) Multiply by 2.
- (b) Then add 3.
- (c) Finally multiply by -1 .

3. Draw transformations of the integer number line which represent the two-step process given by the formula $-2x-3$. (You should have three number lines in total.) Compare your final number line to the one from the previous problem. Explain what you see.

Hint. That formula represents the following two-step process applied to an input integer x .

- (a) Multiply by -2 .
- (b) Then add -3 .

4. The **triangle inequality** is the inequality $|x+y| \leq |x|+|y|$, which is true for any numbers x and y . Confirm the instance of the triangle inequality for the following pairs of numbers x and y . Also place the values x , y , and $x+y$ on a number line, and show on your picture the distances that correspond to $|x|$, $|y|$, and $|x+y|$.

- $x = 2$ and $y = 5$,
- $x = -2$ and $y = -5$,
- $x = -2$ and $y = 5$.

5. When we talked about natural numbers, we defined exponentiation as repeated multiplication:

$$a^n = \underbrace{a \cdot a \cdot \cdots \cdot a}_{n \text{ many}}$$

The same definition makes sense when a is any integer, i.e. possibly nega-

tive. Note however, that n must be ≥ 0 . (We can make sense of negative exponents using fractions, but that will come later.) Give a rule for when a^n is positive versus negative.

Hint. Your rule should break into two cases when a is positive or negative. Handle what happens in each case separately.

6. A **square root** is the opposite of squaring: y is a square root of x means $y^2 = x$. For example, 2 is a square root of 4 because $2^2 = 4$. Explain why every positive number has exactly two square roots, 0 has exactly one square root, and every negative number has no square roots.
7. With clocks it makes sense to add or subtract—what time will it be after three hours? what time was it three hours ago?—but multiplication doesn't quite make sense. What would three o'clock times five o'clock even mean? Nonetheless you could try to make sense of it by doing multiplication modulo 12. Do so to try to make sense of the following>
 - three o'clock times five o'clock,
 - three o'clock times seven o'clock,
 - three o'clock times eleven o'clock.

Can you give another example where you can do a mathematical calculation but it doesn't correspond to anything in reality?

8. Recall that in the integers every number has an additive inverse—a number you add to it to get a sum of 0. (This was one of the axioms of commutative rings.) This is also true in the integers modulo n . Check this in the case of the integers modulo 12 by finding the additive inverses for each number. That is, for each $a < 12$ find the number $b < 12$ so that $a + b \equiv 0 \pmod{12}$. Can you give a general rule for what $-a$, the additive inverse of a , is in the integers modulo 12?
9. A fun fact about the integers modulo a prime p is that every nonzero number has a **multiplicative inverse**. That is, for any $0 < a < p$ there is $0 < b < p$ so that $ab \equiv 1 \pmod{p}$. Check this fact for the integers modulo 5 by finding, for each nonzero number < 5 the number that multiply to get $1 \pmod{5}$.
10. Another fun fact about modular arithmetic is that if $n > 1$ is composite then there are **zero divisors**—nonzero numbers which multiply to have a product of 0. That is, if n is composite then there are nonzero $a, b < n$ so that $ab \equiv 0 \pmod{n}$. Check this fact for the integers modulo 12 by finding two numbers that multiply to get $0 \pmod{12}$. How many pairs of zero divisors can you find?

Chapter 5

The Rational Numbers

This chapter is about the class of numbers known as the rational numbers. These are those numbers which can be formed as fractions of integers. Along the way we will study the arithmetic of fractions and how to do computations involving rational numbers.

5.1 Motivating the rational numbers

The motivation is much the same as with \mathbb{Z} . With the integers, we wanted to always be able to subtract. Equivalently, we wanted to be able to always solve equations like $x + a = b$. Our solution was to add in negative numbers—in jargon, additive inverses. Now we want to always be able to divide. That is, we want to always be able to solve equation like $ax = b$. This presents a little extra complication.

To illustrate the complication, let's try the analogous thing we did for subtraction. Let's add in **multiplicative inverses**. For a number x , its multiplicative inverse, also called a **reciprocal** and written $\frac{1}{x}$, is the number you multiply it by to get 1, the additive identity. As an equation: $x \cdot \frac{1}{x} = 1$. If you take the integers and throw in all multiplicative inverses you still don't have enough numbers. To see this, consider the equation $2x = 3$. You should know using algebra that the solution should be $x = \frac{2}{3}$. But $\frac{2}{3}$ isn't the reciprocal of any whole number; indeed, it's the reciprocal of $\frac{3}{2}$.

What this example shows is that we need to add in more than just multiplicative inverses. Another complication is, there is an integer which cannot have an additive inverse, namely 0. One of our basic rules of arithmetic is that $0 \cdot x = 0$ for any number x . So adding a number which multiplies with 0 to produce 1 would render that rule void. The solution is simple: we don't want 0 to have a reciprocal.

These two obstacles are the only ones to overcome. With that in mind, let's now see the formal definition of the rational numbers.

5.2 The definition of the rational numbers

The new numbers we add in are to come from dividing, so we need to introduce a new notation to talk about division. A **fraction** $\frac{n}{d}$ is a shorthand for division. It represents the value of the **numerator** n divided by the **denominator** d .

A fraction might result in a number which we have another way to write. For example, the fraction $\frac{6}{2}$ is shorthand for the division of 6 by 2, which you

know gives the number 3. But in general this won't happen. In effect, what we are doing is giving names to these new numbers formed from division, and what better name than to directly describe how we created them?

We will see later how we can use these new names to do arithmetic. For now, let's officially define the rational numbers.

Definition 5.2.1 **The rational numbers \mathbb{Q} .** The **rational numbers** are those numbers which can be formed as fractions of integers. More precisely, a rational number is one which can be written in the form $\frac{n}{d}$ where n and d are integers, with $d \neq 0$. We write \mathbb{Q} to refer to the set of all rational numbers. ('Q' stands for "Quotient", the result of division.) \blacklozenge

Warning 5.2.2 You can make fractions with any kind of number as the numerator and denominator. For example, you could write the fraction $\frac{\pi}{2}$ to represent the result of dividing π by 2. However, only fractions of integers count for being rational numbers. It is a difficult mathematical result that π , and hence also $\frac{\pi}{2}$, cannot be written as a fraction of integers.

As a first step, let's see that this is an expansion of the integers.

Theorem 5.2.3 *Every integer is a rational number. In the language of sets, $\mathbb{Z} \subseteq \mathbb{Q}$.*

Proof. Because the integer n can be written as the fraction $\frac{n}{1}$. \blacksquare

With integers, there was a unique name for each new number—just take a positive whole number and slap a minus sign in front of it. With rational numbers that's no longer the case. For example, you should be able to convince yourself that $\frac{3}{1}$ and $\frac{6}{2}$ name the same number (namely 3). Our first task is to explain when two fractions name the same number. That is, we need to talk about equality.

Because fractions represent division, two fractions should name the same number if the divisions result in the same value. How can we check that? Let's cheat a little by remembering what we already know about fraction arithmetic. Consider an equality like

$$\frac{2}{4} = \frac{3}{6}.$$

You can rewrite this equation by moving both denominators to the top. That is, you can multiply both sides by both 4 and 6 to get an equivalent equation with no division

$$2 \cdot 6 = 4 \cdot 3.$$

And you can check these really are equal by doing the multiplication. In general, this **cross multiplying** trick works for any equality of fractions:

$$\begin{aligned} \frac{n}{d} &= \frac{m}{c} \\ nc &= md. \end{aligned}$$

What this means is, checking that two fractions name the same number can be reduced to integer multiplication, something we already understand.

Definition 5.2.4 **Fraction equality.** Two fractions $\frac{n}{d}$ and $\frac{m}{c}$ are **equal**, written $\frac{n}{d} = \frac{m}{c}$, if $nc = md$. \blacklozenge

I just gave you a definition of equality of fractions, but why should you trust me? How could you check that this is a reasonable definition?

One thing you could do is check that this has the correct basic properties equality should have. Here's three important properties.

Theorem 5.2.5 *Equality of fractions has the following properties.*

Reflexivity $\frac{n}{d} = \frac{n}{d}$ for any fraction.

Symmetry If $\frac{n}{d} = \frac{m}{c}$ then $\frac{m}{c} = \frac{n}{d}$ for any fractions.

Transitivity If $\frac{n}{d} = \frac{m}{c}$ and $\frac{m}{c} = \frac{\ell}{b}$ then $\frac{n}{d} = \frac{\ell}{b}$ for any fractions.

The reason fraction equality has these properties is because it's defined using equality of integers, which already has those properties. For example, if you unroll the definition then Symmetry states that if $nc = md$ then $md = nc$.

A final important property, **Substitution**, says that if you replace a fraction with an equal fraction in a formula then you get an equal result. This amounts to saying that when computing the arithmetic operations you get the same answer if you use an equal fraction.

That's the segue into the next topic, how to do fraction arithmetic.

5.3 Arithmetic of fractions

In this section we talk about how to perform arithmetic with fractions. As part of this, we will see the important fact that if you add, subtract, multiply, or divide two rational numbers then you get another rational number. This tells us that if we want to solve equations with rational numbers then we don't need to introduce yet more new numbers.

We will first talk about how to multiply and divide fractions. Adding and subtracting will come later, as those are in fact more difficult.

Definition 5.3.1 Algorithm for fraction multiplication. Consider two fractions $\frac{n}{d}$ and $\frac{m}{c}$. Then their product is

$$\frac{n}{d} \cdot \frac{m}{c} = \frac{nm}{dc}.$$

That is, to multiply fractions you multiply their numerators to get the product's numerator and you multiply their denominators to get the product's denominator. ◆

Here's a way to think about this definition. When you multiply, you multiply the numerators, but why do you also multiply the denominators? Fractions, as a key point of this chapter that has been repeatedly emphasized, are a way to denote that division is happening. What you're asking is, if you divide twice what is the total amount by which you are dividing? The answer is, your total division is by the product of the two.

Because exponentiation is repeated multiplication, this definition also tells you how to compute exponents of fractions. Namely,

$$\left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}.$$

The idea is, you multiply $\frac{a}{b}$ by itself n many times. So that multiplication becomes multiplying a by itself in the numerator and b by itself in the denominator.

Next we discuss division. To understand it, let's first understand what it has to do with multiplicative inverses. Consider for example the division $6/3 = 2$. This is equivalent to the multiplication $6 \cdot \frac{1}{3} = 2$. In general, division is the same as multiplication by the inverse.

Definition 5.3.2 Reciprocal of a fraction. The **reciprocal** or **multiplicative inverse** of a fraction $\frac{n}{d}$ is the fraction $\frac{d}{n}$. That is, to take the reciprocal of a fraction swap the numerator and denominator. ◆

You can check that $\frac{d}{n}$ is the multiplicative inverse of $\frac{n}{d}$ by multiplying them:

$$\frac{n}{d} \cdot \frac{d}{n} = \frac{nd}{dn} = 1.$$

Definition 5.3.3 Algorithm for fraction division. Consider two fractions $\frac{n}{d}$ and $\frac{m}{c}$. Then their quotient is

$$\frac{n/d}{m/c} = \frac{n}{d} \cdot \frac{c}{m} = \frac{nc}{dm}.$$

That is, to divide fractions you multiply by the reciprocal, using the rule for how to multiply fractions. ◆

Writing division of fractions can be annoying, since it involves fractions inside fractions! Good practice is to write the outer fraction in $\frac{N}{D}$ form and the inner fractions in n/d form. If you write them both the same way it can be hard to tell the order. For example, which division happens first in

$$\frac{\frac{1}{2}}{3} \quad \text{or} \quad 1/2/3?$$

Example 5.3.4 To calculate

$$\frac{3/4}{2/5}$$

we turn it into a multiplication, namely the multiplication of the numerator by the reciprocal of the denominator:

$$\begin{aligned} \frac{3}{4} \cdot \frac{5}{2} &= \frac{3 \cdot 5}{4 \cdot 2} \\ &= \frac{15}{8}. \end{aligned}$$

Let's now turn to addition (and its sister subtraction). When the denominators are the same, the distributivity rule tells us what to do. Consider the following example. ▲

$$\begin{aligned} \frac{2}{3} + \frac{5}{3} &= 2 \cdot \frac{1}{3} + 5 \cdot \frac{1}{3} \\ &= (2 + 5) \cdot \frac{1}{3} \\ &= \frac{2 + 5}{3}. \end{aligned}$$

In general, if two fractions have the same denominator then their sum is obtained by adding their numerators and keeping the denominator. (And the rule for subtraction is similar.)

$$\frac{n}{d} + \frac{m}{d} = \frac{n + m}{d}.$$

Checkpoint 5.3.5 Write the rule for subtracting fractions with the same denominator.

But you shouldn't expect to only add fractions which already have the same denominator. In this case, you have to rewrite the fractions to get a common

denominator. Note that if you have any fraction multiplying the top and bottom by the same value gives an equal fraction. For example, $\frac{2}{3} = \frac{2}{3} \cdot \frac{4}{4} = \frac{8}{12}$. If we want to rewrite two fractions to have the same denominator we can do this process for each.

Definition 5.3.6 Rewriting fractions to have a common denominator. Suppose you have two fractions $\frac{n}{d}$ and $\frac{m}{c}$. Here are a few different ways you can rewrite them to have a common denominator.

1. If one denominator is a multiple of the other, say $d = kc$, then d is a common denominator. In this case, keep $\frac{n}{d}$ the same and multiply the other fraction by $\frac{k}{k}$.

$$\frac{n}{d} \text{ and } \frac{m}{c} \cdot \frac{k}{k} = \frac{mk}{d}.$$

2. The smallest common denominator is always the least common multiple $\ell = \text{lcm}(d, c)$ of the denominators, which you can find by the methods of [Chapter 3](#). You then have to determine which you multiply each denominator by to get this least common multiple. Give these names: $\ell = yd$ and $\ell = xc$. These are what you multiply the fractions by.

$$\frac{n}{d} \cdot \frac{y}{y} = \frac{ny}{\ell} \text{ and } \frac{m}{c} \cdot \frac{x}{x} = \frac{mx}{\ell}.$$

3. If you don't want to find the least common multiple you can always take the product of the two denominators as a common denominator.

$$\frac{n}{d} \cdot \frac{c}{c} = \frac{nc}{dc} \text{ and } \frac{m}{c} \cdot \frac{d}{d} = \frac{md}{dc}.$$

◆

The different methods have different advantages. The first requires the least work but it doesn't always apply. The second gives you as small a common denominator as possible, but it might be a bit of work to find the least common multiple. The last method avoids that work but has the downside that your common denominator may be much larger than necessary.

Also useful is the opposite direction, namely simplifying a fraction by canceling out common terms from the numerator and denominator. After performing an addition you may have a fraction that can be written more simply, and it's good to know how to do that.

Definition 5.3.7 Reduced form of a fraction. A fraction $\frac{n}{d}$ is in **reduced form** if there are no common factors of the numerator and denominator, that is if $\text{gcd}(n, d) = 1$. If a fraction is not already in reduced form, you can turn it into reduced form by dividing this GCD out from both parts:

$$\frac{n}{d} = \frac{n/\text{gcd}(n, d)}{d/\text{gcd}(n, d)}.$$

Note that in this second fraction both parts are whole numbers.

◆

With these in hand, we can now see how to add/subtract any two fractions.

Definition 5.3.8 Algorithm for fraction addition and subtraction. Consider two fractions $\frac{n}{d}$ and $\frac{m}{c}$. To add (or subtract) them, first rewrite them to have a common denominator. Then add (or subtract) the two fractions with a common denominator. If you like, then simplify your result to reduced form.

◆

Having a common denominator also lets us compare fractions. First however, let's observe that you can always rewrite a fraction to have a positive denominator.

Definition 5.3.9 Rewriting a fraction to have a positive denominator. Consider a fraction $\frac{n}{d}$.

1. If d is already positive, there's nothing to be done.
2. If both n and d are negative, then the negatives cancel out. For example, $\frac{-2}{-3} = \frac{2}{3}$.
3. If n is positive but d is negative, then you can move the negative to the numerator. For example, $\frac{3}{-2} = \frac{-3}{2}$.

◆

If two fractions have a positive denominator, then to compare them you simply compare their numerators. The larger numerator gives the larger fraction. If the fractions don't have a common denominator, rewrite them to have a common denominator and then compare.

Definition 5.3.10 Algorithm for comparing fractions. Consider two fractions $\frac{n}{d}$ and $\frac{m}{c}$. Here's how you determine which one is larger.

1. First ensure both denominators are positive.
2. Then rewrite the fractions, if necessary, to have the same denominator.
3. Once you have a common denominator, compare the fractions by comparing the numerator. The one with the larger numerator is the larger fraction, and if the numerators are equal the fractions are equal.

◆

Note that this algorithm also gives you a way to check whether two fractions, probably written in different bases, are equal. First you rewrite them to have a common denominator, then they are equal if and only if the numerators are equal.

Now that we have an idea of how to compare fractions, let's think about what the number line of rational numbers looks like. Remember that the number line for integers consisted of equally spaced dots marching off to infinity in both directions.

With the rational numbers we are putting in more dots in the gaps. For instance, between any two adjacent integers a and $a + 1$ there are infinitely many rational numbers, such as

$$a + \frac{1}{2}, a + \frac{1}{3}, a + \frac{2}{3}, a + \frac{1}{4}, \dots$$

In general, between any two dots on the rational number line you can find another dot. If x and y are two rational numbers then their average $\frac{x+y}{2}$ sits between them. Since you can repeat this process indefinitely, taking more and more averages to get more and more intermediate points, this means between x and y are infinitely many dots. Thus, the rational number line looks like an infinitely dense line of dots.

Let's close out this section by mentioning an important algebraic fact about the rational numbers. With integers, adding, subtracting, or multiplying always gave you another integer, but division might take you outside the integers. This doesn't happen with rational numbers.

Theorem 5.3.11 Closure of the rational numbers under arithmetic operations. *The rational numbers are closed under addition, subtraction, multiplication, and division. That is, if you take two rational numbers then their sum, difference, product, and quotient can all be written as a single fraction of integers. (Of course we exclude the case of division by zero.)*

We will expand our number system again in the next chapter, but the reason for this expansion cannot be the basic arithmetic operations. It must be for something new.

5.4 Percents

To close out this chapter we look at a commonly used kind of fraction, namely percents. The meaning is in the name: percent = per cent = per hundred (think cent as in centimeter).

Definition 5.4.1 Percents. Let X be a number. Then $X\%$, pronounced X percent, is the fraction

$$\frac{X}{100}.$$



Percents are useful when talking about parts of a whole. For instance, when setting up direct deposit for your paycheck you can choose how to split it between multiple accounts. Putting 100% in one account would be putting the whole paycheck in one place. Or you might instead split it 30% and 70% between two accounts; the first account would receive $\frac{30}{100}$ of your paycheck while the second would receive $\frac{70}{100}$.

Checkpoint 5.4.2 What percent represents one quarter? One fifth? One tenth?

One way to remember the meaning of the percents is to look at the symbol %. The slash represents a fraction and the two 0's are the zeroes in 100.

One place you may see percents used to describe relative changes. A population of a city might increase by 10% over a decade. If the initial population was 10,000 people, then the extra 10% is $10,000 \cdot \frac{10}{100} = 1000$ people, so the new population is 11,000. You could equivalently calculate the new population by

$$10,000 \cdot 110\% = 10,000 \cdot \frac{110}{100} = 11,000,$$

where the 100% represents the initial population and the 10% is the increase.

In the other direction, percents also can describe decreases. Your car might depreciate in value by 5% in a year. If its initial value was 20,000 dollars then its new value is

$$20,000 \cdot \frac{95}{100} = 20,000 - 20,000 \cdot \frac{5}{100} = 19,000$$

dollars. Because the value is decreasing we subtract the 5% from the 100% representing the initial value.

There's nothing special about the number 100. It's used for percents because it's a round number that's familiar from our base 10 number system. You could use the same idea but pick a different number for the denominator. Though less commonly used than percents, some of these have been used enough to have names.

Definition 5.4.3 Permilles. Let X be a number. Then $X\%$, pronounced

X **permille**, is the fraction

$$\frac{X}{1000}.$$

Think: per mille = per thousand (mille as in milligram). ◆

For example, 100‰ is the fraction $\frac{100}{1000} = \frac{1}{10}$. It is equivalent to 10%.

You also don't need to use a power of ten. You could, say, break things up into sixtieths. If this seems weird to you, it's how we break up hours and minutes. Or you could break things up into 360ths, like how the degree measurement for angles breaks a circle up into 360ths of the whole.

These examples have special names and notation. If you wanted to be talk about it for any denominator, you arrive back at fractions. Indeed, you can use this as a way to think about what fractions are. We divide wholes into parts—per 100 for percents, per 60 for seconds, and so on. Fractions are a generalization of that idea to divide wholes into any number of parts.

The reason why 60 and 360 get used is that they are highly divisible. Many different integers evenly divide into them. On the other hand, percents have fewer divisors. For instance, you cannot express $\frac{1}{3}$ as a whole number fraction. (You can express it using decimals, but even then writing it exactly requires infinitely many digits. We will talk about these in [Chapter 6](#).) On the other hand, percents fit nicely with our base-ten numeral system.

5.5 Exercises

Exercises

1. Check an instance of the closure of the rational numbers by showing that $\frac{3}{4} + \frac{3}{7}$ can be written as a fraction of integers. Do the same for their product $\frac{3}{4} \cdot \frac{3}{7}$.
2. Check an instance of the closure of the rational numbers by showing that $\frac{2}{3} - \frac{5}{4}$ can be written as a fraction of integers. Do the same for their quotient $\frac{2/3}{5/4}$.
3. If William Shakespeare were a mathematician, he might have written the following couplet for *Ratio and Juliet*: “What’s in a name? That which we call a fraction / By any other name would smell as sweet.” Check that $\frac{3}{5}$ would smell just as sweet if we called it $\frac{72}{120}$ by showing that these two fractions name the same number.
4. A student simplifies $\frac{19}{95}$ by canceling out the 9s to get $\frac{1}{5}$. They then simplify $\frac{16}{64}$ by canceling out the 6s to get $\frac{1}{4}$. Show that these two calculations happen to get the correct answer. Then show that this isn't a valid method in general by showing that $\frac{12}{26} \neq \frac{1}{6}$.
5. Check that the average of $\frac{3}{4}$ and $\frac{5}{6}$ is between them.

Hint. First compute the average using the formula $\frac{x+y}{2}$. Then convert everything to a common denominator to check the order.

Now that we have fractions we can make sense of negative exponents. Recall that positive exponents were defined as repeated multiplication: a^n is the product of a with itself n many times. Negatives, being the opposite of positives, should in the exponent have the opposite meaning of multiplication. That is, negative exponents represent repeated division. If $-n$ is a negative number, then

$$a^{-n} = \frac{1}{a^n}.$$

6. Calculate 2^{-3} , $(\frac{1}{3})^{-2}$, and $(\frac{3}{2})^{-3}$.

Hint 1. For the last two, it can be useful to think of it as $a^{-n} = (\frac{1}{a})^n$. That is, think of it as taking a reciprocal then a positive power.

Hint 2. When taking powers of fractions, a helpful rule is

$$\left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}.$$

7. The usual rules for exponents still work given this definition of negative exponents. Check an instance of the rule $a^m \cdot a^n = a^{m+n}$ by showing that

$$\left(\frac{3}{5}\right)^{-3} \cdot \left(\frac{3}{5}\right)^2 = \left(\frac{3}{5}\right)^{-1}.$$

Show all steps of your calculations so that it is clear the calculations represented by each side really do give the same answer.

8. Check an instance of the rule $(a^m)^n$ by showing that

$$\left(\left(\frac{3}{2}\right)^{-2}\right)^{-2} = \left(\frac{3}{2}\right)^4.$$

Show all steps of your calculations so that it is clear the calculations represented by each side really do give the same answer.

9. Each of the following values is given as a fraction, a percent, or a permille. For each of them, rewrite it into the other two forms. Please fully simplify all fractions.

$$\frac{3}{5}, \quad 24\%, \quad 720\text{‰}$$

10. Each year, an investment increases in value from its current value by 10%. How much has it increased from its initial value after two years? If it is initially worth 10,000 dollars, how much is it worth after two years?

Hint. The answer is not 20%.

Chapter 6

The Real Numbers

This chapter is about the real numbers, an extension of the rational numbers used to measure continuous quantities. Before we officially arrive at these new numbers, however, we will build on the positional number system of [Chapter 2](#) to refer to non-whole numbers. This will provide us with a convenient calculational tool for many applications.

Briefly, the real numbers are those which can be written in decimal, possibly with infinitely many digits past the decimal point.

6.1 Writing fractions in decimal

Recall the idea of the Hindu–Arabic numeral system. The positions of the digits correspond to powers of 10, and the digit in that place tells you how many copies of that power of 10 to include. For example,

$$4,071 = 4 \times 10^3 + 0 \times 10^2 + 7 \times 10^1 + 1 \times 10^0.$$

To expand upon this idea to represent non-whole numbers, rather than just having positions which represent repeated multiplication by 10 we want to also have positions which represent repeated division by 10.

It is notationally convenient to use negative powers to represent repeated division. If $-n$ is a negative number then

$$b^{-n} = \frac{1}{b^n}$$

represents dividing by b n many times. Recall that the usual rules for exponents still work under this definition.

With this notation in hand, what we want is to expand our positional notation system to have positions not just for positive powers of 10 but also negative powers. The first thing you might try presents an issue. Suppose you want to write 713 to mean $7 \times 10^1 + 1 \times 10^0 + 3 \times 10^{-1}$. How do you know that you don't mean $7 \times 10^2 + 1 \times 10^1 + 3 \times 10^0$? Or $7 \times 10^0 + 1 \times 10^{-1} + 3 \times 10^{-2}$? When we only had positive powers there was no ambiguity—start with power 0 on the right. Now there is ambiguity.

The solution is simple: mark in some way where the zeroth power position is. The way we do this is via a **decimal point**, written as a period. Left of the decimal point is the positive power positions and right of the decimal point is the negative power positions. If only positive powers are included we don't write the decimal point. Thus we can distinguish between the following.

$$71.3 = 7 \times 10^1 + 1 \times 10^0 + 3 \times 10^{-1}$$

$$713 = 7 \times 10^2 + 1 \times 10^1 + 3 \times 10^0$$

$$7.13 = 7 \times 10^0 + 1 \times 10^{-1} + 3 \times 10^{-2}$$

Definition 6.1.1 Hindu–Arabic numeral system, with fractional part.

The Hindu–Arabic numeral system can be expanded to allow fractional parts as follows. Here, each d_i is a digit, a numeral 0 through 9. First, an example with a fixed number of digits.

$$d_2 d_1 d_0 . d_{-1} d_{-2}$$

$$= d_2 \times 10^2 + d_1 \times 10^1 + d_0 \times 10^0 + d_{-1} \times 10^{-1} + d_{-2} \times 10^{-2}$$

And an example with an arbitrary number of digits, N before the decimal point and n after the decimal point.

$$d_N \cdots d_1 d_0 . d_{-1} \cdots d_{-n}$$

$$= d_N \times 10^N + \cdots + d_1 \times 10^1 + d_0 \times 10^0 + d_{-1} \times 10^{-1} + \cdots + d_{-n} \times 10^{-n}$$

We call the part before the decimal point the **integer part** and the part after the **fractional part**. ♦

Any number written in this form can be written as a single fraction. For example, look at 7.13. The fractional part represents the number of hundredths $= \frac{13}{100}$ s, so this number is $7 + \frac{13}{100}$ or $\frac{713}{100}$. In general, if there are n digits after the decimal point then the number can be written as a fraction with denominator 10^n .

Going in the other direction, any fraction whose denominator is a power of 10 can be written in this form. If the denominator is 10^n then you need n digits past the decimal point. If there are not enough digits in the numerator you fill in the missing slots with 0s. For example:

$$\frac{4513}{1000} = 4.513$$

$$\frac{13}{1000} = 0.013$$

A reasonable question is, can any fraction be written in this form? If the fraction can be written to have a denominator which is a power of 10, the answer is yes. That can happen just in case the denominator's only prime factors are 2 or 5, as those are the prime factors of 10. For example:

$$\frac{1}{2} = \frac{5}{10} = 0.5$$

$$\frac{31}{25} = \frac{124}{100} = 1.24$$

For fractions with other denominators, however, it would take infinitely many digits past the decimal point to write exactly. For example, $\frac{1}{3}$ is written in decimal as 0.33333... where the 3s repeat forever. The upshot of this is, the real numbers can't just consist of numbers with finitely many digits past the decimal. For another example, the circle constant π requires infinitely many digits to write.

Those numbers which require infinitely many digits present complications for calculations. How do you add two numbers if there are infinitely many digits to look at? For practical purposes, however, this isn't a problem. If you're out in the real world you don't know quantities to that level of precision. Thus it suffices to be able to work with finitely many digits. Our next topic is how to modify our algorithms for arithmetic to deal with these, and then we will talk about approximation.

6.2 The arithmetic of Hindu–Arabic numerals

The algorithms for arithmetic of whole numbers written as Hindu–Arabic numerals can be modified to work with fractional parts. With addition and subtraction the new complication is, we need to make sure the decimal points line up. The reason is, when you add in columns for the algorithm each column needs to represent the same power of ten. Here’s an example.

$$\begin{array}{r}
 1\ 1\ 1 \\
 4\ 3\ 2.5\ 3 \\
 +\ 2\ 7\ 8.9 \\
 \hline
 7\ 1\ 1.4\ 3
 \end{array}$$

Figure 6.2.1 The sum $432.53 + 278.9$.

In general, the procedure for addition with fractional parts work the same as [Chapter 2](#)’s case without fractional parts. You start at the rightmost column and add columns as you move to the left, carrying as necessary. As usual, if a number is blank in a column that represents zero copies of that power of 10. What is new is that now blanks can appear on the right, after the decimal point, as opposed to only on the left.

Let’s see a couple more examples of addition.

$$\begin{array}{r}
 1 \\
 1.4\ 1\ 8 \\
 +\ 2.0\ 0\ 3 \\
 \hline
 3.4\ 2\ 1
 \end{array}$$

Figure 6.2.2 The sum $1.418 + 2.003$.

$$\begin{array}{r}
 0.0\ 0\ 3\ 4 \\
 +\ 0.0\ 2\ 1 \\
 \hline
 0.0\ 2\ 4\ 4
 \end{array}$$

Figure 6.2.3 The sum $0.0034 + 0.021$.

Subtraction works much the same, except you subtract in each column rather than add. As with integers, you may need to borrow from the next column. Let’s see a couple examples.

$$\begin{array}{r} 2.003 \\ - 1.418 \\ \hline 0.585 \end{array}$$

Figure 6.2.4 The difference $2.003 - 1.418$.

$$\begin{array}{r} 0.0201 \\ - 0.0034 \\ \hline 0.0167 \end{array}$$

Figure 6.2.5 The difference $0.0201 + 0.0034$.

Multiplication largely works as before, but presents an extra complication with where to place the decimal point in the product. To illustrate the issue, let's think of our numbers as fractions. For example, write 2.3×3.1 as $\frac{23}{10} \times \frac{31}{10}$. Whatever the numerator of the product, it is clear that the denominator ought be 100. That is, instead of the product having one digit past the decimal point it should have two. Once we know where it goes, the rest of the work goes through the same. Namely, multiply by the rightmost digit, then as you move left through the digits shift your position left each time, and finally sum up all the pieces.

$$\begin{array}{r} 2.3 \\ \times 3.1 \\ \hline 23 \\ 69 \\ \hline 7.13 \end{array}$$

Figure 6.2.6 The product 2.3×3.1 .

In general, we can determine where the place the decimal point of the product by counting the digits after it in each multiplicand. Consider the product 2.51×3.131 . There are a total of $2 + 3 = 5$ digits after the decimal point, which correspond to a least ten's place of $10^{-2} \times 10^{-3} = 10^{-5}$, so the product will have five digits after the decimal point.

$$\begin{array}{r}
 2.51 \\
 \times 3.131 \\
 \hline
 251 \\
 753 \\
 251 \\
 753 \\
 \hline
 7.85881
 \end{array}$$

Figure 6.2.7 The product 2.51×3.131 .

Division creates a problem. You can do the long division algorithm, taking the partial remainders farther and farther out in the decimal expansion. But there's no guarantee that this process ever stops. With long division of integers, the process eventually stopped because you got to the units digit and there were no more digits to take from. Here there's always more space. Indeed, most of the time, the process will never stop. You will get an infinite loop of a calculation that goes on forever, producing more and more digits. This happens with an example as simple as $1/3$ or $1/7$.

Later in this chapter we will talk about real numbers that need infinitely many digits to write. For now let me leave this as a signpost of the need for infinite decimal expansions.

Checkpoint 6.2.8 Do the division $1/3$, where instead of stopping with a remainder of 1 you take from past the decimal point to keep going. Stop after the pattern becomes clear.

6.3 Approximations

In real life, you never know quantities to exact precision. If your speedometer reads that your car is going 35 miles per hour, it could be 35.4 miles per hour, or 29.8 or 35.0. Indeed, that level of precision isn't helpful, so you would just round all of these to 35 miles per hour.

A useful feature of writing numbers in decimals is that it makes it straightforward to handle approximations. You can look at, say 4.12 and immediately see it is close to 4. In contrast, it would take a little thought to see that $\frac{103}{25}$ is close to 4.

Let's first talk about rounding to a whole number.

Definition 6.3.1 Rounding to an integer. **Rounding** a real number x to an integer is a way to replace x with an integer which is close to x . There are few ways you can do this. **Rounding toward zero**, also called **truncation**, is the closest integer between x and 0. If x is written in decimal, this is deleting all the digits past the decimal point. For instance, the truncation of -32.1 is -32 .

- The **floor** of x is the closest integer $\leq x$. This is also called **rounding toward $-\infty$** or **rounding down**. If x is positive this is equivalent to truncation.
- The **ceiling** of x is the closest integer $\geq x$. This is also called **rounding**

toward ∞ or rounding up.

- **Rounding to the nearest integer** replaces x with the integer closest to x . That is, if the first digit of x is 0 through 4 then you round toward zero, and if the first digit of x is 5 through 9 then you round away from zero.



Example 6.3.2 Consider the number $\pi \approx 3.14$. Rounding toward zero, rounding down, and rounding to the nearest integer all round to 3. Rounding up rounds to 4.

Or consider $x = 3.6$. Rounding toward zero and rounding down both round to 3. Rounding up and rounding to the nearest integer both round to 4. ▲

Example 6.3.3 With positive numbers, rounding toward zero and rounding down are equivalent. The distinction between them shows up for negative numbers. Consider $x = -3.4$. Rounding down rounds to -4 while rounding toward zero rounds to -3 . On the other hand, for negative numbers rounding up and rounding toward zero are equivalent. ▲

Sometimes people simply say “rounding” to mean rounding to the nearest integer. If you see it said without specifying the method this is probably what they mean.

When rounding to the nearest integer, what do you do when a number is exactly halfway between two integers? Do you round 3.5 to 3 or to 4? The two integers are tied for how close they are, so how do you decide the tie-breaker?

The rule given above is to always round up in case of a tie. For some purposes, however, this rule can introduce bias. To illustrate this, imagine you are using a digital scale which gives weights accurate to one tenth of the gram and you are rounding to the nearest gram. For instance, it might display a weight of 23.4 grams, which you would round to 23 grams. One possible digit after the decimal point, namely 0, doesn’t give you any error when you round. Of the remaining nine, eight of them are balanced. Rounding down from 23.1 introduces an error of 0.1 below the measured weight, but rounding up from 23.9 introduces an error of 0.1 above the measured weight. If you do lots of measurements on average these errors will cancel. The remaining digit, 5, always rounds upward, introducing an error of 0.5 above the measured weight. This one error is not canceled out, and so if you weight many different things and added them up the errors would tend to bias upward.

This is, admittedly, quite in the weeds. If you’re using your digital scale for, say, home baking then the error is too small to matter. But for more precise matters, such as scientific or engineering uses, this bias can matter. In such a case, you’ll want to use a different tie-breaker rule.

The lesson here is, there is no one rule that’s the best for all circumstances. There are trade-offs of convenience and precision, and what is best depends on what you’re using it for. This is common. Mathematics may furnish many possible methods one could use, and no one of them is better than the others for all applications.

Indeed, this is true for rounding. Is it best to round up? Round down? Round to the nearest integer? The answer is: it depends.

Rounding to an integer is the same as rounding to the units place. You can do the same process, but rounding to any other place. When you round to the units digit, all digits to the right—that is, all digits after the decimal point—become 0s. The same applies when rounding to other places. For example, if you round to the hundreds place then the tens and units digits, as well as all

digits after the decimal, become 0. Or if you round to the tenths place, then all digits after that—hundredths, thousandths, etc.—get set to 0.

As with rounding to an integer, there are multiple methods to round to the nearest tens, etc. You can round up to the next thousand, you can round to the nearest thousand, you can round toward zero to the next thousand, and so on. If you round to the nearest, you can tell which direction to round by looking at the next place. For example, to round to the nearest tenth you look at the digit in the hundredths place. The digits 0 to 4 you round toward zero, and the digits 5 to 9 you round away from zero.

Example 6.3.4 Consider the number 231.78. Rounded to the nearest tenth it is 231.8. Rounded down to the tenths place it is 231.7. Rounded to the nearest ten it is 230. Rounded up to the next ten it is 240. Rounded to the nearest thousand it is 0 while rounded up to the next thousand it is 1000. ▲

Physicists tell us that the speed of light is 299,792,458 meters per second. This value is known exactly. But when doing calculations, you probably don't need the number to that precision. It's enough to say it's about three hundred million meters per second. Rounding gives you an easier number to say and work with.

The reason we know the speed of light exactly is because that is the modern definition of the meter: it is defined to be $\frac{1}{299,792,458}$ of the distance traveled by light in one second. This is not the original definition of the meter. Older definitions were based on distances around the globe, so as to create a standard that could be used internationally. Advances in physics allowed this to be replaced by the current, light-based definition.

Most quantities in the real world aren't known to that level of precision. For example, suppose you measure the sides of a square box to be about 1 foot. Using the Pythagorean theorem, you calculate that the length of the diagonal of the box is $\sqrt{2} = 1.414213562\dots$ feet. It wouldn't make sense to report that length to that level of precision. The sides of the box are not exactly 1 foot, and so the diagonal is not exactly $\sqrt{2}$ feet. For example, if the true length of the side was 1.05 feet then the diagonal would be about 1.48 feet long. For this reason, it makes sense to round the value; reporting too precise a measure gives the misimpression that you know the value to that many decimal points.

Significant digits are a convenient way to organize this information. The idea is, the number of digits you write down represents how precisely you know a value. These are the digits that are **significant**, and any other digits are noise and shouldn't be reported. When doing calculations, you round your final values to the same number of significant digits. You can use the following rules to read and use significant digits.

- Nonzero digits within the level of precision are always significant. You should round so digits outside the level of precision aren't reported. For example, if you have two digits of precision then you should report 1.34 as 1.3.
- Zeroes left of the first nonzero digit are never significant.
- Zeroes after the decimal point are significant and should be reported if they are within the level of precision. For example, if you have three digits of precision then you should write 23.0 instead of 23, to communicate that the first digit past the decimal point is significant.
- Zeroes before the decimal point may or may not be significant. For example, 4200 might represent a value with two digits of precision, or it

might represent one with three or four digits of precision. Some underline or overline the last significant zero to indicate the level of precision. For instance, one could write $42\underline{0}0$ to represent three digits of precision. But this is not a universal convention.

- When doing calculations with multiple numbers, values should use the smallest number of significant digits from your inputs.

Example 6.3.5 You have two strips of cloth, one 420 centimeters long and the other 370 centimeters long. (For both of these, the numbers are at two digits of precision.) If you need 2 centimeters of length from each to sew them together, how long will your sewn-together strips be?

To answer this, calculate

$$420 + 370 - 2 \cdot 2 = 786.$$

You then round to two digits of precision, giving you 790 centimeters. Note that this is the same answer you'd get if you just calculated $420 + 370$ and ignored the amount of cloth needed for the stitching.

If instead the cloths were 42 and 37 centimeters long, with the same length needed for sewing, you would calculate

$$42 + 37 - 2 \cdot 2 = 75.$$

This is already at the correct level of precision so you don't need to round to report the total length will be 75 centimeters. Note that this is not the same value as $42 + 37 = 79$. ▲

Intuitively, it should make sense that if the strips are really long then the length needed for sewing them together is a rounding error. Whereas if the lengths of the cloths are close to the amount reserved to sew them together then that difference is significant.

Checkpoint 6.3.6 Suppose you measure that a 2×4 toy building brick is 31.8 millimeters long. If you placed 100 of them in a row, with no space between them, how long would it be? What if you did 23 of them

Solution. The calculations are each a single multiplication. For the first, $31.8 \cdot 100 = 3180$ millimeters. This is already the correct level of precision—three digits, matching the input—so you don't have to round. For the second, $31.8 \cdot 23 = 731.4$. This is more digits than your level of precision, so you should round to 731 millimeters.

Checkpoint 6.3.7 You have a rope which you measured to be 8.0 meters long. You cut off 1.25 meters of rope from it. How much length is left over?

Solution. Calculate

$$8.0 - 1.25 = 6.75.$$

This is three digits of precision, whereas your input measurements had two and three digits of precision. You should round to the smaller of those, so round to two digits of precision: you have 6.8 meters of rope left over.

The rounding here represents the uncertainty in your initial measurements. You measured that the rope was originally 8.0 meters, but it could be 8.00 or 8.05 or 8.09. You don't know, and these will give slightly different lengths for what's left after cutting. You can't be more sure of your calculated length than you are of the measurements that went into it.

A practical way of dealing with both significant figures and magnitudes—is the number approximately one million? one trillion? one thousandth?—is given by **scientific notation**. Let me give an example first. The speed of light

is, to four digits of precision, 299,800,000 meters per second. The magnitude of this number is some number of hundred millions, or 10^8 s. How many 10^8 s is it? The first significant digit gives the whole number, and the rest are the fractional part. Namely, this is 2.998 many 10^8 s. We represent this succinctly as a multiplication: the speed of light is approximately 2.998×10^8 meters per second.

Definition 6.3.8 Scientific notation. A number is written in scientific notation when it is written in the form

$$D.d_1d_2 \cdots d_n \times 10^k,$$

where D, d_1, d_2, \dots, d_n are digits 0–9 and k is an integer. The number k is called the **exponent** and the digits are called the **significand**. Phrased another way, a number is written in scientific notation when it is written in the form

$$s \times 10^k,$$

where s is a number between 1 and 10 and k is an integer. ◆

Some calculators use a shorthand for scientific notation, writing $2.43E5$ to mean 2.43×10^5 . (The E naturally stands for “Exponent”.)

Example 6.3.9

- You write 12.3 in scientific notation as 1.23×10^1 .
- You write 4.789 in scientific notation as 4.789×10^0 .
- You write 0.086 in scientific notation as 8.6×10^{-2} .
- You write 800,800 in scientific notation as 8.008×10^5 .
- The number 345×10^{100} isn’t quite in scientific notation. To write it that way, it should be 3.45×10^{102} .

▲

The virtue of scientific notation is that it lets you quickly see how large a quantity is and its number of significant digits. It is commonly used, as the name suggests, in science. For example, in chemistry **Avogadro’s constant** is used to convert between mass of a chemical and the number of particles therein. It is approximately 6.022×10^{23} . This is too large in magnitude to comfortably write out in the Hindu–Arabic numeral system:

602200000000000000000000

(This number is actually wrong; it’s 6.022×10^{24} so it’s off by an order of magnitude. But checking that requires counting digits and is prone to mistakes!)

Another way this problem is handled is with prefixes for units in the metric system. Rather than write 6000 grams one writes 6 kilograms. Or instead of saying that visible light has a wavelength of about 5×10^{-7} meters you can say it’s about 500 nanometers.

6.4 Defining the real numbers

Having spent a bunch of time on working practically with real numbers, let’s close the section with a little theory. How do we make sense of what the real numbers are?

Natural numbers we made sense of in terms of what they do—they count discrete, finite collections. With integers and rational numbers we built on top of the natural numbers. Integers are natural numbers with sign (plus or minus) to represent direction. Rational numbers are formed as fractions of integers. These let us solve the problems, respectively, of not always being able to subtract and not always being able to divide.

The problem that real numbers solve is the problem of holes in the number line. While there are lots and lots of rational numbers and you can get them as close as you like, there's still holes—numbers that can't be written as a fraction of integers at all. We call these numbers **irrational**, because they are not rational numbers. Examples of these include $\sqrt{2}$ and the circle constant π .

It can be very difficult to show that a number is irrational. To get a taste, here is an explanation of why the square root of 2 is irrational.

Do not feel bad if you look at the following proof and don't understand it. The logic of it can be difficult to wrap your head around. I included it not because I expect you to read and immediately understand it but instead to illustrate the challenge in concluding that a number is irrational.

Theorem 6.4.1 *The number $\sqrt{2}$ is irrational. That is, there are no integers p, q so that $\sqrt{2} = \frac{p}{q}$.*

Proof. How do you show that there are no such integers p, q ? One way is to assume that there were such p, q and show that that is impossible. That is the approach we will take. To that end, suppose that p and q are integers so that $\sqrt{2} = \frac{p}{q}$. As we discussed in the chapter on rational numbers, any fraction can be written in reduced form where there are no common factors between the numerator and denominator. Let's thereby assume that we've already done any canceling out so p and q have no common factors.

Now we do a little algebra.

$$\begin{aligned}\sqrt{2} &= \frac{p}{q} \\ 2 &= \frac{p^2}{q^2} \\ 2q^2 &= p^2\end{aligned}$$

What this shows is, p^2 is even. This means that p is also even, because if it were odd then its square would also be odd. In symbols, $p = 2k$ for some integer k . Substitute that into the equation and do a little more algebra.

$$\begin{aligned}2q^2 &= (2k)^2 \\ 2q^2 &= 4k^2 \\ q^2 &= 2k^2\end{aligned}$$

This shows us that q^2 is even, so like before q is also even. But this is impossible! If p and q are both even then they have a common factor of 2, when they were supposed to have no common factor.

We have seen that our assumption that there are integers p, q so that $\sqrt{2} = \frac{p}{q}$ is impossible, so $\sqrt{2}$ must be irrational. ■

With some heavy-duty number theory you can generalize this fact. Say that an integer is a **perfect square** if it is a square of an integer. And an integer is a **perfect n -th power** if it is an n -th power of an integer. Note that the square root of a perfect square is a rational number, because it is a whole

number. And similarly for n -th roots of perfect n -th powers.

Theorem 6.4.2 *The number \sqrt{k} is irrational if k is not a perfect square. And $\sqrt[n]{k}$ is irrational if k is not a perfect n -th power.*

The upshot is, there are lots and lots of numbers we can't represent as fractions of integers, and so the rational numbers aren't enough.

We can fill some of these holes in the number line by declaring that $\sqrt{2}$, $\sqrt{3}$, $\sqrt[3]{2}$, and so on are all numbers too. But how do we know that we didn't miss any? For example, if we just include n -th roots of integers then we won't have thrown in π . (The fact that π isn't a root of an integer is a difficult theorem.) If we want to fill all the holes, we need something less piecemeal.

Decimal expansions give us a way to do this. We can say that a real number is any number you can write in decimal, possibly with infinitely many digits past the decimal point. This will include $\sqrt{2}$, π , and so on.

Definition 6.4.3 The real numbers \mathbb{R} . The **real numbers** are all numbers which can be expressed with infinite decimal expansions. That is, a real number is one of the form

$$\begin{aligned} & d_N \cdots d_1 d_0 . d_{-1} \cdots d_{-n} \cdots \\ & = d_N \times 10^N + \cdots + d_1 \times 10^1 + d_0 \times 10^0 + d_{-1} \times 10^{-1} + \cdots + d_{-n} \times 10^{-n} + \cdots, \end{aligned}$$

where each digit d_i is a whole number between 0 and 9. We write \mathbb{R} to refer to the set of real numbers. \blacklozenge

Example 6.4.4

- The fraction $1/2$ can be written as 0.5. If you want to make the infinite decimal expansion clear, this is 0.5000... where the 0s repeat infinitely.
- The fraction $1/3$ can be written as 0.333... where the 3s repeat infinitely.
- The fraction $1/7$ can be written as 0.1421857... where the digits 142857 repeat forever.
- The number π can be written as

$$\pi = 3.1415926535897932384626433832795028841971693993751 \dots$$

The decimal expansion of π never repeats. No matter how far out you go it will always be a new pattern. \blacktriangle

Numbers written with finitely many digits represent sums. For example, the number 623.4 represents the sum

$$6 \times 10^2 + 2 \times 10^1 + 3 \times 10^0 + 4 \times 10^{-1}.$$

Numbers with infinite decimal expansions also represent sums, but now the sums have infinitely many terms. For example,

$$\frac{1}{3} = 0.33 \dots = 3 \times 10^{-1} + 3 \times 10^{-2} + \dots$$

represents an infinite sum, with each term getting smaller and smaller. Infinite sums require calculus to fully understand. Without going into the calculus ideas, let me give you a taste of the challenges.

When adding finitely many numbers, the sum always makes sense. This is not the case when adding infinitely many numbers. For instance, consider the

sum of all positive integers:

$$1 + 2 + 3 + 4 + \cdots + n + \cdots .$$

This doesn't add up to a number. You keep adding larger and larger values, so the overall sum would have to be infinite in size, but all numbers are finite. That is, there's no number you can consistently declare to be the sum of these infinitely many numbers.

For another example, consider the sum

$$1 - 1 + 1 - 1 + 1 - 1 \pm \cdots .$$

This also doesn't add up to a number. If you try to evaluate this sum by adding more and more terms, you'll see that the partial sum flipflips between 1 and 0. Because it keeps flipflipping between two different values there's no single number to be the value of the entire infinite sum.

On the other hand, some infinite sums do add up to a single finite value. Imagine you have a cake and a scheme to feed infinitely many friends with it. You cut it in half and take one half for yourself. Then you take the remainder and cut it in half again, giving about one of the two pieces to your first friend. You cut the remainder in half and give it to the next friend. You repeat this process, so each friend gets half as much as the previous one, feeding infinitely many people. (Although some might be insulted at the tiny piece of cake you gave them.

If the original cake had a size of 1, how big is each piece? Your piece has size $\frac{1}{2}$. Your first friend gets half of that, namely $\frac{1}{4}$. The next friend gets $\frac{1}{8}$, then $\frac{1}{16}$. In general, the n -th person to get a slice had it cut in half n many times before it gets to them, so they get $\frac{1}{2^n}$ much cake. Add all these pieces together and you get the sum

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots + \frac{1}{2^n} + \cdots .$$

Since these infinitely many pieces together comprise the entire cake, the value of this infinite sum must be the size of the cake, namely the sum is 1.

In mathematician-speak, infinite sums that have a value are called **convergent** while those that don't are called **divergent**. The good news is, if you're just dealing with the sums from numbers written in decimal you don't have to worry about this, because everything is convergent.

Theorem 6.4.5 *Every real number represents a convergent sum. That is, every real number really is a number.*

It's perhaps helpful to contrast real numbers with something you could write that isn't a number. What if you tried to write infinitely many digits to the left of the decimal point? Consider the "number" $\cdots 1111$. This would represent the sum

$$1 + 10 + 100 + \cdots + 10^n + \cdots$$

Much like the $1 + 2 + 3 + \cdots$ sum from before, this infinite sum diverges to infinity, since the values you are adding grow without bound.

In summary, infinitely many digits after the decimal point is good, while infinitely many before the decimal point is bad.

You might ask at this point, what about the holes? Did this really fill all of them? The answer is yes. The real numbers \mathbb{R} give a complete number line with no gaps. Making this mathematically precise and proving it is, on the other hand, a really big task. The undergraduate math class in which one would learn all these details, called real analysis, is one which is too high

level to be taught at community colleges. As such I will leave us with that statement.

Is this the end of the story for number systems then? The answer is no. There are interesting and useful number systems which go beyond the real numbers. For example, the **complex numbers** were invented to make sense of square roots of negative numbers. (Remember that $\sqrt{-1}$ doesn't make sense in the real numbers; if you multiply a real number by itself it will always be ≥ 0 and so won't be -1 .) These form a two-dimensional plane of numbers, and enjoy use in electrical engineering, physics, and more.

There are yet other, more exotic number systems of interest to few outside of mathematics. Hyperreal numbers, surreal numbers, quaternions, ordinal numbers, and other number systems get used for specialized purposes. But for most people, the real numbers are all you will ever need.

6.5 Characterizing rational and irrational numbers

It's neat that decimal expansions can be used to characterize which real numbers are rational versus irrational. This is of limited practical use—you can't survey an entire infinite string of digits in any reasonable time—but it serves as a nice coda for this chapter.

Theorem 6.5.1 Characterizing the rational numbers with decimal expansions. *A real number is rational, meaning it can be written as a fraction of integers, if and only if its decimal expansion eventually repeats. Thus, an irrational number is one whose decimal expansion goes forever without repeating.*

Let's look at a few different examples. We will write a bar over the sequence of digits which repeat, rather than relying on unclear dotdotdots. For example, we will write $0.090909\dots$ as $0.\overline{09}$.

The number 0.5 has a decimal expansion which eventually repeats—after the 5 it's all 0 s, which we don't write. Thus it represents a rational number. Namely, this is $\frac{1}{2}$.

The number $0.\overline{3}$ has its decimal expansion consisting entirely of 3 s repeated. It represents the rational number $\frac{1}{3}$.

The number $0.00\overline{142857}$ represents the rational number $\frac{1}{700}$. Note that the repeating sequence of digits doesn't start right after the decimal point.

The decimal expansion for $\sqrt{2}$ starts with $\sqrt{2} \approx 1.41421$. Since this number is irrational it will never repeat, no matter how far out you go.

The decimal expansion for π starts with $\pi \approx 3.14159$. This number is irrational, so it will never repeat.

Note the order of the logic here. We know π is irrational, therefore we know the decimal expansion never repeats. In theory you could go the other way, but how would you conclusively show that its decimal expansion never repeats? It would take infinitely long to look at the whole thing. Much quicker to just spend years studying math to understand a proof that π is irrational.

On the other hand, you can create patterns that don't repeat and thereby produce yet more irrational numbers. However, these numbers aren't useful for much. Consider the number

$$0.7070070007\dots$$

where between each pair of 7 s there is one more zero than last time. This decimal expansion never repeats—there's always longer and longer blocks of

0s the farther out you go. So this number is irrational.

Checkpoint 6.5.2 Give a definition of an irrational number and explain how you know it's irrational.

6.6 Exercises

Exercises

1. Calculate the following sum and difference. Show all steps in your calculation.

$$13.4 + 29.9, \quad 4.1043 - 3.255$$

2. Each year, an investment increases in value from its current value by 5%. If it is initially worth 1000 dollars, how much is it worth after two years? After six years? After twelve years? Round all your answers to the nearest hundredth.

Hint 1. You can write 5% in decimal as 0.05 and 105% is 1.05. Using this, write a formula which describes the value of the investment after n years.

Hint 2. If you're completely stuck, reread the section from about percents.

3. For the following products, state how many digits past the decimal point the product will have.

$$12.3 \times 123, \quad 0.13 \times 0.43, \quad 1.0001 \times 1.0000001, \quad 123 \times 0.00123$$

4. Calculate the product $3.4 \cdot 4.3$ by the box method. That is, individually do the four products

$$3 \times 4, \quad 3 \times 0.3, \quad 0.4 \times 4, \quad 0.4 \times 0.3$$

and add them up to get the sum. For each of the four intermediate products, how many digits past the decimal point does it have?

Hint. One way to determine the number of digits past the decimal point is to write things as fractions: $0.3 = \frac{3}{10}$ and $0.4 = \frac{4}{10}$. Then use rules for multiplying fractions.

5. For each of the following numbers, round to the nearest ten and the nearest tenth using each of: rounding toward zero, rounding up, rounding down, and rounding to the nearest ten/tenth. Clearly label your rounding method for each.

$$-73.97, \quad 1349.15$$

6. Consider a new rule for rounding, namely **rounding away from zero**. That is, to round a number x away from zero to the nearest integer you move away from zero toward the next integer. Demonstrate this new rule by rounding away from zero to the next integer for all of the following.

$$14.2, \quad -3.9, \quad 16, \quad -4.2$$

For which numbers is this rule equivalent to rounding down? For which numbers is this rule equivalent to rounding up?

7. Write the following numbers in scientific notation. For each one, how many significant digits does it have?

$$134.5, \quad 0.0004130, \quad 920100000, \quad 723 \times 10^{-6}$$

8. Come up with a rule for how to multiply two numbers written in scientific notation. Demonstrate your rule by calculating the product

$$1.2 \times 10^3 \cdot 2.0 \times 10^5.$$

Hint. A product like $1.2 \times 10^3 \cdot 2.0 \times 10^5$ can be thought of as multiplying four numbers together. Use rules for multiplication and exponents to rewrite.

Numbers with a repeating decimal expansion are examples of what mathematicians call **geometric sums**. Namely, a geometric sum is one of the form

$$a + ar + ar^2 + ar^3 + \cdots + ar^n + \cdots.$$

The rule for geometric sums is, if $|r| < 1$ then the sum converges to $\frac{a}{1-r}$. For example,

$$\begin{aligned} 1.\bar{1} &= 1 + \frac{1}{10} + \frac{1}{100} + \cdots + \frac{1}{10^n} + \cdots \\ &= \frac{1}{1 - \frac{1}{10}} \\ &= \frac{1}{9/10} \\ &= \frac{10}{9} = 1 + \frac{1}{9}. \end{aligned}$$

In this case, we had $a = 1$ and $r = \frac{1}{10}$.

9. Write the numbers $3.\bar{3}$ and $9.\bar{9}$ as infinite sums and use the rule for geometric sums to determine their values. What do your answers tell you about the rational numbers these are equal to?

Hint. Follow the same steps as the example.

10. Consider the two numbers x and y given by the following patterns for their decimal expansions. Each number has an integer part of 0.

- The digits after the decimal point for x are obtained by writing the integers 1 through 9 in order one after another, looping back to 1 whenever they reach 9.
- The digits after the decimal point for y are obtained by writing the integers starting at 1 in order one after another, never looping back.

For each of these numbers, is it rational or irrational? Explain your answers.

Hint. If you are not sure, begin by writing out the decimal expansion of the number. What patterns can you notice?

Colophon

This book was authored in PreTeXt.