Number Systems

An axiomatic foundation for algebra, number theory, and analysis

Wayne Aitken and Linda Holt

January 2023 edition

Contents

Introduction

This book carefully and rigorously develops the basic number systems including the following.

- The set of natural numbers $\mathbb{N}$.
- The ring of integers $\mathbb{Z}$.
- The ring $\mathbb{Z}_m$ of integers module a positive integer m, including the field $\mathbb{Z}_p = \mathbb{F}_p$ of integers modulo a prime number p.
- The field of rational numbers $\mathbb{Q}$.
- The field of real numbers $\mathbb{R}$.
- The field of complex numbers $\mathbb{C}$.

In addition to developing these important number systems, this course emphases important ideas in mathematics more generally.

- *The axiomatic method.* We start with Peano's axioms (due to Dedekind as well as Peano), and derive everything from these using classical deductive logic and basic set theory. This book aims to be completely self-contained as long as the reader has some facility with logic and very basic set theory. Every result will be proved from the axioms, or earlier results in this book.
- *Mathematical structures.* We start with the notion of an ordered set, and then move on to algebraic structures such as groups, rings, integral domains, fields, and ordered fields. We do not go too deeply into these topics, but we give enough exposure that a reader will be more than prepared for a standard course in abstract algebra. In fact, the reader acquires a standard supply of examples and ideas that should make abstract algebra quite accessible.

- *Limits and continuity*, including Cauchy sequences, suprema and infinima and so on. These topics are central to advanced calculus and analysis. We do not go into as much depth as in an analysis course, for instance we do not treat limits of functions but only limits of sequences, but still a student who masters the material here will be in an excellent position to tackle a rigorous course in analysis.

- *Divisibility, prime numbers, and congruences.* Some of the foundational ideas of number theory are covered in this course. A student who masters the material here will be in good shape to take an introductory number theory course, even one that uses some ideas from abstract algebra.

- *Building structures with equivalence relations.* We build up $\mathbb{Z}$ using equivalence classes of ordered pairs of elements of $\mathbb{N}$, we build up $\mathbb{Z}_m$ and $\mathbb{Q}$ using equivalence classes of objects built from $\mathbb{Z}$, we build $\mathbb{R}$ from equivalence classes of Cauchy sequences of elements of $\mathbb{Q}$. Using multiple examples, this book illustrates the usefulness of equivalence classes for building mathematical structures. After mastering this material, students should have no problem with more advanced constructions such as quotient groups and rings.

- *Practice with induction and other proof techniques.* The student who works through all the exercises will have a great amount of practice honing their skills in induction and a wide variety of proof techniques. This should make more advanced mathematics courses much more accessible.

The above indicates how this book will provide a good foundation for further mathematical study. This book is also useful in preparing future teachers. The subject matter here is basic to high school level, and even pre-high school level mathematics, and future teachers who master this material will have a deeper understanding of this material, allowing them to be more confident and versatile teachers.

We hope that this book will find an audience among experienced people in mathematics and the sciences who are interested in going back to the basics and seeing how the foundations are developed rigorously. Many scientists, graduate students, and mathematicians take this material for granted, even if they have a high standard of rigor in their own research. Many of these people will naturally be curious about the foundations of the number systems. This book should help satisfy this curiosity and provide a strong foundation for their work in mathematics.

This book has a few important appendices.

- An appendix called "Chapter 0" which summarizes the logic and set theory needed to start the project of building the number systems.

- An appendix that continues the development of the complex numbers up to De Moivre's theorem and the complex exponential function. This is offered as an appendix since it uses results from outside this book (especially the sine and cosine functions, and the real exponential function). We hope to write a sequel to this book someday that develops this material in complete rigor.

- An appendix that concerns polynomials leading up to the fundamental theorem of algebra. Only some of the results are proved, but they do lead to greater insight into the real and complex number systems. We hope to write a sequel to this book someday that develops this material in complete rigor.

- An appendix (not in this edition) outlining alternative constructions of the number systems including the Dedekind construction of the real numbers via Dedekind cuts. This appendix will also discuss various isomorphism theorems. For example, all models of Peano's axioms are isomorphic, all complete ordered fields are isomorphic, and so on.

In the spirit of the axiomatic method, our development of the number systems will be rigorous and self-contained: we will give careful proofs for our results. There are, however, two exceptions, outside the appendices, where we will allow results without proof:

1. *Axioms.* These are fundamental statements that are accepted without the need for formal justification. Sometimes they are presented as "self-evident", but technically they do not need to be obvious. They are, however, accepted as true for the purpose of proving further results.

   In this course the only axioms are the Dedekind-Peano axioms and the iteration axiom. In an optional section near the end of Chapter 1 the iteration axiom will be shown to be a consequence of the other axioms, so the only axioms that are necessary for this course are the Dedekind-Peano axioms.[1] In more advanced mathematics, the axiom of choice, and certain advanced set theoretic axioms are also sometimes needed.

[1]These axioms, coupled with some basic set theory, suffice for a large part of mathematics. Even geometry can be developed from these axioms. For example, once you have developed the real numbers $\mathbb{R}$, you can define the plane to be $\mathbb{R}^2$ and three-dimensional space to be $\mathbb{R}^3$. In this approach you develop all the theorems of Euclidean geometry using the coordinate point of view and no new geometric axioms are needed. This is in contrast to Euclid's original approach, updated by Hilbert, which develops geometry using geometric axioms that do not rely on the real numbers.

2. *Principles of logic and elementary set theory.* From the axioms, we will derive other results using logic. So we will take as given the knowledge of classical deductive logic. This logic can be used freely to derive new results. For example, we assume the basic principles related to connectives ($\wedge$, $\vee$, $\implies$, $\neg$, $\iff$) quantifiers ($\forall$, $\exists$, $\exists!$), and equality ($=$).

   We will regard elementary set theory as part of our logical background and toolkit. These includes concepts, rules, and facts that are in common use in modern mathematics. Included under the heading of set theory are principles concerning ordered pairs, functions, and relations as well as sets (in fact, ordered pairs, functions, and relations can be modeled as certain types of sets). See Appendix A for a complete description of what we take as given. We do not include the topic of cardinality as part of the set-theory background since we will develop the theory of finite cardinality here (in Chapter 3). Note that set theory can be developed axiomatically from a small set of axioms, but we will not do so here. We simply take them as given.[2]

Aside from the axioms, and the basic facts of logic and set theory, every statement we wish to establish or use must be proved. Even something as simple as the commutative law of addition, or even the equation $1 + 1 = 2$, will be proved.

Likewise, every *concept* not occurring in the axioms, logic, or elementary set theory must be defined before it can be used. Such a definition must use only set theoretical and logical concepts as well as previously established concepts. For example, we will define addition and multiplication using the concept of function from basic set theory. Similarly, will provide definitions for all the number systems except the natural numbers using various set-theoretical ideas applied to previously established number systems. The set of natural numbers is an exception; it will not be defined. Since the set $\mathbb{N}$ is part of the axioms, it does not actually need to be defined. In general, terms used in the axioms do not need to be defined, and such undefined terms are called *primitive terms*.

The intent is to develop all the results in a self-contained manner, but not all the results are proved explicitly in the text. There is a division of labor where we, the authors, prove some of the results and you, the reader, prove others. Thus many of the results are left as exercises, but there should be enough hints and context to make it possible for the reader to prove theses results. One common situation is where we, the authors, give the steps of the

[2]The best known axiomatic development of set theory uses the Zermelo-Fraenkel axioms including the axiom of choice. This is a very powerful axiom system and is overkill for what we do here. If we were to axiomatize the set theory needed, we would use a weaker form of these axioms without, for example, the axiom of replacement or the axiom of infinity.

proof but not always with explicit justifications. You, the reader, should be able to find previous results or axioms that support each such step. Some of the exercises will be labelled as "informal". These do not have to be done using the strict axiomatic method, but can use ideas and results from outside of this book. These are designed to help the reader achieve a deeper understanding the concepts by applying them outside of the scope of the book.

Officially the material here is self-contained, but in fact we do expect that the reader is familiar with much of the material on at least an informal level. Much of this material is "elementary mathematics from an advanced standpoint". Readers' previous exposure to these ideas should be very helpful in allowing them to understand the material and provide valuable motivation and intuition. This knowledge may help readers construct proofs. However, the reader should be very careful not to let their prior knowledge cause them to make unjustified leaps of logic. Every result should be justified by the prior results in the book, using logic and basic set theory, and not by principles that they "know" are true.

Chapter 1

The Peano Axioms

1.1 Introduction

We begin our exploration of number systems with the most basic number system: the natural numbers $\mathbb{N}$. Informally, natural numbers are just the ordinary whole numbers $0, 1, 2, \ldots$ starting with 0 and continuing indefinitely.[1] For a formal description, see the axiom system presented in the next section.

Throughout your life you have acquired a substantial amount of knowledge about these numbers, but do you know the reasons behind your knowledge? Why is addition commutative? Why is multiplication associative? Why does the distributive law hold? Why is it that when you count a finite set you get the same answer regardless of the order in which you count the elements? In this and following chapters we will systematically prove basic facts about the natural numbers in an effort to answer these kinds of questions. Sometimes we will encounter more than one answer, each yielding its own insights. You might see an informal explanation and then a formal explanation, or perhaps you will see more than one formal explanation. For instance, there will be a proof for the commutative law of addition in Chapter 1 using induction, and then a more insightful proof in Chapter 3 involving the counting of finite sets.

We will use the axiomatic method where we start with a few axioms and build up the theory of the number systems by proving that each new result follows from earlier results. In the first few chapters of these notes there will be a strong temptation to use unproved facts about arithmetic and numbers that are so familiar to us that they are practically part of our

[1] Warning: some authors do not include 0 in the set of natural numbers. This will be discussed in the next section.

mental DNA. Resist this temptation! In the context of a formal proof, take the attitude that such familiar facts are not certain until they are proved. So they cannot be used in a formal proof until after they have been proved. A similar thing can be said of definitions: pretend that your intuitive ideas of even basic things such as $+$ and $<$ are inaccessible until you can have a formal definition. In the beginning, the only terms that can be used are terms from logic and set theory, explained in Chapter 0, and the primitive terms. The only facts that can be used are the axioms together with facts from logic and set theory as summarized in Chapter 0, including general facts about equality, functions, and relations.[2]

The system of axioms we use here is a famous system called the *Dedekind-Peano axioms* (Section 1.2), or the *Peano axioms* for short. We will add to this an axiom about iterating functions (Section 1.3), but in an optional section (Section 1.9) to this chapter, we will see that this iteration axiom is not necessary since it can actually be proved from Peano's axioms. Thus it is strictly speaking a convenient "temporary" axiom: one could replace the iteration axiom by a theorem that says the same thing. We take it as a temporary axiom in these notes since the proof of the iteration axiom is a bit subtle, and is at a higher level than most of the other theorems of the chapter. We do not want to start off the chapter by scaring away readers.

Remark 1. Although we will be strict about not using unproved assertions in the formal development, you do not need to be so shy about using your prior knowledge in the *informal* exercises. Such prior knowledge is also useful for temporarily guiding your thinking until a firmer foundation is laid down in the formal development.

This distinction between formal and informal is especially important in the many exercises that will arise in these notes. The informal exercises will be labeled as such. The rest are considered to be formal exercises.

The formal exercises may require you to fill in details of sketchy proofs or even to write complete proofs for theorems whose proofs are not too hard or are similar to earlier proofs. These constitute part of the official development of the number systems, and the facts established in them can be used in future proofs. On the other hand, the informal exercises are designed to help familiarize you with facts or definitions, or to lead you in interesting but tangential directions. These do not have to be solved with a formal proof, and can appeal to prior knowledge. They are considered to be outside the logical development of the number systems, and so cannot be cited in a later formal proof.

[2]In these notes, we start almost at the very beginning of mathematics, but you should be aware that there are other approaches that start with less and begin by proving theorems about set theory first before developing the number systems. For example, set theorists typically start with the Zermelo-Fraenkel axioms for set theory, and from there develop set theory, the number systems, and (most of) the rest of mathematics.

For example, suppose an informal exercise asks for an example of an associative binary operation that is not commutative. Suppose you know about matrix multiplication from a linear algebra course. Then you can use your knowledge of linear algebra to help solve the problem. On the other hand, you cannot use matrix multiplication in a formal exercise since matrices are not developed in this course.

Remark 2. In the above discussion, the term *theorem* refers to any result that has a proof. Keep in mind that other terms for theorems are commonly used including *proposition*, *lemma*, and *corollary*. The term *lemma* is used for a theorem that is only important as a stepping stone in proving other theorems, and a *corollary* is a theorem that follows fairly easily, for example as an interesting special case, from a previous theorem. Some authors also make a distinction between the terms *theorem* and *proposition*, using the label *proposition* for more ordinary theorems and using *theorem* only for the more important theorems. These are informal guidelines: one can find exceptions.

Remark 3. As mentioned above, in the formal development of the natural numbers we begin by assuming that everything about the natural numbers is as yet unknown territory. On the other hand, we do allow logic as expressed in everyday, but careful, language. This leads to a point that needs to be clarified: even though we are developing the natural numbers from scratch, we will allow ourselves to use a few number-related terms such as "pair", "unique", "first", "second", and so on. We do so because we can safely treat such basic terms as forming part of our *logical* vocabulary.[3] We will also use numerals for the labeling of sections, theorems, exercises, and such. These labels have no arithmetic content, and could have just as easily been any string of symbols. They are being used informally to help keep the chapter organized. On the other hand, we will not take any truly mathematical or arithmetic fact for granted, for example facts about addition and multiplication. These all must be proved.

1.2 The axioms

Forget everything you think you know about the natural numbers, even something as basic as $1+1=2$. Pretend you don't even know the definition of addition. In what follows, we will recreate all this knowledge on a solid logical foundation by *proving* all the elementary theorems and *definining* all

[3]For example, the statement "the set S has at least two elements" does not really require the number 2. It can be translated easily into basic logic as follows:

$$\exists x\, \exists y \, \Big((x \in S) \ \wedge \ (y \in S) \ \wedge \ (x \neq y) \Big).$$

the basic ideas. (Of course this self-imposed forgetting should be confined to the official formal development of the natural numbers, and the formal proofs. Your past knowledge will come in handy for thinking up strategies for proofs, for helping you mentally digest definitions, and for warning you when you are about to make an error.)

At this point, the only thing that you are officially allowed to know concerning the natural numbers is what is expressed in the following axioms. They function partially as descriptions of the primitive terms, and partially as a list of facts that we can use in later proofs. These axioms are called the *Dedekind-Peano axioms* since they are based on the axioms of the German mathematician Richard Dedekind (1831 – 1916) and the the Italian mathematician Giuseppe Peano (1858 – 1932).[4]

We begin with the primitive terms described in the axioms. They are called *primitive* because they do not have to be formally defined, but instead are described in the axioms. All other terms, such as $+$ or $<$ must be defined. Such definitions can build on primitive terms, notions from Chapter 0, or any previously defined term.

Primitive Terms. The three primitive terms are $\mathbb{N}$, 0, and σ.

The axioms then tell you everything you are allowed to assume about the meaning of these terms. For example, the first axiom tells you broadly what type of object these terms denote.

Axiom 1. (i) $\mathbb{N}$ is a set, (ii) 0 is an element of $\mathbb{N}$, and (iii) σ is a function

$$\sigma : \mathbb{N} \to \mathbb{N}$$

with domain and codomain equal to $\mathbb{N}$.

We call $\mathbb{N}$ the "set of natural numbers", and we call its elements "natural numbers". We call 0 the "zero element", or just "zero". We call σ the "successor function". If $n \in \mathbb{N}$ we call σn the "successor of n." Informally, the successor of n is the next number following n. This is informal since we have not yet defined an order $<$ on $\mathbb{N}$.

Axiom 2. The image of $\sigma : \mathbb{N} \to \mathbb{N}$ does not contain 0:

$$\neg\Big(\exists\, n \in \mathbb{N},\ \ \sigma n = 0\Big).$$

In other words, 0 is not the successor of a natural number.

[4] There are several variations of these axioms. We use a version of what is sometimes called the *second-order Peano axioms* which allows the notion of subsets of $\mathbb{N}$. There is another, more elementary system called the *first-order* Peano axioms which does not quantify over sets of natural numbers. If you encounter the Peano axioms outside these notes, you might see the first order version with axioms that refer directly to addition and multiplication. In our second-order version the operations of addition and multiplication are not mentioned in the axioms, but must be defined in terms of the successor function.

Axiom 3. The function $\sigma : \mathbb{N} \to \mathbb{N}$ is injective.[5] In other words, distinct natural numbers have distinct successors.

$$\forall\, x, y \in \mathbb{N},\ \ x \neq y \implies \sigma x \neq \sigma y$$

or equivalently

$$\forall\, x, y \in \mathbb{N},\ \ \sigma x = \sigma y \implies x = y.$$

Axiom 4 (Induction)**.** Suppose S is a subset of $\mathbb{N}$ such that **(i)** $0 \in S$, and **(ii)** $n \in S$ implies $\sigma n \in S$ for arbitrary $n \in \mathbb{N}$. Then $S = \mathbb{N}$.

$$S \subseteq \mathbb{N} \ \wedge\ 0 \in S \ \wedge\ \Big(\forall n\, (n \in S \ \Rightarrow\ \sigma n \in S)\Big) \quad \Longrightarrow \quad S = \mathbb{N}$$

Informal Exercise 1. Go through the axioms one by one, and convince yourself that they do indeed hold for your conception of the natural numbers

$$0,\ 1,\ 2,\ 3,\ \ldots\ .$$

Informally think of σn as the next number after n, or as $n + 1$. Since the exercise is informal, you may appeal to your earlier knowledge of arithmetic, knowledge of which will be formally proved later in the course. If it helps to justify the Induction Axiom (Axiom 4), think about why there could not be a smallest natural number $m \notin S$ given (i) and (ii) are known for $S \subseteq \mathbb{N}$.

Remark 4. As discussed in the introduction, basic concepts related to logic, sets, functions, and equality are all taken as given. They constitute the *logical background* to the development while the Dedekind-Peano axioms above are what we take to be the first real *mathematical* assumptions.[6]

Remark 5. Since $\mathbb{N}$, 0, and σ are primitive, they do not have to be defined. We start with some undefined terms to avoid circular definitions. All that we know about these terms at the moment is what is set forth in the axioms. By taking $\mathbb{N}$ to be primitive, we are avoiding the question "what are the natural numbers really". Our answer is just that they are elements of $\mathbb{N}$ where $\mathbb{N}$ is some set satisfying the axioms. Mathematicians regard the question "what are the natural numbers really" as not a mathematical question but as a philosophical question. Such questions have actually played an important role in in the history of philosophy for thousands of years, and continue to be discussed in contemporary philosophy.

[5] The reader is expected to be familiar with the term *injective*, or the equivalent term *one-to-one*. These terms describe functions f that map distinct elements to distinct images.

[6] The real location of the line between logic and mathematics is an interesting philosophical issue with no one predominate answer. The line drawn here is convenient for our purposes.

Remark 6. Some authors, especially of older texts, view the natural numbers as starting with 1. The axioms are then written in terms of 1 instead of 0. It makes sense to begin with 1 from a historical point of view since it took many years for mathematicians to get comfortable with the number 0. So in some sense 0 is not as natural as the positive integers. On the other hand, one of the main reasons for developing the natural numbers is for counting the size, or cardinality, of finite sets. Today the empty set $\emptyset$ is in common use, and we need 0 to describe its cardinality.

Remark 7. You might have seen induction presented in a slightly different style than that given above. Our axiom is in terms of *sets*, but you may have seen induction described in terms of *properties* instead. Perhaps it was stated in terms of identifying a certain property or statement that you want to prove for all of $\mathbb{N}$ by (i) proving it for 0 (the base case) and (ii) assuming it for n (the inductive hypothesis) and then proving it for the next number after n (which we call σn, but it is commonly called $n+1$).

The two versions of induction, however, are really the same. To see why, we need to think about the connection between properties and sets. A basic fact of set theory is that every property of natural numbers defines a subset of $\mathbb{N}$. Showing that a property holds for all natural numbers is the same as showing that the corresponding subset S is all of $\mathbb{N}$. To see an example, consider the following problem from number theory (using ideas we haven't defined formally yet). Suppose you want to prove that every natural number is the sum of four squares. Then instead of using the property "n is the sum of four squares" for an inductive hypothesis, our version of induction (Axiom 4) would use the set

$$S = \{n \in \mathbb{N} \mid n \text{ is the sum of four squares}\}.$$

The base step is to show $0 \in S$. By the definition of S this actually amounts to showing that 0 is the sum of four squares ($0 = 0^2 + 0^2 + 0^2 + 0^2$), so it amounts to the same thing as the base case of the other form of induction. Next you need to establish (ii) by assuming $n \in S$ and showing $\sigma n \in S$. By definition of S this means that you assume that n is the sum of four squares (the inductive hypothesis), and somehow try to show that the successor σn is also the sum of fours squares (this is the hard part of the proof). Once the base step (i) and the inductive step (ii) have been established, the induction axiom shows that $S = \mathbb{N}$. In other words, all natural numbers are the sum of four squares. As this illustrates, using a set instead of a property is just a very minor change of outlook, the actual work is the same.

In a later chapter we will discuss another type of induction, strong induction, which is truly different from that described above. We will also discuss versions where 0 is replaced by other "base cases". Our first version of induction is an axiom (Axiom 4), but the later versions will be theorems. The later versions cannot be used until they are proved.

Remark 8. The induction axiom is more complicated than the others. There is a cleaner way of stating this axiom using the notion of "closed" which we now explain. If A is a subset of $\mathbb{N}$ then the image set $\sigma[A]$ is necessarily also a subset of $\mathbb{N}$ since σ is a function $\mathbb{N} \to \mathbb{N}$. The subset $A \subseteq \mathbb{N}$ is said to be *closed under successor* if $\sigma[A] \subseteq A$. In other words, σ cannot move you out of A: for all $n \in A$, we have $\sigma n \in A$.

Using this concept, we can express the axiom as follows:

If $A \subseteq \mathbb{N}$ contains 0 and is closed under successor then $A = \mathbb{N}$.

Informal Exercise 2. Describe three distinct subsets of $\mathbb{N}$ that are closed under σ but that are not all of $\mathbb{N}$. By the above remark, none of your examples can contain 0. This shows the importance of checking the "base case" since all of these satisfy (ii) but not (i) of Axiom 4. Hint: since this is informal you have available the formula $\sigma n = n + 1$ even though it has not been proved yet. Also, one of your examples can be the empty set.

We are ready for our first formal definition.

Definition 1. Define 1 as $\sigma 0$. Define 2 as $\sigma 1$.

Exercise 3. Give formal definitions of $3, 4, 5, 6, 7, 8, 9$. Now we have names for at least a few numbers. We will wait until Chapter 5 before we develop the familiar base ten notation for naming the rest of the natural numbers.

Remark 9. Symbolic names for numbers are called *numerals*. There is a difference between numbers and numerals since several names can refer to the same number. Suitably defined, IV and 4 refer to the same number, namely $\sigma(\sigma(\sigma(\sigma 0)))$. So 'IV' and '4' are two different names, or numerals, for the same number.[7]

Informal Definition 2 (Stroke numerals)**.** One can regard any string of strokes | as a numeral via the following rule: replace each stroke with σ and then apply these to 0. For example, |||||| is just $\sigma(\sigma(\sigma(\sigma(\sigma(\sigma 0)))))$. Thus '||||||' and '6' are two numerals, from two different systems, for the same number. The Babylonians and the Egyptians used stroke numerals for numbers up to nine, so would have used something like '||||||' for six (but on two rows). The Romans would have used 'VI' for the same number. The stoke numerals gives a way of naming all natural numbers. However, this convention is not very efficient or practical! Fortunately we will later develop the more efficient base ten positional notation.

[7] A random person on the street might think of numerals and numbers as the same thing. But numerals are symbols. If numbers are not symbols, what are they? This comes back to the philosophical question: *what are numbers really*? As mentioned above we sidestep this as follows: numbers are what the axioms postulate to exist. The axioms do not specify what they really are, they just specify some of their properties. Numerals, on the other hand, are names we give to the objects described by the axioms. In summary, the axioms supply the numbers, but we supply the numerals to refer to these numbers, and can do so any way we choose

We now end this section by using the axioms to study the concept of *predecessor*. While the successor was primitive and did not have to be defined, predecessor needs to be defined. It is defined in terms of σ:

Definition 3 (Predecessor)**.** Suppose $a, b \in \mathbb{N}$. We say that "a is a predecessor of b" if $\sigma a = b$. We say that "b has a predecessor in $\mathbb{N}$" if[8] there exists an $x \in \mathbb{N}$ such that x is a predecessor of b.

Next we see the first official theorem of the course. It is a simple proof by contradiction.

Theorem 1. *The natural number* 0 *does not have a predecessor in* $\mathbb{N}$*.*

Proof. Suppose otherwise that 0 has predecessor $x \in \mathbb{N}$. By Definition 3 we have $\sigma x = 0$. This contradicts Axiom 2. Thus 0 has no predecessor in $\mathbb{N}$. □

Now we see the first proof by induction. It is subtle in one respect. One might want S to be the set $\{x \in \mathbb{N} \mid x \text{ has a predecessor in } \mathbb{N}\}$ for the induction, but this definition of S does not contain 0. So Axiom 4 cannot be used! We do not yet have a form of induction that starts at 1 (we will establish such an induction later). So instead we just artificially put 0 in S by using $\{x \in \mathbb{N} \mid (x = 0) \vee (x \text{ has a predecessor in } \mathbb{N})\}$. We only use this trick when we want to prove something about everything but 0.

Theorem 2. *Every nonzero element of* $\mathbb{N}$ *has a predecessor in* $\mathbb{N}$*.*

Proof. Our goal is to use the induction axiom (Axiom 4). To do so we need to define a set:

$$S \stackrel{\text{def}}{=} \{x \in \mathbb{N} \mid (x = 0) \vee (x \text{ has a predecessor in } \mathbb{N})\}.$$

Observe (i) $0 \in S$ by definition of S.

Next we will establish that (ii) $n \in S \implies \sigma n \in S$ for all $n \in \mathbb{N}$. So assume $n \in S$. Since $\mathbb{N}$ is the codomain of σ we have $\sigma n \in \mathbb{N}$. Observe that n is a predecessor of σn by Definition 3, so σn has a predecessor. Thus $\sigma n \in S$ by definition of S.

Now that we have established (i) and (ii) above, we can use Axiom 4 to conclude that $S = \mathbb{N}$. Since $S = \mathbb{N}$, every element of $\mathbb{N}$ is either 0 or has a predecessor in $\mathbb{N}$. So if $n \in \mathbb{N}$ and $n \neq 0$ we have that n has a predecessor in $\mathbb{N}$. □

We now consider the question of uniqueness. Successors are unique simply because they are values of a function. On the other hand we did not define predecessors as values of a function. Note $\exists!$ denotes "there exists a unique."[9]

[8] In *definitions* "if" really means "if and only if" in common mathematical writing.

[9] We do not use ! by itself to mean "unique". The use of the exclamation mark to mean "unique" is only used after $\exists$.

Exercise 4. Prove that if $n \in \mathbb{N}$ has a predecessor in $\mathbb{N}$ then the predecessor is unique. In other words, show the following for all $b \neq 0$ in $\mathbb{N}$:

$$\exists!\ a \in \mathbb{N},\ \ a \text{ is a predecessor of } b.$$

Hint: use Axiom 3.

Definition 4 (Positive)**.** A *positive natural number* is a nonzero element of $\mathbb{N}$. Let $\mathbb{N}^+$ be the set of positive natural numbers.

The following is an immediate consequence of the above theorem, exercise, and definition.

Corollary 3. *If $n \in \mathbb{N}^+$ then n has a unique predecessor in $\mathbb{N}$.*

Definition 5 (Predecessor function)**.** We define the *predecessor function*

$$\pi : \mathbb{N}^+ \to \mathbb{N}$$

as follows: given $n \in \mathbb{N}^+$ we define πn to be the unique predecessor of n.

In one sense the predecessor function and the successor function are inverses since one undoes the effect of the other. However this cannot be literally true. Since π is a function $\mathbb{N}^+ \to \mathbb{N}$, its inverse (if it exists) must be a function $\mathbb{N} \to \mathbb{N}^+$. To deal with this technicality we define a *modified successor function.*

Definition 6 (Modified successor)**.** We define the modified *successor function* $\sigma' : \mathbb{N} \to \mathbb{N}^+$ as follows: Given $n \in \mathbb{N}$ we define $\sigma' n$ to be σn. Since σn is not 0 (Axiom 2), and since $\sigma' n$ is just σn, we know that $\sigma' n$ is in the set $\mathbb{N}^+$. So this definition yields a function with codomain $\mathbb{N}^+$ as desired.

Observe that $\sigma' n = \sigma n$ for all $n \in \mathbb{N}$. The only difference between the functions is the codomain.

Exercise 5. Let $a \in \mathbb{N}$ and $b \in \mathbb{N}^+$. Show that $\pi b = a$ if and only if $\sigma' a = b$.

Exercise 6. Let $a \in \mathbb{N}$. Show that $\pi(\sigma' a) = a$. Hint: let $b = \sigma' a$ and substitute for b in the above exercise.

Exercise 7. Let $b \in \mathbb{N}^+$. Show that $\sigma'(\pi b) = b$.

Exercise 8. Show that π and σ' are inverse functions. Conclude that they are both bijections.

Hint: recall that $f : A \to B$ and $g : B \to A$ are called *inverse functions* if (i) $g(f(x)) = x$ for all $x \in A$, and (ii) $f(g(y)) = y$ for all $y \in B$. Recall also that a function is a bijection if it is both injective and surjective. Finally, recall that a function $f : A \to B$ is bijective if and only if it has an inverse function (from B to A).

Exercise 9. We know that σ' is bijective. Show that σ is not a bijection.

Exercise 10. The mathematician Dedekind defined a set S to be *infinite* if there is a bijection $S \to T$ where T is a proper subset of S. Explain why $\mathbb{N}$ is infinite according to Dedekind's definition.[10] (We will give another definition of infinite in Chapter 3).

Exercise 11. Show that if $n \in \mathbb{N}$ then $n \neq \sigma n$. Do so by defining a certain set $S \subseteq \mathbb{N}$ and using the induction axiom to show $S = \mathbb{N}$.

In particular this shows that $0 \neq 1$, and $1 \neq 2$, and so on. It does not mean $0 \neq 2$ though, this has to be proved separately!

1.3 Iteration

At this point the only operations we have are successor and predecessor. But any self-respecting theory of arithmetic also needs addition and multiplication. Our strategy for developing these operations is simple: we define addition in terms of iterated successor, and multiplication in terms of iterated addition. Continuing on, we will define exponentiation in terms of iterated multiplication. These definitions all rely on the general concept of *iteration*, so in order to reach our goal of basic arithmetic, we need to take a side trip through iteration.

Informally, we can think of iteration in terms of repeating an action or processes.[11] In these notes we think of operations, actions, processes, and such in terms of functions. So iteration will mean repeatedly applying a function.

For example, applying the function $f : S \to S$ twice to an element $x \in S$ yields $f(f(x))$, which is the same as appying the composition $f \circ f$ once to x. Likewise, applying $f \circ f \circ f$ to x gives the third iteration, and so on. We see that there is a close relationship between repeated composition and iteration. Note that in order to be able to compose a function with itself, it must have a codomain that matches its domain. So we want f to be a function $S \to S$ for some set S. In summary:

Informal Definition 7. Let $f : S \to S$ be a function. Observe that we are restricting ourselves to a function whose domain and codomain agree. The *second iteration* f^2 is $f \circ f$, the *third iteration* f^3 is $f \circ f \circ f$, the *fourth iteration* f^4 is $f \circ f \circ f \circ f$, and so on. In general, if $n \geq 2$ is a natural number, the nth iteration f^n is obtained by composing f with itself n times.

Remark 10. This is just an informal definition because some ideas in it, such as "composing f with itself n times", have not been formally defined.

Remark 11. We assume that the reader is already knowledgeable about composition of functions (Chapter 0). Recall that $f \circ g$ is only defined if

[10] A *proper* subset of S is a subset that is not equal to S.

[11] The verb *iterate* comes from the Latin verb *itero* meaning 'repeat'.

the codomain of g is equal to the domain of f. Another important fact: composition is associative (when it is defined):

$$f \circ (g \circ h) = (f \circ g) \circ h.$$

This fact allows us to drop parenthesis without introducing ambiguity. So the expression $f \circ g \circ h$ can refer to either $f \circ (g \circ h)$ or $(f \circ g) \circ h$, but both possibilities are equal by the associative law for composition.

What is f^n if n is 0 or 1? Informally, it makes sense to define f^1 as f itself since if you apply this function 1 time to an x in the domain, you get $f(x)$. What if you apply f to x zero times? You will just have x. So it makes sense, informally speaking, to define f^0 as the identity function.

Here is the formal axiom:

Axiom 5 (Iteration)**.** Let $f : S \to S$ be a function from a set S to itself, and $n \in \mathbb{N}$. Then the nth iteration of f is a function from S to itself. We write the nth iteration of f as $f^n : S \to S$. Such functions satisfy the following: (i) f^0 is the identity function on S, and (ii) $f^{\sigma n} = f \circ f^n$.

Remark 12. Here iteration is regarded as a primitive notion. In a later section (Section 1.9), however, we will see that there is a way in which the nth iteration can be *defined* and the the properties (i) and (ii) *proved*. Thus, for those willing to do some extra work, the above can be converted from an axiom to a theorem.

We have decided to move the proof to Section 1.9 because it is fairly long and a bit tricky, and because we want to get to basic arithmetic as soon as possible. Ultimately, however, it is an "eliminatable axiom".

Exercise 12. Use this axiom to prove that

$$f^1 = f, \quad f^2 = f \circ f, \quad f^3 = f \circ (f \circ f).$$

Informal Exercise 13. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by the rule $x \mapsto 3x$. Give a formula for the fifth iterate. In other words, describe f^5. What is g^3 if $g : \mathbb{R} \to \mathbb{R}$ is defined by the formula $g(x) = 2x^2 + 1$? Here $\mathbb{R}$ is the set of real numbers (to be developed later in the course).

Exercise 14. Let $f : S \to S$ be given. Prove that f^2 is the identity function on S if and only if $f = f^{-1}$.

Note: Here we are using '-1' as a symbolic expression to mark the inverse function; it does not yet refer to a number. We will not define negative numbers until Chapter 4. Recall that f^{-1} is defined to be the inverse function of f, which exists if and only if f is bijective.

Exercise 15. Prove that $\sigma^3(2) = 5$ where $\sigma : \mathbb{N} \to \mathbb{N}$ is the successor function.

Informal Exercise 16. What is $\sigma^n(m)$?

Informal Exercise 17. Propose an informal definition of addition in terms of iteration of the successor function. Discuss how multiplication can be explained of in terms of iteration of addition, and how exponentiation can be explained in terms of iteration of multiplication.

1.4 Addition

As mentioned above, we define addition in terms of iteration of successor. Informally, you get $m + n$ by starting with m and taking the successor n times. This idea motivates the formal definition.

Definition 8 (Addition)**.** Let $m, n \in \mathbb{N}$. Let $\sigma^n : \mathbb{N} \to \mathbb{N}$ be the nth iteration of the successor map. Then

$$m + n \stackrel{\text{def}}{=} \sigma^n(m).$$

Observe that addition defines a function $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$.

Remark 13. Functions $S \times S \to S$ are called *binary operations*. Thus $+$ is a binary operation on $\mathbb{N}$.

Remark 14. This is not the only way of viewing addition. In Chapter 3, we will show how $+$ can be understood in terms of counting the elements in a disjoint union.

The following are consequences of the iteration axiom and Definition 8.

Theorem 4. *For all* $m \in \mathbb{N}$

$$m + 0 = m.$$

Proof. By definition $m + 0 = \sigma^0(m)$. Recall that σ is a function $\mathbb{N} \to \mathbb{N}$. By the iteration axiom, σ^0 is the identity $i : \mathbb{N} \to \mathbb{N}$. Thus $\sigma^0(m) = m$ by definition of identity function. So $m + 0 = m$ by transitivity of equality. □

Lemma 5. *For all* $m, n \in \mathbb{N}$

$$m + \sigma n = \sigma(m + n).$$

Exercise 18. Prove the above lemma.

Remark 15. As mentioned earlier a lemma is a kind of theorem whose purpose in life is to help prove more important theorems. The above result is relegated to the role of lemma not because it is not of independent interest, but because it will be superseded by a more general theorem (the associative law of addition), so its usefulness is only temporary.

Lemmas are not always simple. In fact, many times in mathematics a lemma will be more complicated to state or harder to prove than the main theorem. Part of the art of mathematics is to decide what lemmas to prove in order to make the proofs of the important theorems as clear and elegant as possible.

Remark 16. Many authors define addition in terms of recursion instead of iteration of successor. The above theorem and lemma are the two recursion conditions used in this approach.

Informally we know that successor σ is just addition by one. The following makes this official:

Theorem 6. *For all* $m \in \mathbb{N}$

$$m + 1 = \sigma m.$$

Exercise 19. Prove the above theorem.

Remark 17. From now we can replace σm with $m + 1$ whenever we want. Based on the above theorem, these two expressions are completely interchangeable.

Exercise 20. Use the above theorem to prove that $1 + 1 = 2$.

Exercise 21. Prove that $2 + 2 = 4$. Prove $2 + 3 = 5$. Prove $3 + 2 = 5$.

Now we come to the first major theorem of the chapter.

Theorem 7 (Associative Law)**.** *For all* $x, y, z \in \mathbb{N}$

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

Proof. Fix $x, y \in \mathbb{N}$, and let $S_{x,y} \subseteq \mathbb{N}$ be the set of $z \in \mathbb{N}$ with the property that $x + (y + z) = (x + y) + z$.

First we observe that $0 \in S_{x,y}$ since, by Theorem 4 (twice),

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

Now assume $z \in S_{x,y}$. By Lemma 5 (several times) and our assumption,

$$\begin{aligned} x + \big(y + \sigma z\big) &= x + \sigma(y + z) \\ &= \sigma\big(x + (y + z)\big) \\ &= \sigma\big((x + y) + z\big) \\ &= (x + y) + \sigma z. \end{aligned}$$

So $\sigma z \in S_{x,y}$.

By the induction axiom, $S_{x,y} = \mathbb{N}$. This is true for any $x, y \in \mathbb{N}$. So if $x, y, z \in \mathbb{N}$ are arbitrary, $z \in S_{x,y}$ which implies $x+(y+z) = (x+y)+z$. □

Remark 18. This proof by induction is valid, but, like many induction proofs, is weak on conveying an understanding *why* associativity is true. In Chapter 3 we give a second, more insightful proof involving the set theoretic identity $A \cup (B \cup C) = (A \cup B) \cup C$.

Warning: we do not yet have the commutative law. Thus the next two lemmas are not redundant. They do not merely repeat Theorems 4 and 6, but assert something truly new. They are lemmas since, once the commutative law is proved, they will become redundant. So they are only of temporary use.

Lemma 8. *If $n \in \mathbb{N}$ then $\sigma^n(0) = n$. In particular $0 + n = n$.*

Lemma 9. *If $n \in \mathbb{N}$ then $1 + n = \sigma n$.*

Proof. Let $S = \{x \in \mathbb{N} \mid 1 + x = \sigma x\}$. So $0 \in S$ since $1 + 0 = 1 = \sigma 0$.

Suppose $n \in S$.

$$\begin{aligned} 1 + \sigma n &= 1 + (n + 1) \\ &= (1 + n) + 1 \\ &= \sigma n + 1 \\ &= \sigma(\sigma n). \end{aligned}$$

So $\sigma n \in S$.

We conclude that $S = \mathbb{N}$. □

Exercise 22. Prove Lemma 8. Complete the above sketchy proof of Lemma 9 by justifying every step by referring to earlier results, definitions, assumptions, or axioms, or by referring to the definition of S.

Theorem 10 (Commutative Law)**.** *If $x, y \in \mathbb{N}$ then*

$$x + y = y + x.$$

Proof. Fix $x \in \mathbb{N}$, and let $S_x = \{u \in \mathbb{N} \mid x + u = u + x\}$. By Theorem 4 and Lemma 8, we get $0 \in S_x$.

Now assume $n \in S_x$. So

$$\begin{aligned} x + \sigma n &= \sigma(x + n) \\ &= \sigma(n + x) \\ &= 1 + (n + x) \\ &= (1 + n) + x \\ &= \sigma n + x \end{aligned}$$

We conclude that $\sigma n \in S_x$.

By the induction axiom, $S_x = \mathbb{N}$. This is true of any $x \in \mathbb{N}$ since x was chosen to be any arbitrary element of $\mathbb{N}$. Now let $x, y \in \mathbb{N}$ be any two elements of $\mathbb{N}$. Since $S_x = \mathbb{N}$, we have $y \in S_x$. By definition of S_x we conclude that $x + y = y + x$. □

Remark 19. In Chapter 3 we see a more insightful proof of the commutative law involving the set theoretic identity $A \cup B = B \cup A$.

Exercise 23. Justify every step in the above proof by referring to earlier results, assumptions, or axioms, or by referring to the definition of S_x.

Exercise 24. Prove $(x + y) + z = (x + z) + y$ without using induction.

1.5 Multiplication

As mentioned above, our strategy for defining multiplication is to use iteration of addition. To understand, how this works, first consider the following familiar informal definition:

$$m \cdot n \quad = \quad \underbrace{m + m + \cdots + m + m}_{n \text{ times}}.$$

We can interpret the phrase "n times" in terms of iteration. To see this, notice how we can build up to this sum in n steps:

Step 1:	Add m to 0:	$0 + m$
Step 2:	Add m to previous result:	$(0 + m) + m$
Step 3:	Add m to previous result:	$\big((0 + m) + m\big) + m$
	$\vdots$	

Observe that we are just iterating the function $x \mapsto x + m$ as we go through the steps: every step involves applying $x \mapsto x + m$ where x is the result of the previous step. Observe also that the nth step results in $m \cdot n$, and that we start with $x = 0$ in the first step. (If we started with $x = m$, which might seem more natural, we would only use $n-1$ steps. We prefer to take exactly n steps, so we want to start at $x = 0$). We call the function $x \mapsto x + m$ the "addition by m" function or the "translation by m" function, and we write it as α_m. Multiplication is obtained by iterating α_m. For the product $m \cdot n$, we iterate n times.

This informal discussion motivates the following formal definition:

Definition 9 (Multiplication)**.** Let $m, n \in \mathbb{N}$. Let $\alpha_m : \mathbb{N} \to \mathbb{N}$ be defined by the rule $x \mapsto x + m$, and let α_m^n be the nth iteration of α_m. Then

$$m \cdot n \stackrel{\text{def}}{=} \alpha_m^n(0).$$

In particular, multiplication defines a binary operation $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$. As is common, we do not always need to write the dot $\cdot$, but can use juxtaposition to indicate multiplication.

Remark 20. This is not the only way of viewing multiplication. In Chapter 3, we will show how multiplication can be thought of in terms of counting the elements in the Cartesian product of two finite sets. Another popular approach is through recursion (using the equations of Theorem 11 and Lemma 12).

Exercise 25. Prove the following theorem, lemma, and theorem using the iteration axiom and the definition of multiplication.

Theorem 11. *For all* $m \in \mathbb{N}$

$$m \cdot 0 = 0.$$

Lemma 12. *For all* $m, n \in \mathbb{N}$

$$m \cdot \sigma n = (m \cdot n) + m.$$

Theorem 13. *For all* $m \in \mathbb{N}$

$$m \cdot 1 = m.$$

Exercise 26. We already have proved that $2 + 2 = 4$ and $3 + 2 = 5$ (Exercise 21). Use the results concerning addition to give proofs of the following: $4 + 2 = 6$, $5 + 2 = 7$, $6 + 2 = 8$, $7 + 2 = 9$, $3 + 3 = 6$, $4 + 3 = 7$, $5 + 3 = 8$, $6 + 3 = 9$, $4 + 4 = 8$, $5 + 4 = 9$. Use these addition facts together with Theorems 11 and 13 and Lemma 12 to show the following: $0 \cdot 0 = 0$, $0 \cdot 1 = 0$, $0 \cdot 2 = 0$, $1 \cdot 1 = 1$, $1 \cdot 2 = 2$, $2 \cdot 0 = 0$, $2 \cdot 1 = 2$, $2 \cdot 2 = 4$, $2 \cdot 3 = 6$, $2 \cdot 4 = 8$, $3 \cdot 2 = 6$, $3 \cdot 3 = 9$.

Theorem 14 (Distributive Law: part 1)**.** *For all* $x, y, z \in \mathbb{N}$

$$(x + y)z = xz + yz.$$

Remark 21. We adopt the usual conventions for dropping parentheses. Thus, when the parentheses and the dots are restored, the above equation is

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

Exercise 27. Prove the distributive law. Do so by defining, for any fixed $x, y \in \mathbb{N}$, a set $S_{x,y} \subseteq \mathbb{N}$. Show that $S_{x,y} = \mathbb{N}$ by the axiom of induction. In order to give a complete and rigorous proof, *do not leave any parentheses out in this proof.*

Remark 22. This induction proof is valid but, like many induction proofs, weak on conveying an understanding *why* the result is true. In Chapter 3 we will see a second proof using the set theoretic identity

$$(A \cup B) \times C = (A \times C) \cup (B \times C).$$

Lemma 15. *If $n \in \mathbb{N}$ then $0 \cdot n = 0$.*

Lemma 16. *If $n \in \mathbb{N}$ then $1 \cdot n = n$.*

Exercise 28. Prove the above two lemmas using the induction axiom.

Theorem 17 (Commutative Law)**.** *For all $x, y \in \mathbb{N}$*

$$xy = yx.$$

Proof. Fix $x \in \mathbb{N}$. Let $S_x = \{u \in \mathbb{N} \mid xu = ux\}$. We wish to show $y \in S_x$. We do so by showing all natural numbers are in S_x (via induction).

By Theorem 11 and Lemma 15, we get $0 \in S_x$.

Now assume $n \in S_x$. Then

$$\begin{aligned} x \cdot \sigma n &= xn + x \\ &= nx + x \\ &= n \cdot x + 1 \cdot x \\ &= (n+1) \cdot x \\ &= \sigma n \cdot x. \end{aligned}$$

We conclude that $\sigma n \in S_x$.

By the induction axiom, $S_x = \mathbb{N}$. Thus $y \in S_x$ which implies $xy = yx$. □

Exercise 29. Justify every step in the above proof.

Remark 23. In Chapter 3 we give a more insightful proof involving the natural bijection from $A \times B$ to $B \times A$.

Corollary 18 (Distributive Law: part 2)**.** *For all $x, y, z \in \mathbb{N}$*

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

Exercise 30. Prove the above corollary using the commutative law, and without using induction.

Exercise 31. Try to prove the following without looking at the given proof. If you get stuck, take a short peek at the proof for ideas. Now compare your proof to the given proof. Justify every step in the given proof.

Theorem 19 (Associative Law)**.** *For all $x, y, z \in \mathbb{N}$*

$$x(yz) = (xy)z.$$

Proof. Let $S_{x,y} = \{u \in \mathbb{N} \mid x(yu) = (xy)u\}$.

First we check that $0 \in S_{x,y}$. This follows from Theorem 11.

Now assume $n \in S_{x,y}$. So

$$\begin{aligned} x\big(y(n+1)\big) &= x\big((y \cdot n) + (y \cdot 1)\big) \\ &= x\big((yn) + y\big) \\ &= \big(x(yn)\big) + (xy) \\ &= \big((xy)n\big) + \big((xy)1\big) \\ &= (xy)(n+1) \end{aligned}$$

So $\sigma n = n + 1$ is in $S_{x,y}$.

By the induction axiom, $S_{x,y} = \mathbb{N}$. In particular, $z \in S_{x,y}$. □

1.6 Exponentiation

Just as repeated addition gives multiplication, repeated multiplication gives exponentiation. In other words, you can define exponentiation via the iteration of a multiplication function. How we do this for exponentiation is similar to how we developed multiplication, so the details will be left to the reader.

Definition 10. Let $m, n \in \mathbb{N}$. Let $\mu_m : \mathbb{N} \to \mathbb{N}$ be defined by the rule

$$x \mapsto xm.$$

Let μ_m^n be the nth iteration of μ_m. Then

$$m^n \stackrel{\text{def}}{=} \mu_m^n(1).$$

Remark 24. One amusing aspect of our approach is that exponential notation is used for iteration (Section 1.3) before it is used in the traditional way for exponentiation itself (here in Section 1.6). This is a symptom of the large emphasis we place on functions and their iterates. Our convention is that when an exponent is used with a function it refers to iteration, but when it is used with a number it refers to exponentiation.

Informal Exercise 32. In contrast with the previous section, we start with 1 instead of 0 in our iterative definition. What would happen if we used 0 instead of 1 in Definition 10?

Remark 25. This is not the only way of viewing exponentiation. In Chapter 3, we will see how it can be defined in terms of counting the number of functions between two sets. It can also be defined using recursion.

Informal Exercise 33. Do you expect $(m, n) \mapsto m^n$ to be a commutative binary operation $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$? Do you expect it to be associative? If you said 'no' to either question, back up your answer with a counter-example.

Theorem 20. *For all* $m \in \mathbb{N}$

$$m^0 = 1.$$

Lemma 21. *For all* $m, n \in \mathbb{N}$

$$m^{\sigma n} = m^n \cdot m$$

Theorem 22. *For all* $m \in \mathbb{N}$

$$m^1 = m.$$

Exercise 34. Prove the above theorems and lemmas.

Exercise 35. Use the above theorems and lemmas to show the following: $0^0 = 1, \quad 2^2 = 4, \quad 2^3 = 8.$

Warning. Although the equation $0^0 = 1$ is valid in our current context, there are some parts of mathematics where 0^0 is regarded as undefined. This is related to the use of limits in calculus where we have to be careful with limits that converge to indeterminate expressions of the form $0/0$, ∞/∞, $\infty - \infty$, or even 0^0. Limits of expressions in indeterminate form do not consistently converge to any fixed value. In fact, some limits in indeterminate form diverge, and some converge, and those that converge do not all converge to the same value. The problem with limits with indeterminate form 0^0 is related to the fact that the function $f(x, y) = x^y$ is not continuous at $(0, 0)$. So in calculus and other contexts, 0^0 is often left undefined.

Theorem 23. *If* $x, y, n \in \mathbb{N}$ *then*

$$(xy)^n = x^n y^n.$$

Exercise 36. Prove the above using induction on n. In other words, apply the induction axiom to a certain set $S_{x,y}$.

Theorem 24. *If* $x, m, n \in \mathbb{N}$ *then*

$$x^{m+n} = x^m x^n.$$

Exercise 37. Prove the above using induction on n. In other words, apply the induction axiom to a certain set $S_{x,m}$.

Theorem 25. *If* $n \in \mathbb{N}$ *is not* 0 *then*

$$0^n = 0.$$

Exercise 38. Prove the above *without* induction. Use Theorem 2 to first show that $n = \sigma m$ for some m.

Theorem 26. *If* $n \in \mathbb{N}$ *then*

$$1^n = 1.$$

Theorem 27. *If* $x, n, m \in \mathbb{N}$ *then*

$$\left(x^m\right)^n = x^{mn}.$$

Exercise 39. Prove the above two theorems.

1.7 Other properties of addition

There are a few more facts about addition that we will need in the next chapter.[12]

Theorem 28. *If $m, n \in \mathbb{N}$ are such that $0 = m + n$ then $m = n = 0$.*

Proof. Suppose $n \neq 0$. Then $n = \sigma(x)$ for some $x \in \mathbb{N}$. Thus

$$0 = m + n = m + \sigma(x) = \sigma(m + x).$$

This contradicts Axiom 2. From this contradiction, we conclude $n = 0$.

From $n = 0$ and $0 = m + n$ we get $m = 0$ as well using Theorem 4. □

Exercise 40. Supply the missing justifications in the above proof, including every step in the chain of equalities

$$0 = m + n = m + \sigma(x) = \sigma(m + x).$$

By definition of addition, we know that $\mathbb{N}$ is closed under addition. In other words, if $a, b \in \mathbb{N}$ then $a+b \in \mathbb{N}$. The following shows that $\mathbb{N}^+$ is closed as well. Recall that $\mathbb{N}^+$ is the set of nonzero natural numbers (Definition 4).

Corollary 29. *The set $\mathbb{N}^+$ of positive natural numbers is closed under addition. In other words, if $m, n \in \mathbb{N}^+$ then $m + n \in \mathbb{N}^+$.*

Exercise 41. Prove the above corollary.

Now we turn our attention to the cancellation law for addition. Both addition and multiplication have cancellation laws. For addition the law states that if $x + z = y + z$ then $x = y$. In other words, we cancel z. Contrary to what one might think, the operation of subtraction is not needed to state or prove this law. For multiplication the law states that if $xz = yz$ and $z \neq 0$ then $x = y$. Note the extra condition $z \neq 0$. The cancellation law for multiplication will be proved in the next chapter.

First we consider the contrapositive form (which is a bit easier to prove):

Theorem 30. *Suppose $x, y \in \mathbb{N}$ are distinct: $x \neq y$. Then $x + z \neq y + z$ for all $z \in \mathbb{N}$.*

Proof. Fix $x, y \in \mathbb{N}$ distinct, and let $S_{x,y} = \{z \in \mathbb{N} \mid x + z \neq y + z\}$. Since $x + 0 = x$ and $y + 0 = y$, we have $x + 0 \neq y + 0$. So $0 \in S_{x,y}$.

Suppose that $n \in S_{x,y}$. We wish to show that $\sigma n = n + 1$ is in $S_{x,y}$, i.e., that $x + \sigma(n) \neq y + \sigma(n)$. Since $n \in S_{x,y}$ we have $x + n \neq y + n$. Since σ is injective, we have $\sigma(x + n) \neq \sigma(y + n)$. Observe that by Lemma 5

$$\sigma(x + n) = x + \sigma(n), \text{ and}$$

[12] Nothing in this section uses the results of Section 1.5 or 1.6.

$$\sigma(y+n) = y + \sigma(n).$$

Thus $x + \sigma(n) \neq y + \sigma(n)$. In other words, $\sigma(n) \in S_{x,y}$.

By the induction axiom, $S_{x,y} = \mathbb{N}$. Thus, $x + z \neq y + z$ for all $z \in \mathbb{N}$. □

Theorem 31 (Cancellation Law)**.** *Suppose $x, y, z \in \mathbb{N}$. Then*

$$x + z = y + z \qquad \textit{implies} \qquad x = y.$$

Proof. Suppose that $x + z = y + z$, but that $x \neq y$. By Theorem 30, this implies that $x + z \neq y + z$, a contradiction. □

In Exercise 11 we saw that $n \neq n + 1$. The following generalizes this to other sums. It can be used to show, for instance, that the natural numbers we have defined so far, $0, 1, 2, 3, 4, \ldots, 9$, are pairwise distinct.

Theorem 32. *Suppose $n, m, b \in \mathbb{N}$. If $n = m + b$ where $b \neq 0$ then $n \neq m$.*

Proof. Suppose $n = m + b$ where $b \neq 0$, but that $n = m$. So

$$b + m = m + b = n = m = 0 + m.$$

By the cancellation law $b = 0$, contradicting our hypothesis. □

1.8 The universal property of $\mathbb{N}$ (optional)

Earlier we adopted the iteration axiom and used it to help define and prove the basic properties of addition, multiplication, and such. However, a promise was made to show that the iteration axiom is not needed since it can be derived from the Peano axioms. In this section we prepare for a proof of the iteration axiom by making a careful study of iteration.

In order to avoid circularity we will appeal only to the Peano axioms, and not to any theorems proved with the assistance of the iteration axiom. In fact, this section and the next could be cut and pasted immediately after Section 1.2 with no loss of logical rigor. We did not do this since we felt that the development would go more smoothly if we *applied* iteration to define and prove things about addition, multiplication, and such before giving the more elaborate proofs justifying iteration.

Let's begin with an informal discussion about iteration. The raw materials of iteration consists of a set A and a function $s : A \to A$. If you choose a starting element $z \in A$ and repeatedly apply s you will get

$$z, \quad s(z), \quad s(s(z)), \quad s(s(s(z))), \quad s(s(s(s(z)))),$$

and so on. We can informally think of this as a sort of "path", and we can think of s as determining a "step". Metaphorically, you are starting with z and stepping along a path away from z.

An example of this is the definition of $m + n$ by iteration of $\sigma : \mathbb{N} \to \mathbb{N}$. In that case we start with $z = m$ and iterate σ a total of n times. In other words, we take n steps along the path to reach our goal of $m+n$. Here each step consists of applying σ to the previous result.

Fix $z \in A$ and $s\colon A \to A$ as above. The following theorem defines a function $\varphi : \mathbb{N} \to A$ that in some sense identifies the result of taking n steps starting from z. In other words,

$$\varphi(0) = z, \quad \varphi(1) = s(z), \quad \varphi(2) = s(s(z)), \quad \varphi(3) = s(s(s(z))),$$

and so on. This function will depend, of course, on z and s, and we could write it as $\varphi_{z,s}$ if we want to make this dependency clear. It turns out that the condition $\varphi \circ \sigma = s \circ \varphi$ is what is needed to force φ to be such a "stepping function" (see the corollary).

This informal discussion helps motivate the following theorem. It states that is is possible to find such a function φ.

Theorem 33 (Universal property of $\mathbb{N}$)**.** *Suppose A is a set, $z \in A$ is an element, and $s : A \to A$ is a function. Then there is a unique function $\varphi : \mathbb{N} \to A$ such that $\varphi(0) = z$ and $\varphi \circ \sigma = s \circ \varphi$.*

Proof. Let $\sigma_2 : \mathbb{N} \times A \to \mathbb{N} \times A$ be defined by the rule $(n, a) \mapsto (\sigma n, sa)$. (It is called σ_2 since it takes pairs to pairs). For a subset R of $\mathbb{N} \times A$, we say R is *z-closed* if (i) $(0, z) \in R$ and (ii) $\sigma_2[R] \subseteq R$. This second condition means that if $(n, a) \in R$ then $(\sigma n, sa)$ must be in R.

Observe also that the intersection of z-closed sets is z-closed. Observe that whole set $\mathbb{N} \times A$ is z-closed. Let N be the intersection of all z-closed sets (the *z-closure*), so N is itself closed. Since it is the intersection of all z-closed set, N is contained in any given z-closed set.

Since $(0, z) \in N$ and $\sigma_2[N] \subseteq N$, it follows that $\sigma_2[N] \cup \{(0, z)\} \subseteq N$. Check that $\sigma_2[N] \cup \{(0, z)\}$ is z-closed. So, by the minimality of N,

$$N = \sigma_2[N] \cup \{(0, z)\}.$$

We say $n \in \mathbb{N}$ is *paired-up* if N has a unique pair (n, a) with first coordinate n. In this case, we say that *a corresponds to n*. Claim: all $n \in \mathbb{N}$ are paired up. We prove this claim by induction.

Since $N = \sigma_2[N] \cup \{(0, z)\}$, it follows that 0 is paired-up: Axiom 2 implies that no pair of the form $(0, a)$ is in the image of σ_2. Also, z corresponds to 0.

Suppose that n is paired-up. Then there is a unique pair $(n, a) \in N$. Thus $(\sigma n, sa)$ is in N since N is z-closed. We now want to show $(\sigma n, sa)$ is the unique pair with first coordinate σn. So, suppose $(\sigma n, b)$ is also in N. Since $N = \sigma_2[N] \cup \{(0, z)\}$ and since $\sigma n \neq 0$ it follows that $(\sigma n, b) \in \sigma_2[N]$. So $(\sigma n, b) = (\sigma m, sc)$ for some pair $(m, c) \in N$. Since σ is injective, $m = n$.

So $(n, c) \in N$. Observe that (n, a) and (n, c) are in N. But n is paired-up, so, by uniqueness, $c = a$. Thus $(\sigma n, b) = (\sigma m, sc) = (\sigma n, sa)$. This concludes the argument for uniqueness and shows that σn is paired-up.

Let $S \subseteq \mathbb{N}$ be the subset of paired-up elements. By the induction axiom, $S = \mathbb{N}$. Thus every natural number is paired-up.

Let φ be defined by the rule $n \mapsto a$ where a corresponds to n. In other words, φn corresponds to n, so $(n, \varphi n) \in N$. Since N is z-closed we have $(\sigma n, s(\varphi n)) \in N$. Thus $s(\varphi n)$ corresponds to σn. So $\varphi(\sigma(n)) = s(\varphi(n))$ by the definition of φ. This holds for all $n \in \mathbb{N}$, so we have established that $\varphi \circ \sigma = s \circ \varphi$. Since z corresponds to 0, we have $\varphi(0) = z$. We have now established the existence of the desired φ.

We now show uniqueness. Suppose that $\varphi' : \mathbb{N} \to A$ is such that $\varphi'(0) = z$ and $\varphi' \circ \sigma = s \circ \varphi'$. We need to show that $\varphi = \varphi'$. If W is the set of $n \in \mathbb{N}$ such that $\varphi(n) = \varphi'(n)$, we need to show $W = \mathbb{N}$. We do this by induction.

Observe that $\varphi(0) = z = \varphi'(0)$, so $0 \in W$. Now assume that $n \in W$, so $\varphi(n) = \varphi'(n)$. Then $s(\varphi(n)) = s(\varphi'(n))$. However,

$$\varphi(\sigma n) = (\varphi \circ \sigma)(n) = (s \circ \varphi)(n) = s(\varphi(n)),$$

and

$$\varphi'(\sigma n) = (\varphi' \circ \sigma)(n) = (s \circ \varphi')(n) = s(\varphi'(n)).$$

So $\varphi(\sigma n) = \varphi'(\sigma n)$. In particular, $\sigma n \in W$.

By the induction axiom, $W = \mathbb{N}$. So, $\varphi = \varphi'$. □

Exercise 42. Let A, s, z, φ be as in the above theorem. Show the following

$$\varphi(0) = z, \quad \varphi(1) = s(z), \quad \varphi(2) = s(s(z)), \quad \varphi(3) = s(s(s(z))).$$

Informal Exercise 43. Describe φ in the case where $A = \{0, 1\}$, $z = 0$, and $s : A \to A$ is defined by the rule $0 \mapsto 1$ and $1 \mapsto 0$. What if $z = 1$ instead? What if s is the identity function instead?

Remark 26. This theorem is called the *universal* property of $\mathbb{N}$. To explain this, we need to discuss some ideas related to category theory.

In the last 60 years or so, mathematicians have become more concerned with the notion of *structure*. Roughly speaking, a structure is typically a set equipped with special relations, functions, binary operators, elements, and the like. The field of mathematics that is used to compare structures is *category theory*. We will not discuss category theory in general, but will illustrate some ideas of category theory in the context of the above theorem.

Recall that the first of the Peano axioms describes $\mathbb{N}$ as a set equipped with two things (i) a starting element $0 \in \mathbb{N}$ and (ii) a function (called the successor function) $\sigma : \mathbb{N} \to \mathbb{N}$, that can be used to "take steps" away from the starting point 0. Similarly, the set A in the theorem is given with a function $s : A \to A$ and a starting element z. We can view $\mathbb{N}$ and such a

set A as examples of a certain basic type of structure. Let's make up some fancy terminology and call such a structure a *path structure* since we indicate a starting point and from there can go on a path through the set by using the function to take steps. More precisely, a *path structure* is a set A equipped with (i) a *starting point* $z \in A$ and (ii) a *stepping function* $s : A \to A$.

A simple example of a path structure is the set $A = \{0, 1\}$ where we declare the starting point to be 0 and declare the stepping function to be the function $s : A \to A$ defined by the rule $0 \mapsto 1$ and $1 \mapsto 0$. Repeating s gives you the path $0, 1, 0, 1, 0, \ldots$. As you can see, we do not require that every path structure satisfy all the Peano axioms. For example, in this structure 0 *is* in the image of the stepping function.

We use different path structures for different situations. The path structure most appropriate for defining m^n is that given by taking the set $\mathbb{N}$, but declaring the starting point to be 1, and declaring the stepping function to be $\mu_m(x) = xm$. Following a path in this structure would give you $1, m, m^2, m^3, m^4, \ldots$.

The collection of all possible path structures forms something that mathematicians call a *category*.

Now, among all path structures, $\mathbb{N}$ is very special: Theorem 33 shows it maps (uniquely) to any other path structure in a special way. More specifically, given any other path structure A, there is a function $\varphi : \mathbb{N} \to A$ such that (i) $0 \mapsto z$ and (ii) $\varphi \circ \sigma = s \circ \varphi$. The first condition says that φ sends the start to the start. The second condition matches σ with s. We can illustrate the second condition with the following *commutative diagram*:

$$\begin{array}{ccc} \mathbb{N} & \xrightarrow{\varphi} & A \\ \downarrow{\scriptstyle \sigma} & & \downarrow{\scriptstyle s} \\ \mathbb{N} & \xrightarrow{\varphi} & A \end{array}$$

What this diagram expresses is that both ways of going from the top left set to the bottom right set gives the same image.[13] This diagram expresses the equation $\varphi \circ \sigma = s \circ \varphi$.

The existence of this special φ is called the *universal property of* $\mathbb{N}$. In other words, $\mathbb{N}$ has the universal property of being able to map uniquely into any other path structure (in a path compatible way).

Remark 27. The universal property (Theorem 33) is important for other reasons besides describing iteration. In fact, it makes it easy to show that any two models of the Peano axioms are "isomorphic". However, we will skip this important isomorphism theorem since explaining in precisely will lead us too far afield.

[13] In category theory the map φ is called a *morphism* or a *homomorphism* because it in some sense preserves the form ("morph") of the structures. Different categories have different types of morphisms. For example, in the category of vector spaces, the morphisms are linear maps: they preserve the linear structure.

1.9 Eliminating the iteration axiom (optional)

We now use Theorem 33 to show that the iteration axiom can be dispensed with. In other words, it can be proved as a theorem.

Theorem 34. *Let $f : S \to S$ be a function. Then one can assign to every $n \in \mathbb{N}$ a function $f^n : S \to S$ such that (i) f^0 is the identity function on S, and (ii) $f^{\sigma n} = f \circ f^n$.*

The idea behind this theorem is to use Theorem 33 to describe iteration. If $a \in S$ then we want to consider the iteration process giving

$$a, \quad f(a), \quad f(f(a)), \quad f(f(f(a))), \quad f(f(f(f(a)))),$$

and so on. So we apply the theorem to the case where $A = S$, $z = a$, and $s = f$. Then $f^n(a)$ is obviously $\varphi(n)$. We give the details below:

Proof. Define f^n by the rule $a \mapsto \varphi_a(n)$ where φ_a is the function φ described in Theorem 33 given by choosing $A = S$, $s = f$, and $z = a$. Observe that f^n sends elements $a \in S$ to elements of S.

Observe also that $f^0(a) = \varphi_a(0) = a$ (because $z = a$ in this case). Thus f^0 is the identity function.

Finally, use Theorem 33 to observe that

$$f^{\sigma n}(a) = \varphi_a(\sigma n) = f(\varphi_a(n)) = f(f^n(a)) = f \circ f^n(a)$$

(recall $s = f$). This holds for arbitrary $a \in S$, so $f^{\sigma n} = f \circ f^n$. □

There is another proof that is interesting (also based on Theorem 33, but with different choice of A, z and $s : A \to A$):

Proof. (Second proof) Let A be the set of functions from S to itself. Let z be the identity function on S. Let $s : A \to A$ be the map that sends a function g to $f \circ g$.

Let $\varphi : \mathbb{N} \to A$ be as in Theorem 33. Define f^n to be $\varphi(n)$. Since A consists of functions from S to itself, f^n maps S to S. Since $\varphi(0) = z$, we have that f^0 is the identity function on S. Finally, since $\varphi \circ \sigma = s \circ \varphi$,

$$f^{\sigma n} = \varphi(\sigma n) = \varphi\big(\sigma(n)\big) = s\big(\varphi(n)\big) = s(f^n) = f \circ f^n.$$

This completes the proof. □

Remark 28. There is also a uniqueness result. Let $maps(S, S)$ be the set of all functions $S \to S$ (written A in the second proof). Then the theorem describes the existence of a function $\mathbb{N} \to maps(S, S)$, given by $n \mapsto f^n$, that satisfies certain properties. The second proof above can be modified and extended to show the uniqueness of the function $\mathbb{N} \to maps(S, S)$ with the desired properties.

Chapter 2

The Natural Numbers $\mathbb{N}$ as an Ordered Set

In this chapter we continue the study of $\mathbb{N}$ with a focus on the definition of order $<$ and matters related to this order. For instance we will show that $\mathbb{N}$ is well-ordered. We will define subtraction in a manner related to the order. We will discuss the set $\{1, \ldots, n\}$ which requires the order on $\mathbb{N}$ to define. Finally, we will discuss recursion which, in a sense, also depends on order since a value of a function $f(n)$ is defined in terms of $f(m)$ for one or more $m < n$.

We start with the general concept of a strict linear order. Then we consider the specific order relation on $\mathbb{N}$, and then treat various topics related to this order.

2.1 Order relations

Several of the number systems considered in this course can be thought of as having a linear arrangement. Such number systems include $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$. In this section we defined the general notion of a linear order. In the next section we define the specific linear order for $\mathbb{N}$.

Definition 1. A *strict linear order* on a set S is a binary relation, commonly written with the symbol $<$, which satisfies the following two conditions.

1. *The transitivity law*: for all $x, y, z \in S$,

$$\text{if } x < y \text{ and } y < z \text{ then } x < z.$$

2. *The trichotomy law*: for all $x, y \in S$ exactly one of the following three holds
$$x < y, \qquad x = y, \qquad y < x.$$

Remark 1. When $x < y$ holds we say that x *is strictly less than* y. Strict linear orders are also called *strict total orders*, in contrast to a more general type of order relation important in mathematics called *partial orders*.

Remark 2. (Review) Order relations are one of the two types of binary relations used in this course. The other type are the *equivalence relations*. Equivalence relations are relations which are reflexive, symmetric, and transitive.

We assume that that reader is familiar with the general idea of a binary relation. Informally, a binary relation R on a set S is a condition linking elements of S. We write xRy to indicate that $x, y \in S$ are linked according to the relation R, and $\neg xRy$ to indicate otherwise. For each pair $(x, y) \in S \times S$ we either have xRy or $\neg\, xRy$.

In modern set theory, it is usual to think of a relation R on S as a subset of $S \times S$. For such a relation $R \subseteq S \times S$ we have $(x, y) \in R$ if and only if the relation holds between x and y. In other words, xRy just means that $(x, y) \in R$. If $(x, y) \notin R$ then we write $\neg\, xRy$ instead.

The convention of putting the symbol R between the elements x, y is called *infix notation*. We write xRy when using infix notation, versus writing $R(x, y)$ or $R\,x\,y$ using prefix notation. For equivalence relations, we often use symbols such as $\equiv$ or $\sim$ for the relation R. For strict linear orders the common symbol for the relation R is $<$, as in the definition above.

Exercise 1. Suppose that $<$ is a strict linear order on a set S. Show that this relation is *anti-reflexive*. In other words, show that $\neg(x < x)$ for all $x \in S$.

Exercise 2. There are other equivalent ways of defining linear orders. For example, instead of trichotomy, one could require "weak trichotomy" together with the anti-reflexive law. With this in mind, prove the following lemma.

Lemma 1. *Suppose $<$ is a binary relation on a set S such that the following three conditions hold:*

1. *The transitivity law: for all $x, y, z \in S$,*
$$\text{if } x < y \text{ and } y < z \text{ then } x < z.$$

2. *The weak trichotomy law: for all $x, y \in S$*
$$(x < y) \vee (x = y) \vee (y < x).$$

3. *The anti-reflexive law: for all $x \in S$*
$$\neg(x < x).$$

Given these all hold, then $<$ is a strict linear order on S.

Informal Exercise 3. What is the main difference between the trichotomy law and the weak trichotomy law? Which implies the other?

Definition 2. Let $<$ be a strict linear order on a set S. Then we define the associated nonstrict order relation $\leq$ as follows: For each $x, y \in S$ the relation $x \leq y$ is defined to hold if and only if

$$(x < y) \ \vee \ (x = y)$$

holds. When $x \leq y$ holds we say that x *is less than or equal to* y.

Exercise 4. Let $\leq$ be as in the above definition and $x, y \in S$. Show the following:

1. $(x \leq y) \vee (y \leq x)$.
2. If $x \leq y$ and $y \leq x$ then $x = y$.
3. $\neg(x \leq y)$ if and only if $y < x$.
4. $\neg(x < y)$ if and only if $y \leq x$.

Exercise 5. From the transitivity of $<$ we can derive other transitivity laws. Do this by proving the following two theorems.

Theorem 2 (mixed transitivity)**.** *Suppose that $<$ is a strict linear order on a set S. Let $\leq$ be as in Definition 2. Then*
(i) For all $x, y, z \in S$, if $x < y$ and $y \leq z$ then $x < z$.
(ii) For all $x, y, z \in S$, if $x \leq y$ and $y < z$ then $x < z$.

Theorem 3 (transitivity of $\leq$)**.** *Suppose that $x, y, z \in S$ and that $<$ is a strict linear order on a set S. Let $\leq$ be as in Definition 2. If $x \leq y$ and $y \leq z$ then $x \leq z$.*

Definition 3. Suppose that $<$ is a strict linear order on a set S. Then we define $>$ as follows: for all $x, y \in S$ the relation $x > y$ holds if and only if $y < x$ holds. In this case we say that x *is strictly greater than* y.

We define $\geq$ as follows: for all $x, y \in S$ the relation $x \geq y$ holds if and only if $(x > y) \vee (x = y)$ holds. In this case we say that x *is greater than or equal to* y.

Remark 3. Since $<$ and $\leq$ are transitive, it follows easily that $>$ and $\geq$ are transitive as well.

Definition 4. The notation $a < b < c$ is short for $(a < b) \wedge (b < c)$. Observe that by transitivity $a < b < c$ implies $a < c$.

A similar notation is adopted for $>$, $\leq$, and $\geq$. We adopt a similar notation also for more than three terms, and we sometimes combine $<$ and $\leq$ (or $>$ and $\geq$). Thus $a \leq b < c < d$ is short for $(a \leq b) \wedge (b < c) \wedge (c < d)$.

2.2 The standard order relations on $\mathbb{N}$

Our definition of order for $\mathbb{N}$ is based on addition:

Definition 5. We define the binary relation $<$ on $\mathbb{N}$ by the following rule: Given $m, n \in \mathbb{N}$, we have $m < n$ if and only if there is a nonzero $b \in \mathbb{N}$ such that $n = m + b$. In symbols:

$$m < n \iff \exists b \in \mathbb{N}\big((b \neq 0) \wedge (n = m + b)\big).$$

We get two quick consequences of this definition, which we state for future reference.

Theorem 4. *If $x, y \in \mathbb{N}$ and if $y \neq 0$ then $x < x + y$.*

Theorem 5. *If $m \in \mathbb{N}$ then $m < m + 1$.*

Exercise 6. Give short proofs of the above two theorems. Hint: do we know that $1 \neq 0$?

Theorem 6 (transitivity of $<$)**.** *Suppose $x, y, z \in \mathbb{N}$. If $x < y$ and $y < z$ then $x < z$.*

Proof. Since $x < y$, there is a $b \neq 0$ such that $y = x + b$, and since $y < z$ there is a $c \neq 0$ such that $z = y + c$. So

$$z = y + c = (x + b) + c = x + (b + c).$$

In Chapter 1 we proved that $b + c \neq 0$. Therefore, $x < z$ by Definition 5 □

Lemma 7. *If $n \in \mathbb{N}$ then $(0 < n) \vee (0 = n)$.*

Proof. If $n = 0$ then the result holds. So assume $n \neq 0$. In this case $0 < 0+n$ by Theorem 4. Then since $0 + n = n$, we get $0 < n$. □

We now give the proof of the weak trichotomy law. Its proof is a fairly tricky use of induction.

Lemma 8 (Weak trichotomy)**.** *Suppose $m, n \in \mathbb{N}$. Then*

$$(m < n) \vee (m = n) \vee (n < m).$$

Proof. Let $n \in \mathbb{N}$ be fixed. Let S_n be the set of elements $x \in \mathbb{N}$ that satisfy the following condition:

$$(x < n) \vee (x = n) \vee (n < x).$$

In other words, S_n is the set of all m for which the lemma holds (with fixed n).

The base case $0 \in S_n$ follows from Lemma 7.

Now suppose that $k \in S_n$. We wish to show $\sigma k \in S_n$. Since $k \in S_n$ we have three cases: (1) $k < n$, (2) $k = n$, and (3) $n < k$.

Case 1: $k < n$. By definition $n = k + b$ for some $b \neq 0$. Since b is not 0, it has a predecessor $c \in \mathbb{N}$. So $b = c + 1$ and

$$n = k + b = k + (c + 1) = k + (1 + c) = (k + 1) + c = \sigma k + c.$$

If $c = 0$ then $\sigma k = n$, so $\sigma k \in S_n$. If $c \neq 0$ then $\sigma k < n$ by Definition 5. so $\sigma k \in S_n$. So whatever c is, we have $\sigma k \in S_n$.

Case 2: $k = n$. Observe that $k < k + 1$ by Theorem 5. So $n < \sigma k$ since $k = n$ and $\sigma k = k + 1$. So $\sigma k \in S_n$.

Case 3: $n < k$. Observe that $k < k + 1$ by Theorem 5. So $n < k + 1$ by the transitivity of $<$. In other words $n < \sigma k$. Thus $\sigma k \in S_n$.

By the induction axiom, $S_n = \mathbb{N}$. This is true of arbitrary $n \in \mathbb{N}$. So for any $m, n \in \mathbb{N}$, we have $m \in S_n$. The result follows. □

Exercise 7. Use the cancellation law for addition (or related properties) to show the following: if $n \in \mathbb{N}$ then $\neg(n < n)$. Now use this fact, together with transitivity and weak trichotomy to conclude the following:

Theorem 9. *The relation $<$ is a strict linear order on $\mathbb{N}$.*

Remark 4. We define $>$, $\leq$, and $\geq$ in terms of $<$ as in the previous section.

We can now show that 0 is the least element of $\mathbb{N}$:

Theorem 10. *If $n \in \mathbb{N}$ then $0 \leq n$.*

Proof. This follows from the definition of $\leq$ and Lemma 7. □

Theorem 11. *Suppose $m, n \in \mathbb{N}$. Then*

$$m \leq n \iff \exists b \in \mathbb{N}\,(n = m + b).$$

Proof. First suppose $m \leq n$. We wish to find a b such that $n = m + b$. By Definition 2, we have two cases: (i) $m < n$ and (ii) $m = n$. In case (i) the existence of b follows from Definition 5. In case (ii) we can take $b = 0$. In either case, we have a suitable $b \in \mathbb{N}$.

Next suppose $n = m + b$ for some b. We must show that $m \leq n$. If $b = 0$ then $m = n$, which implies $m \leq n$ by Definition 2. If $b \neq 0$ then $m < n$ by Definition 5. Thus $m \leq n$ by Definition 2. □

2.3 Basic properties of the order on ℕ

In Chapter 1 we defined the set $\mathbb{N}^+$ of positive natural numbers in terms of the condition $n \neq 0$. However, most people use the condition $n > 0$. Both work for $\mathbb{N}$, but in fact $n > 0$ is the right condition for other number systems. We didn't use the condition $n > 0$ in Chapter 1 because the relation $>$ had not been defined yet. We now show that both conditions are equivalent for $\mathbb{N}$, so it doesn't matter which you use to define positive natural numbers.

Theorem 12. *Suppose $n \in \mathbb{N}$. Then $n \neq 0$ if and only if $n > 0$.*

Proof. Part 1. Suppose that $n \neq 0$. Observe that $n = 0 + b$ where $b = n$. So $b \neq 0$. Thus, by Definition 5, $n > 0$.

Part 2. Suppose that $n > 0$. By Definiton 5, we have $n = 0 + b$ where $b \neq 0$. Thus $n = b$. Since $b \neq 0$, we have $n \neq 0$. □

Corollary 13. *A natural number n is positive if and only if $n > 0$, and*

$$\mathbb{N}^+ = \{n \in \mathbb{N} \mid n \neq 0\} = \{n \in \mathbb{N} \mid n > 0\}.$$

The following, like many results in this chapter, is so ingrained into our thinking that it is easy to forget to prove it:

Theorem 14. *Let $n \in \mathbb{N}$. There is no $x \in \mathbb{N}$ such that $n < x < n + 1$. In other words, are no natural numbers between n and $n + 1$.*

Proof. (By contradiction). Suppose that $n < x$ and $x < n+1$. Since $n < x$, there is a positive b such that $x = n+b$. Since $x < n+1$, there is a positive c such that $n+1 = x+c$. Since $b \neq 0$, it has a predecessor. So $b = \sigma d = 1+d$ for some $d \in \mathbb{N}$. Putting this together gives

$$x = n + b = n + (1 + d) = (n + 1) + d = (x + c) + d = x + (c + d).$$

Thus, $0 + x = (c + d) + x$. By the cancellation law, $c + d = 0$. By a result in Chapter 1, we must have $c = d = 0$. But $c = 0$ is a contradiction. □

Exercise 8. Identify the results and definitions used in the above proof.

Exercise 9. Show that the only $x \in \mathbb{N}$ such that $x < 2$ are $x = 0$ and $x = 1$. Hint: use addition facts to show $x = 0$ and $x = 1$ work. Now suppose that $x \neq 0, 1$. Divide into cases: $x < 1$ and $1 < x$.

Theorem 15. *Suppose that $x, y, z \in \mathbb{N}$. If $x \leq y$ then $xz \leq yz$.*

Proof. Suppose that $x \leq y$. By Theorem 11 there is a $b \in \mathbb{N}$ where $y = x+b$. By the distributive law, $yz = (x + b)z = xz + bz$. Thus $xz \leq yz$ by Theorem 11. □

Exercise 10. Prove the following theorem using Definition 5 and Theorem 11.

Theorem 16. *Suppose that $x, y, z \in \mathbb{N}$.*
Then $x < y$ if and only if $x + z < y + z$.
Similarly, $x \leq y$ if and only if $x + z \leq y + z$.

Exercise 11. The remaining results of this section are given with sketchy proofs. Rewrite them in a more complete, organized form.

Theorem 17. *Suppose that $x, y, z \in \mathbb{N}$ where $z > 0$. If $x < y$ then $xz < yz$.*

Proof. Assume $x < y$. There is a $b \in \mathbb{N}$ such that $z = b + 1$. So

$$xz = x(b + 1) = xb + x \leq yb + x < yb + y = y(b + 1) = yz.$$

□

Theorem 18. *Suppose that $x, y, z \in \mathbb{N}$. If $xz < yz$ then $x < y$.*

Proof. Suppose $xz < yz$, but that $x < y$ fails. Then $y \leq x$ by the trichotomy law. Now use Theorem 15 to derive a contradiction. □

Exercise 12. Suppose $m_1, m_2, n_1, n_2 \in \mathbb{N}$, $m_1 < m_2$ and $n_1 < n_2$. Show that $m_1 + n_1 < m_2 + n_2$ and $m_1 n_1 < m_2 n_2$. Hint: for the last inequality, it helps to first show that $m_2 > 0$.

2.4 Cancellation law for multiplication

In Chapter 1 we proved the cancellation law for addition, but we postponed the multiplicative cancellation law until we developed properties of $<$. These properties allow for a quick and easy proof of the law.

Theorem 19 (Cancellation Law for Multiplication)**.** *Suppose $x, y, z \in \mathbb{N}$. If $xz = yz$ and $z \neq 0$ then $x = y$.*

Proof. By the trichotomy law, either $x = y$, $x < y$ or $y < x$. The last two cases lead to contradictions via Theorem 17 and the trichotomy law. Thus $x = y$. □

Another important theorem is the following:

Theorem 20. *Suppose $m, n \in \mathbb{N}$. If $mn = 0$ then $m = 0$ or $n = 0$.*

Proof. Assume $mn = 0$, but that m and n are both nonzero. Since $mn = 0$ and $0n = 0$, we have $mn = 0n$. Thus $m = 0$ by the cancellation law for multiplication. This is a contradiction. □

Exercise 13. Prove that $\mathbb{N}^+$ is closed under multiplication. Show this as a corollary to the above theorem.

Exercise 14. Suppose $n, B \in \mathbb{N}$ where $B \neq 0$. Show $B^n \neq 0$ for all $n \in \mathbb{N}$.

2.5 The set $\{1, \ldots, n\}$

The purpose of this section is to define and develop basic properties of the set $\{1, \ldots, n\}$ where $n \in \mathbb{N}$. This set will be important in the next chapter when we use it to define the cardinality of a finite set.

Definition 6. Let $n \in \mathbb{N}$ Then $\{1, \ldots, n\}$ is defined as the set

$$\{x \in \mathbb{N} \mid 1 \leq x \leq n\}.$$

More generally if $m, n \in \mathbb{N}$ the $\{m, \ldots, n\}$ is defined as $\{x \in \mathbb{N} \mid m \leq x \leq n\}$. Warning: this is the empty set if $m > n$.

We allow common notational variants. For example, $\{1, 2, \ldots, n\}$ is also defined as $\{x \in \mathbb{N} \mid 1 \leq x \leq n\}$.[1]

Here are facts from set theory, repeated for convenience:

$$\{a\} = \{x \mid x = a\}$$

$$\{a, b\} = \{x \mid (x = a) \vee (x = b)\}.$$

$$\{a, b, c\} = \{x \mid (x = a) \vee (x = b) \vee (x = c)\}.$$

Theorem 21. *Let $m \in \mathbb{N}$. The set $\{m, \ldots, m\}$ is equal to $\{m\}$.*

Proof. Suppose $x \in \{m, \ldots, m\}$. By Definition 6, $m \leq x \leq m$. However, $m \leq x$ and $x \leq m$ imply $x = m$. From set theory we know that $x = m$ implies $x \in \{m\}$. We conclude that $\{m, \ldots, m\} \subseteq \{m\}$.

Suppose $x \in \{m\}$. By basic set theory, this means $x = m$. Since $x = m$ we have both $m \leq x$ and $x \leq m$. In other words, $m \leq x \leq m$. So, by Definition 6, $x \in \{m, \ldots, m\}$. Thus $\{m\} \subseteq \{m, \ldots, m\}$. □

Theorem 22. *Let $n \in \mathbb{N}$. The set $\{n, \ldots, n+1\}$ is equal to $\{n, n+1\}$.*

Exercise 15. Prove the above theorem.

The next theorem one might want to show is that

$$\{n, \ldots, n+2\} = \{n, n+1, n+2\}.$$

After this, one might want a theorem concerning $\{n, \ldots, n+3\}$, and so on. The next theorem shows a key relationship that helps make it easier to prove such results.

[1] Warning: the notation $\{1, 2, \ldots, n\}$ might suggest to the reader that $n > 2$. To be safe, we should always mention any assumptions about the size of n.

Theorem 23. *If $m, n \in \mathbb{N}$ where $m \leq n$, then*

$$\{m, \ldots, n+1\} = \{m, \ldots, n\} \cup \{n+1\}.$$

Furthermore, $n+1 \notin \{m, \ldots, n\}$ so the union is disjoint.

Proof. First we prove $\{m, \ldots, n+1\} \subseteq \{m, \ldots, n\} \cup \{n+1\}$. So suppose that $x \in \{m, \ldots, n+1\}$. Then, by definition, $m \leq x$ and $x \leq n+1$. So either $x = n+1$ or $x < n+1$.

Case 1: $x = n+1$. So by basic set theory, $x \in \{m, \ldots, n\} \cup \{n+1\}$.

Case 2: $x < n+1$. Observe that $n < x$ implies $n < x < n+1$ which cannot happen (Theorem 14). Thus $x \leq n$. We know that $m \leq x$, so $m \leq x \leq n$. Thus $x \in \{m, \ldots, n\}$. Since $\{m, \ldots, n\} \subseteq \{m, \ldots, n\} \cup \{n+1\}$, we have $x \in \{m, \ldots, n\} \cup \{n+1\}$ as well.

So in either case, $x \in \{m, \ldots, n\} \cup \{n+1\}$. Thus

$$\{m, \ldots, n+1\} \subseteq \{m, \ldots, n\} \cup \{n+1\}.$$

Next we will prove that $\{m, \ldots, n\} \cup \{n+1\} \subseteq \{m, \ldots, n+1\}$. So suppose that $x \in \{m, \ldots, n\} \cup \{n+1\}$ By definition of union, we have two cases.

Case 1: $x \in \{m, \ldots, n\}$. In this case $m \leq x \leq n$. But $n < n+1$ by Theorem 5, Thus $x < n+1$ by (mixed) transitivity. In particular $x \leq n+1$. So $m \leq x \leq n+1$, thus $x \in \{m, \ldots, n+1\}$.

Case 2: $x \in \{n+1\}$ In other words, $x = n+1$. Now $n < n+1$, by Theorem 5. In particular, $n \leq x$. By hypothesis, $m \leq n$. Thus $m \leq x$ by transitivity. Trivially $x \leq n+1$, since $x = n+1$. So $m \leq x \leq n+1$. We conclude that $x \in \{m, \ldots, n+1\}$.

In either case, $x \in \{m, \ldots, n+1\}$. Thus

$$\{m, \ldots, n\} \cup \{n+1\} \subseteq \{m, \ldots, n+1\}.$$

Combining this with the previous inclusion, we conclude the sets are equal.

Finally, we show $n+1 \notin \{m, \ldots, n\}$. Suppose otherwise: $m \leq n+1 \leq n$. Then $n+1 \leq n$. However, by Theorem 5, $n < n+1$. This contradicts trichotomy. □

Exercise 16. Use the above to give another proof that the set $\{n, \ldots, n+1\}$ is just $\{n, n+1\}$. Hint: $\{a\} \cup \{b\} = \{a, b\}$ by basic set theory.

Exercise 17. Show $\{1, \ldots, 3\} = \{1, 2, 3\}$. Hint: $\{a\} \cup \{b\} \cup \{c\} = \{a, b, c\}$ by basic set theory.

The set $\{1, \ldots, n\}$ will be used extensively in the next chapter when we discuss counting, even if $n = 0$. Observe that, by definition, $\{1, \ldots 0\}$ is the empty set. However, the notation $\{1, \ldots 0\}$ is a bit awkward and may even suggest counting down from 1 to 0, which is not what we want. So for convenience we introduce the notation $\langle n \rangle$ as an alternative to $\{1, \ldots, n\}$.

Definition 7. If $n \in \mathbb{N}$ then

$$\langle n \rangle \stackrel{\text{def}}{=} \{x \in \mathbb{N} \mid 1 \leq x \leq n\}.$$

We can translate some of the results above to the new notation:

Proposition 24. *The set* $\langle 0 \rangle$ *is the empty set. The set* $\langle 1 \rangle$ *is* $\{1\}$*. The set* $\langle 2 \rangle$ *is* $\{1, 2\}$*. The set* $\langle 3 \rangle$ *is* $\{1, 2, 3\}$*.*

We can continue the above pattern with the following result:

Proposition 25. *Let* $n \in \mathbb{N}$*. Then the set* $\langle n + 1 \rangle$ *is* $\langle n \rangle \cup \{n + 1\}$*.*

Proof. If $n > 0$ then this just restates Theorem 23.. If $n = 0$ this result just states that $\{1\} = \varnothing \cup \{1\}$ which is a special case of a basic set theoretic result. □

Note. The above results are labelled as "propositions". From a logical point of view, there is no difference between terms such as "lemma", "proposition", "theorem", and "corollary". These terms all refer to mathematical statements that can be proved. Authors use different terms to emphasize the role of the results in the overall narrative, or the relationship between results. The above were called "propositions" to emphasize that they are not new results, but just restatements of results from before. This allows us to reserve the term "theorem" for new results.

2.6 The maximum principle and the well-ordering property

Above we established that $\mathbb{N}$ is linearly ordered by $<$. Several other number systems, including $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$, are also linearly ordered. We now consider a general definition of a linearly ordered set in order to describe the concepts of *minimum* and *maximum*:

Definition 8 (Linearly Ordered Set)**.** An *linearly ordered set* is a set U with a designated strict linear order relation $<$.

Remark 5. If we designate the order relation $<$ from Definition 5, then $\mathbb{N}$ becomes a linearly ordered set. In later chapters we will see how $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ can be considered to be linearly ordered sets. (There are linear orders on $\mathbb{C}$ but they are all somewhat artificial, so it is not useful for us to think of $\mathbb{C}$ as a linearly ordered set: there is no order relation that we would want to specially designate for $\mathbb{C}$.)

Remark 6. From $<$ we define $\leq, >, \geq$ as in Section 2.1. So, in a sense, a linearly ordered set has four designated order relations $(<, \leq, >, \geq)$.

Remark 7. If S is a subset of an ordered set U, then S is itself an ordered set. We just restrict the designated strict linear order to the subset S. Thus any subset of $\mathbb{N}$ is an ordered set.

Definition 9. Let S be a subset of an ordered set U. An element $b \in U$ is called a *lower bound* of S if $b \leq x$ holds for all $x \in S$. An element $B \in U$ is called an *upper bound* of S if $x \leq B$ holds for all $x \in S$.

Definition 10. Let S be a subset of an ordered set U. An element $m \in S$ is called a *minimum* of S if $m \leq x$ holds for all $x \in S$. An element $M \in S$ is called a *maximum* of S if $x \leq M$ holds for all $x \in S$.

Exercise 18. How do the above two definitions differ? Use the answer to this question to prove the following:

Lemma 26. *Suppose that S is a subset of an ordered set U. If b is a lower bound of S that is also an element of S then b is a minimum of S. Similarly, if B is an upper bound of S that is also an element of S then B is a maximum of S.*

Warning. Not all subsets of ordered sets have a minimum and a maximum. For example, if $U = \mathbb{N}$ then $S = U$ has no maximum. In a later chapter we will describe intervals such as $S = (0, 1]$ in $\mathbb{Q}$ that have no minimum. The number 0 is not a minimum of $(0, 1]$ since it is not in S, but 0 is a lower bound for S.

Existence may fail, but if existence holds then uniqueness must as well:

Theorem 27. *The minimum of U, if it exists, is unique. The maximum of U, if it exists, is unique.*

Exercise 19. Prove the above theorem.

Warning. In contrast, upper and lower bounds are not necessarily unique.

Exercise 20. What is the minimum of $\mathbb{N}$? Show that $\mathbb{N}$ has no maximum.

Although $\mathbb{N}$ itself has no maximum, we will now prove that every nonempty subset of $\mathbb{N}$ with an upper bound does have a maximum. This is in contrast to other number systems we will see such as $\mathbb{Q}$ and $\mathbb{R}$ where bounded subsets do not always have maxima.

Informal Exercise 21. Find a nonempty subset of $\mathbb{Q}$ that has an upper bound, but no maximum.

Theorem 28. *Suppose $n \in \mathbb{N}$. Every nonempty subset of $\{0, \ldots, n\}$ has a maximum.*

Proof. We will prove this by induction. Let S be the set of natural numbers n such that every nonempty subset T of $\{0, \ldots, n\}$ has a maximum element.

For the base case, assume T is a nonempty subset of $\{0, \ldots, 0\}$. By Theorem 21,

$$\{0, \ldots, 0\} = \{0\},$$

and so T is a nonempty subset of $\{0\}$. By set theory $T = \{0\}$. In this case 0 is the maximum element of T, since any $x \in T$ satisfies $x = 0$, and so $x \leq 0$.

Next, suppose that $n \in S$, so any nonempty subset of $\{0, \ldots, n\}$ has a maximum. Assume that T is a nonempty subset of $\{0, \ldots, n+1\}$. We need to prove that T has a maximum.

First assume that $n + 1 \in T$. In this case since $T \subseteq \{0, \ldots, n + 1\}$, every element x of T satisfies $0 \leq x \leq n + 1$. Hence by definition of maximum, $n + 1$ is the maximum of T.

Next assume that $n + 1 \notin T$. By Theorem 23,

$$\{0, \ldots, n + 1\} = \{0, \ldots, n\} \cup \{n + 1\},$$

so $T \subseteq \{0, \ldots, n\} \cup \{n + 1\}$. Since $n + 1 \notin T$, by set theory $T \subseteq \{0, \ldots, n\}$. By the induction hypothesis T has a maximum.

By the Induction Axiom $S = \mathbb{N}$ and the result follows. □

Theorem 29 (Maximum Principle)**.** *Suppose T is a nonempty subset of $\mathbb{N}$ with an upper bound. Then T has a maximum.*

Exercise 22. Use Theorem 28 to prove the Maximum Principle.

We now come to a key concept of this chapter:

Definition 11. An ordered set U is said to be *well-ordered* if every nonempty subset $S \subseteq U$ has a minimum.

Warning. Showing that $\mathbb{N}$ has a minimum is not enough to prove it is well-ordered. You must show that every nonempty subset of $\mathbb{N}$ has a minimum. Of course, different subsets can have different minima.

Theorem 30. *The set of natural numbers $\mathbb{N}$ is well-ordered.*

Proof. We will not prove this by induction, but instead will prove it as a corollary of the Maximum Principle. Let S be a nonempty subset of $\mathbb{N}$ that we want to show has a minimum. Let L be the set of lower bounds of S.

Observe that $0 \in L$ (why?), so L is nonempty. Since S is nonempty, there is an element $k \in S$. Observe that k is an upper bound of L. By the Maximum Principle, the set L has a maximum element n. In other words, there is a maximum lower bound of S. In particular, n is a lower bound of S, and we would like to show that n is a minimum of S as well. This requires showing that $n \in S$.

In order to show $n \in S$, suppose otherwise. Since n is a lower bound, this means $n < s$ for all $s \in S$. Since there are no elements strictly between n and $n+1$, this means that $n+1 \leq s$ for all $s \in S$. Observe that this means $n+1 \in L$. This contradicts the definition of n as the maximum of L. Thus $n \in S$, and so n is the minimum of S by Lemma 26. □

Exercise 23. Justify the observations in the above proof: (i) $0 \in L$, (ii) k is an upper bound of L, and (iii) if $n \notin S$ then $n+1 \in L$.

Informal Exercise 24. Is the set of nonnegative rational numbers well-ordered? Is the set of integers $\mathbb{Z}$ well-ordered?

2.7 Subtraction in $\mathbb{N}$

In Chapter 1 we considered addition and multiplication for the natural numbers, but we did not consider subtraction. This is because, $n - m$ is not defined for all $m, n \in \mathbb{N}$. However, now we have an order relation on $\mathbb{N}$, and when $n \geq m$ we *can* define subtraction. (In Chapter 4 we will introduce negative integers, and then we will be able to define $n - m$ for all integers n and m.)

Recall from Theorem 11 that $n \geq m$ if and only if there is a $b \in \mathbb{N}$ such that $n = m + b$. It turns out that this b is unique:

Lemma 31. *Let $n, m \in \mathbb{N}$. If $n \geq m$ then there is a unique $b \in \mathbb{N}$ such that $n = m + b$.*

Exercise 25. Prove the above lemma.

In mathematics, when we defined a new term such as $n - m$, we need to identify a specific object that the term will reference. Both existence and uniqueness are important in identifying a specific object with a given property. For example we can specify b with the property $n = m + b$ when $n \geq m$ since such a b exists, and there is no ambiguity since b is unique.

Definition 12 (Subtraction in $\mathbb{N}$)**.** Let $m, n \in \mathbb{N}$ be such that $n \geq m$. Then $n - m$ is defined to be the $b \in \mathbb{N}$ such that $n = m + b$. We call $n - m$ the *difference* of n and m, and call $-$ the *subtraction operation*.

The subtraction operation is not defined for all $(n, m) \in \mathbb{N} \times \mathbb{N}$, but only for the subset consisting of pairs where $n \geq m$. This means that subtraction is *not* a binary operation on $\mathbb{N}$. (In Chapter 3 we will define $n - m$ without assuming $n \geq m$, but the result will not always be in $\mathbb{N}$. We will see that subtraction *is* a binary operation on $\mathbb{Z}$. Later we will see that it is a binary operation on $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ as well.)

Directly from the definition we have the following.

Theorem 32 (Basic law of subtraction)**.** *Suppose $m, n, b \in \mathbb{N}$ and $n \geq m$. Then $n = m + b$ if and only if $b = n - m$.*

Theorem 33. *Let $n \in \mathbb{N}$. Then $n - n = 0$.*

Proof. Since $n = n + 0$, the conclusion $n - n = 0$ follows from Theorem 32. □

Exercise 26. Prove the following four theorems as consequences of Theorem 32.

Theorem 34. *Given $n, m \in \mathbb{N}$ with $n \geq m$, if $n - m = 0$ then $n = m$.*

Theorem 35. *Given $n \in \mathbb{N}$, then $n - 0 = n$.*

Theorem 36. *Suppose $m, n \in \mathbb{N}$ with $n \geq m$. Then $m + (n - m) = n$.*

Theorem 37. *Suppose $y \leq x$ and $z \leq x$ where $x, y, z \in \mathbb{N}$. Then $x - y = z$ if and only if $x - z = y$.*

Remark 8. Parentheses are often required to determine the meaning of an expression involving subtraction. For example, $(9 - 8) - 2$ is not defined since $1 < 2$. However, $9 - (8 - 2)$ is defined, and is equal to 3. (Note that $8 - 2 = 6$ since $8 = 2 + 6$, and $9 - 6 = 3$ since $9 = 6 + 3$.)

When parentheses are not explicitly written, we follow the usual rules for grouping. One rule we will adopt is that when we are given terms linked by $+$ and $-$, we perform our operations left to right. For example

$$a + b + c - d + e - f + (g + h) - (i - j) + k.$$

is really

$$\Big(\Big(\big(\big(\big((a + b) + c\big) - d\big) + e\big) - f\Big) + (g + h)\Big) - (i - j)\Big) + k.$$

Even though parentheses are usually necessary when using subtraction, the following shows a situation where parentheses can be moved.

Theorem 38. *Suppose $x, y, z \in \mathbb{N}$ are such that $z \leq y$. Then*

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

Proof. Let $c = x + (y - z)$. Observe that

$$c + z = \big(x + (y - z)\big) + z = x + \big((y - z) + z\big) = x + y$$

by the associative and commutative laws of addition (Chapter 1) and Theorem 36. Thus $z + c = x + y$ by the commutative law. By the basic law of subtraction $c = (x + y) - z$. The result follows. □

Here is an application of the above law:

Theorem 39. *Suppose $m, n, c \in \mathbb{N}$. If $n \geq m$ then $n + c \geq m + c$ and*

$$n - m = (n + c) - (m + c).$$

Proof. The first part follows from Theorem 16. For the second part:

$$\begin{aligned}
(m + c) + (n - m) &= ((m + c) + n) - m && \text{(Theorem 38)} \\
&= (n + (c + m)) - m && \text{(Comm law: Ch.1, twice)} \\
&= ((n + c) + m) - m && \text{(Assoc law: Ch.1)} \\
&= (n + c) + (m - m) && \text{(Theorem 38)} \\
&= (n + c) + 0 && \text{(Theorem 33)} \\
&= n + c && \text{(Rule from Ch.1)}
\end{aligned}$$

The result now follows from the basic law of subtraction. □

2.8 Simple recursion (optional)

It is common to define a function $g : \mathbb{N} \to S$ by recursive equations. These are equations that define $g(n)$ in terms of other values $g(m)$ of the same function g. This seems circular, but it is not since we will always require $m < n$.

For example, suppose we want to define a function $g : \mathbb{N} \to \mathbb{N}$ by the equations

$$g(0) = 1, \qquad \text{and} \qquad g(n + 1) = 2g(n) + 1.$$

These equations force $g(0) = 1$, $\quad g(1) = 2g(0) + 1 = 3$,

$$g(2) = 2g(1) + 1 = 2 \cdot 3 + 1 = 7, \quad \text{and so on.}$$

It seem clear that these equations define a unique function $g : \mathbb{N} \to \mathbb{N}$ given our intuitive idea of the natural numbers. How do we prove this? Uniqueness is not hard to show, but what about existence?

Exercise 27. Show uniqueness of g using induction.

The iteration theorem from Chapter 1 gives existence quite easily. First observe that the equation $g(n + 1) = 2g(n) + 1$ makes the next value a function of the previous value. The function f that gives the next value is given by the rule $x \mapsto 2x + 1$. You get the values of g by iterating f starting with 1. So define g by the equation $g(n) = f^n(1)$.

Exercise 28. Let $g : \mathbb{N} \to \mathbb{N}$ be defined by the rule $g(n) = f^n(1)$ where f is as above. Show that $g(0) = 1$ based on the fact that f^0 is the identity function. Show that $g(n+1) = 2g(n)+1$ based on the fact that $f^{n+1} = f \circ f^n$.

The above discussion generalizes to the following theorem.

Theorem 40 (Simple Recursion)**.** *Let S be a set. Suppose that $f : S \to S$ and $a \in S$ are given. Then there is a unique function $g : \mathbb{N} \to S$ satisfying the equations*

$$g(0) = a, \qquad \text{and} \qquad g(n+1) = f(g(n)).$$

Furthermore, g is given by $g(n) = f^n(a)$.

Exercise 29. Prove the above theorem.

2.9 More advanced recursion (optional)

A famous function defined by recursion is the *Fibonnaci function*. This is defined by the recursive equations:

$$F(0) = 0, \quad F(1) = 1, \quad F(n+2) = F(n) + F(n+1).$$

The difference between this and simple recursion is that, in general, a value of F depends not only on the previous value of F, but the previous *two* values of F. Note that the equations force

$$F(0) = 0, \quad F(1) = 1, \quad F(2) = 1 + 0 = 1, \quad F(3) = 1 + 1 = 2,$$

$$F(4) = 1 + 2 = 3, \quad F(5) = 2 + 3 = 5, \quad F(6) = 3 + 5 = 8,$$

and so on. It is obvious that these equations define a unique function $\mathbb{N} \to \mathbb{N}$ given our intuitive idea of the natural numbers. However, can we prove existence and uniqueness given what we have rigorously established?

To use iteration to prove the existence of the Fibonnaci function we use a trick: we switch to $\mathbb{N} \times \mathbb{N}$. We are interested in pairs (x, y) where x is a given Fibonnaci number, and y is the next Fibonnaci number. We also consider the function $\theta : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ defined by the rule $(x, y) \mapsto (y, x + y)$ which advances us from one pair of adjacent Fibonnaci numbers to the next pair of Fibonnaci numbers. For example, $\theta(3, 5) = (5, 8)$.

The initial pair is $(0, 1)$, so consider $\theta^n(0, 1)$ as the nth pair. Define $F(n)$ to be the first coordinate of $\theta^n(0, 1)$.

Exercise 30. Prove that $F(n+1)$ is the second coordinate of $\theta^n(0, 1)$. Hint: by definition of $F(n)$ we have $\theta^n(0, 1) = \big(F(n), y\big)$ for some $y \in \mathbb{N}$. Now apply θ to both sides of this equation to conclude that $y = F(n + 1)$.

From the above exercise, we have that $\theta^n(0, 1) = \big(F(n), F(n + 1)\big)$ for all $n \in \mathbb{N}$. The special case $n = 0$ gives us $\theta^0(0, 1) = \big(F(0), F(1)\big)$.

Exercise 31. Show $F(0) = 0$ and $F(1) = 1$.

Exercise 32. Show that $F(n + 2) = F(n) + F(n + 1)$ for all $n \in \mathbb{N}$. Hint: apply θ to both sides of the equation $\theta^n(0, 1) = \big(F(n), F(n+1)\big)$. Now look at the second coordinate of both sides.

We now have the existence of F. What about the uniqueness?

Exercise 33. Show that there is a unique solution $F : \mathbb{N} \to \mathbb{N}$ to the equations

$$F(0) = 0, \quad F(1) = 1, \quad F(n+2) = F(n) + F(n+1).$$

Hint: suppose F_1 and F_2 are two distinct solutions. Let S be the set of n such that $F_1(n) \neq F_2(n)$. So S has a least element m by the well-ordering theorem. Observe that $m \neq 0$ and $m \neq 1$. Conclude that $m \geq 2$. Now derive a contradiction.

We end with a different sort of recursion. The following equations defines the so-called *triangular numbers*:

$$T(0) = 0, \quad T(n+1) = (n+1) + T(n).$$

The difference between this and simple recursion is that $T(n+1)$ is not a function of $T(n)$ alone, but also depends on n. In other words, you need to know both $T(n)$ and n (or $n+1$) in order to find $T(n+1)$. Note that the equations force

$$T(0) = 0, \quad T(1) = 1 + 0 = 1, \quad T(2) = 2 + 1 = 3, \quad T(3) = 3 + 6 = 6,$$

and so on.[2] It is obvious, from our intuitive idea of the natural numbers, that these equations define a unique function $T : \mathbb{N} \to \mathbb{N}$. How do we prove this?

To use iteration to prove the existence of the function $T : \mathbb{N} \to \mathbb{N}$ we use a trick: we work in $\mathbb{N} \times \mathbb{N}$. We are interested in pairs (x, y) such as $(3, 6)$ where y is the xth triangular number. We also consider the function

$$\psi : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \times \mathbb{N}$$

defined by the rule $(x, y) \mapsto (x+1, (x+1)+y)$ which advances us from one triangular number to the next. For example, $\psi(2, 3) = (3, 6)$.

We then define $T(n)$ to be the second coordinate of $\psi^n(0, 0)$, and prove it satisfies the desired equations.

Exercise 34. Prove, by induction, that the first coordinate of $\psi^n(0, 0)$ is n. Thus

$$\psi^n(0, 0) = \big(n, T(n)\big).$$

Exercise 35. Use the equation $\psi^n(0, 0) = \big(n, T(n)\big)$ to show $T(0) = 0$.

Exercise 36. Show $T(n+1) = (n+1) + T(n)$. Hint: apply ψ to both sides of the equation $\psi^n(0, 0) = \big(n, T(n)\big)$.

[2] It turns out that $T(n) = n(n+1)/2$, so once we have developed division, we do not need to define T recursively. However, the recursive definition captures the idea of a triangle better than the formula $T(n) = n(n+1)/2$.

We see from these exercises the existence of a solution $T : \mathbb{N} \to \mathbb{N}$ to the equations

$$T(0) = 0, \quad T(n+1) = (n+1) + T(n).$$

Here is a generalization of the triangular number example:

Theorem 41. *Let S be a set, c an element of S, and $g : \mathbb{N} \times S \to S$ a function. Then there is a unique function $f : \mathbb{N} \to S$ satisfying the equations*

$$f(0) = c, \quad f(n+1) = g\big(n, f(n)\big).$$

for all $n \in \mathbb{N}$.

Proof. Let $\gamma : \mathbb{N} \times S \to \mathbb{N} \times S$ be defined by the rule $(n, x) \mapsto \big(n+1, g(n, x)\big)$. Define $f(n)$ to the the second coordinate of $\gamma^n(0, c)$. This function can be shown to satisfy the equations (see above discussion). Induction can be used to show uniqueness. □

Definition 13. The *factorial function* $f : \mathbb{N} \to \mathbb{N}$ is defined by using the above theorem with $S = \mathbb{N}$, $c = 1$, and $g(n, m) = (n+1) \cdot m$. In other words, it is the solution to the recursive equations

$$f(0) = 1, \quad f(n+1) = (n+1)f(n).$$

We write $n!$ for $f(n)$. So

$$0! = 1, \quad (n+1)! = (n+1)n!.$$

Chapter 3

Cardinality and Counting

This chapter is devoted to idea of *counting* and *cardinality* for finite sets. These are arguably the most important applications of the natural numbers $\mathbb{N}$. An important result is the *invariance of counting*: it does not matter what order you count a given finite set, the answer will be the same. We also consider cardinality properties of the following: subsets, bijections, injections, surjections, addition, multiplication, subtraction, and exponentiation. We use these ideas to give alternative, set-based proofs of some of the basic laws of arithmetic. These proofs are more insightful than the induction proofs from Chapter 1. Counting principles, such as the pigeonhole principle and the inclusion-exclusion principle, will be also be discussed.

We end with a short discussion of how iteration is related to addition and multiplication. This illustrates an important application for the operations of addition and multiplication, and prepares us for Chapter 4 where multiplication in $\mathbb{Z}$ is developed in terms of iteration.

By the end of the chapter we will have seen three characterizations each of addition and multiplication.
Addition is
(i) iterated successor (Chapter 1),
(ii) what is needed to count disjoint unions,
(iii) what is needed to describe the composition of two iterations.
Multiplication is
(i) iterated addition (Chapter 1),
(ii) what is needed to count cartesian products,
(iii) what is needed to describe the iteration of an iteration.

Remark 1. Our focus on counting and cardinality explains why we include 0 as an element of $\mathbb{N}$. Although classifying 0 as a natural number is quite

common, some adopt the convention that the integer 1 is the first natural number; this is a reflection of the historical fact that 0 was developed much later than the positive integers. However, the empty set is very common in modern mathematics, and we want to be able to count all finite sets including the empty set. We need 0 as a natural number in order to define the size (cardinality) of the empty set.

3.1 The invariance of counting

We count a finite set S by assigning a number to each object in S. We start by picking an object of S and assigning it 1. Then we assign 2 to another object of S (if there are any more). We continue until we have assigned a number, n say, to a final element of S. We then declare that S has n elements. This is a method we all learn as small children counting, say, apples or Halloween candy.

So in the process of counting a set S of n objects, every integer in $\{1, \ldots, n\}$ is assigned to an element of S. In other words, this process defines a function

$$\{1, \ldots, n\} \to S.$$

We assign distinct numbers to distinct objects, so the function is injective (one-to-one). We assign a number to every element of S, so the function is surjective (onto). Thus counting a finite set S is really the same thing as building a bijection $\{1, \ldots, n\} \to S$. Recall that we adopted the notation $\langle n \rangle$ for $\{1, \ldots, n\}$, where $\langle 0 \rangle$ is just the empty set.

This informal discussion motivates the following formal definition.

Definition 1. Let S be a set. If there is a bijection

$$\langle n \rangle \to S$$

with $n \in \mathbb{N}$, then we say that S *is counted by* n. The bijection $\langle n \rangle \to S$ is called a *counting function* or a *counting*.

We can count the objects of a set S in any order we like: we instinctively know that we will get the same result regardless of how we choose to assign the integers. In particular, if we have two countings $f : \{1, \ldots, m\} \to S$ and $g : \{1, \ldots, n\} \to S$ of the same set S, we expect that $m = n$. Why do we expect this to hold? Probably very few people are explicitly taught this principle. Is it based on experience or are we hard-wired to believe it? We will pass over such psychological questions. What is important to our axiomatic approach is that we *not* take it for granted. Instead it is a theorem:

Theorem 1 (Invariance of counting)**.** *Suppose that a set S can be counted by m and n. Then $m = n$.*

Definition 2. A set S is said to be *finite* if there is an $n \in \mathbb{N}$ such that S is counted by n.

Suppose that S is a finite set. Then the *cardinality* or *size* of S is defined to be the element $n \in \mathbb{N}$ such that S can be counted by n. This element is unique by the above theorem, so cardinality is well-defined. We write the cardinality of S as $|S|$ or $\#S$.

We will prove the invariance of counting theorem after establishing some lemmas. The main proof will be by induction. For the base case, we will need Proposition 2, involving functions to and from the empty set. Since these results concern basic set theory, we take them as given (but the proofs are not difficult).

Proposition 2. *Let S be a set. There is a unique function $\varnothing \to S$. It is injective. It is surjective if and only if S is also the empty set.*

On the other hand, there are no functions $S \to \varnothing$ unless S is also the empty set.

A key step in the proof of the main theorem, Lemma 6, will require the following intuitively obvious principle: given any $a \in S$ we can count a last. In other words, suppose that S can be counted by n and that $a \in S$, then we can find a counting function $f : \{1, \ldots, n\} \to S$ where $f(n) = a$. To prove this is possible (Lemma 5), we will first introduce the idea of a *transposition*:

Definition 3. Suppose $a, b \in S$. Consider the function $\tau_{(a\,b)} : S \to S$ defined by the rule $a \mapsto b$, $b \mapsto a$, and $x \mapsto x$ if x is not equal to a or b.

Observe that if $a = b$ then $\tau_{(a\,b)}$ is by definition the identity map. If $a \neq b$ then $\tau_{(a\,b)}$ is called a *transposition*.

In other words, the function $\tau_{a,b}$ just switches a and b: $\tau_{(a\,b)}(a) = b$ and $\tau_{(a\,b)}(b) = a$, but $\tau_{(a\,b)}(c) = c$ when $c \neq a, c \neq b$. These equations can be used to prove the following:

Lemma 3. *Suppose $a, b \in S$. Then $\tau^2_{(a\,b)}$ is the identity function on S. In other words, $\tau_{(a\,b)} : S \to S$ is its own inverse.*

Since $\tau_{(a\,b)} : S \to S$ has an inverse, we have the following:

Corollary 4. *Suppose $a, b \in S$. Then $\tau_{(a\,b)} : S \to S$ is a bijection.*

Lemma 5. *Suppose that S is a finite set with element $a \in S$, and suppose that S can be counted by $n > 0$. Then there is a bijection $f : \{1, \ldots, n\} \to S$ with the property that $f(n) = a$.*

Proof. By Definition 1, there is a bijection $g : \{1, \ldots, n\} \to S$. Form the function $f = \tau_{(a\,b)} \circ g$ where $b = g(n)$. Since $\tau_{(a\,b)}$ and g are bijections, the same is true of the composition f. Finally,

$$f(n) = \tau_{(a\,b)}\big(g(n)\big) = \tau_{(a\,b)}(b) = a.$$

□

The following is critical to the proof of the invariance of counting.

Lemma 6. *Suppose S is a set, and suppose that $S' = S \cup \{a\}$ where $a \notin S$. If S' can be counted by $n+1$, then S can be counted by n.*

Proof. By Lemma 5, we can assume a is counted last. In other words, we have a bijection $f : \langle n+1 \rangle \to S'$ with $f(n+1) = a$. Since f is injective, $f(x) \neq a$ if $x < n+1$. In particular, if $x \in \langle n \rangle$ then $f(x) \in S$.

Define $h\colon \langle n \rangle \to S$ to be the restriction of f to a smaller domain and codomain. In other words, $h(x) = f(x)$ for all $x \in \langle n \rangle$. The only real difference between f and h is the domain and codomain. Observe that h is injective: $h(x) = h(y)$ implies $f(x) = f(y)$, which in turn implies $x = y$ since f is an injection.

Now we will show that h is surjective. Let $y \in S$. If $y \in S$ then, since f is surjective, there is an $x \in \langle n+1 \rangle$ such that $f(x) = y$. Observe that $a \neq y$ since $a \notin S$. So $x \neq n+1$ because $f(n+1) = a \neq y$. Since $x \neq n+1$ and since

$$\langle n+1 \rangle = \langle n \rangle \cup \{n+1\}$$

by a result from Chapter 2, we have that $x \in \langle n \rangle$. So x is in the domain of h, and we have $h(x) = f(x) = y$. We conclude that h is surjective.

Since h is a bijection, S is counted by n. This establishes the lemma. □

Remark 2. The above proof is perfectly valid even for the case $n = 0$, but it is a bit unnatural and overly long for this simple case. In this case, when we restrict f we get a function $h\colon \langle 0 \rangle \to S$ which is automatically injective by Proposition 2. The proof of surjectivity given above can be interpreted as a proof that S is empty: assume $y \in S$ (i.e., assume that S is not empty), then there is an element $x \in \langle 1 \rangle$ mapping to y. But we can show $x \neq 1$. This is a contradiction since $\langle 1 \rangle = \{1\}$, so S is empty. Once we know S is empty, then h is a bijection by Proposition 2.

Exercise 1. The following lemma is important for the base case of the induction proof of the main theorem. Prove all three claims of this lemma using Proposition 2.

Lemma 7. *The empty set $\varnothing$ can be counted by 0. The empty set is the only set that can be counted by 0 (in other words, if a set S is counted by 0, it is empty). Furthermore, the empty set cannot be counted by $n > 0$.*

Now we prove the main theorem:

Proof. (Theorem 1). The proof is by induction. We set up the induction by defininining A be the set of all natural numbers u with the property that *any set T that can be counted by u can only be counted by u.* Our goal is to show that $A = \mathbb{N}$.

First we consider the base case with $u = 0$. Suppose T is a set that can be counted by 0. By Lemma 7, T must be the empty set, and can only be counted by 0. This implies $0 \in A$ (base case).

Now suppose $n \in A$. We want to show that $n + 1 \in A$ by showing that any set that can be counted by $n+1$, can only be counted by $n+1$. So let T be a set that can be counted by $n + 1$. Our goal is to show that if T can by counted by p then $p = n + 1$. Since $n + 1 > 0$, we have T is nonempty, and so $p > 0$ (Lemma 7). Since $p > 0$, we have $p = m + 1$ for some m. Let $a \in T$ (which exists since T is not empty). Let S be the set obtained by removing a from T. By Lemma 6, since T can be counted by $m + 1$, the set S can be counted by m. Similarly, since T can be counted by $n + 1$, the set S can be counted by n. Since $n \in A$, the set S can only be counted by n. So $n = m$. Thus $n + 1 = m + 1$.

We conclude that $n + 1 \in A$ if $n \in A$. By the induction axiom, $A = \mathbb{N}$.

Now we are ready to prove the main statement. Suppose that S is a set that can be counted by m and n. Since $\mathbb{N} = A$ we must have $n \in A$. By definition of A, the set S can only be counted by n. Thus $m = n$. □

3.2 Basic properties of counting

Theorem 8. *Let S be a finite set. The set S has cardinality 0 if and only if it is empty. The set S has positive cardinality if and only if it is nonempty.*

Exercise 2. Use Lemma 7 to prove the above.

Theorem 9. *Let $n \in \mathbb{N}$. The set $\langle n \rangle$ is finite and has cardinality n.*

Exercise 3. Give a very short proof of the above theorem.

Exercise 4. Show that if a set S has cardinality 1 then it has a unique element.

The following theorems state that two finite sets have the same size if and only if there is a bijection between them.

Theorem 10. *Suppose S is finite of cardinality n. If $f : S \to T$ is a bijection, then T is finite and has cardinality n.*

Theorem 11. *Suppose S and T are finite of cardinality n. Then there is a bijection $S \to T$.*

Exercise 5. Use bijections to prove the above two theorems. (Do not prove them by induction.)

Theorem 12. *Suppose S is finite of cardinality n. Suppose a is an element outside S. Then the set $S' = S \cup \{a\}$ is finite of cardinality $n + 1$.*

Proof. By Definition 1, there is a bijection $f\colon \langle n\rangle \to S$. By a result from Chapter 2, we have $\langle n+1\rangle = \langle n\rangle \cup \{n+1\}$. Observe that $n+1 \notin \langle n\rangle$. Our goal is to extend f to a function $f' : \langle n+1\rangle \to S'$.

Define $f' : \langle n+1\rangle \to S'$ as follows: if $x \in \langle n\rangle$ then $f'(x) \stackrel{\text{def}}{=} f(x)$, but let $f'(n+1) \stackrel{\text{def}}{=} a$.

First we show that f' is injective. To do so, suppose that $f'(x) = f'(y)$ where $x, y \in \langle n+1\rangle$. We wish to show that $x = y$. If $x = y = n+1$ then we are done. If one of the two, x say, is $n+1$, but the other is not then $f'(x) = a$ but $f'(y) = f(y) \in S$. Since $a \notin S$, we get a contradiction. The final case is where x and y are both not $n+1$. In this case, $f'(x) = f(x)$ and $f'(y) = f(y)$, so $f(x) = f(y)$. Thus $x = y$ since f is injective.

Finally, observe that f' is surjective. So $f' : \langle n+1\rangle \to S'$ is a bijection. By Definition 1, S' is counted by $n+1$. □

Exercise 6. Show that f' in the above proof is surjective.

Exercise 7. Prove the following corollaries.

Corollary 13. *If $S = \{a\}$ then S has cardinality 1.*

Corollary 14. *If $S = \{a, b\}$ where $a \neq b$, then S has cardinality 2.*

Theorem 15. *If A and B are finite, then so is $A \cup B$.*

Proof. We prove the result by induction on the size of B. So fix a finite set A, and define S as follows

$$S_A \stackrel{\text{def}}{=} \{x \in \mathbb{N} \mid A \cup B \text{ is finite for all sets } B \text{ of size } x\}.$$

Since A is an arbitrary finite set, the theorem will be established once we show that $S_A = \mathbb{N}$ regardless of A.

If B has size 0, it is empty. So $A \cup B = A$. Thus $A \cup B$ is finite since A is finite. Hence, $0 \in S_A$.

Suppose $n \in S_A$. We must show $n+1 \in S_A$. Let B be a set of size $n+1$. We must show that $A \cup B$ is finite. Since B has nonzero size, it is not empty. Let $b \in B$. Then $B - \{b\}$ has size n by Lemma 6. Since $n \in S_A$, we conclude that $A \cup (B - \{b\})$ is finite. If $b \in A$ then $A \cup B = A \cup (B - \{b\})$, so $A \cup B$ is finite. If $b \notin A$, then we use Theorem 12 and the equality

$$A \cup B = A \cup (B - \{b\}) \cup \{b\}$$

to conclude that $A \cup B$ is finite. We have established that $n+1 \in S_A$.

By the induction axiom, $S_A = \mathbb{N}$ regardless of choice of A. The result follows. □

We end this section with two lemmas needed in the next section. These require the concept of ordered pair. Recall from set theory that if A and B are sets, then $A \times B$ is defined to be the set of ordered pairs with first coordinate in A and second coordinate in B. Also recall from set theory that given (a, b) and (a', b') in $A \times B$, we have $(a, b) = (a', b')$ if and only if both $a = a'$ and $b = b'$.

Lemma 16. *Let $m, n \in \mathbb{N}$. There exists disjoint finite sets A and B such that A has cardinality m and B has cardinality n.*

Proof. If $m = 0$ let $A = \varnothing$. Otherwise let

$$A = \{(1, x) \in \mathbb{N} \times \mathbb{N} \mid 1 \leq x \leq m\}.$$

In this case, $x \mapsto (1, x)$ is a function $\{1, \ldots, m\} \to A$ with inverse function given by $(1, x) \mapsto x$. Thus the function is a bijection, so A has size m.

Similarly, if $n = 0$ let $B = \varnothing$. Otherwise let

$$B = \{(2, x) \in \mathbb{N} \times \mathbb{N} \mid 1 \leq x \leq n\}.$$

In this case, we can define a bijection showing B has size n.

Observe that the sets A and B are disjoint since their respective elements have different first coordinates. □

A similar proof gives the following.

Lemma 17. *Let $x, y, z \in \mathbb{N}$. There are pairwise disjoint finite sets A, B, C such that A has cardinality x, B has cardinality y, and C has cardinality z.*

Exercise 8. Define suitable A, B, C for the above lemma.

3.3 New perspective on addition

In Chapter 1, addition was defined in terms of iteration of successor. However, there are other ways to characterize addition. For example, one might explain to a child that $m + n$ is the number of apples you have if you combine m apples with n additional apples. In other words, if A is a set of m objects, and if B is a set of n objects, then, as long as A and B are disjoint, $m+n$ is the size of $A \cup B$. We now prove the validity of this alternative characterization of addition. It basically follows the pattern of Theorem 15 but uses a few basic laws of addition (proved in Chapter 1 before the associative and commutative laws).

Theorem 18. *Suppose that A and B are disjoint finite sets. Let m be the size of A, and let n be the size of B. Then $A \cup B$ has size $m + n$.*

Proof. We prove this by induction on the size of the second set. Let

$$S = \{u \in \mathbb{N} \mid A \cup X \text{ has size } m+u \text{ for all } X \text{ disjoint from } A \text{ of size } u\}$$

We start by showing $0 \in S$. If X has size 0, then X is the empty set, so $A \cup X = A$. Thus

$$|A \cup X| = |A| = m = m + 0.$$

We have established that $0 \in S$.

Suppose $k \in S$. We must show $k+1 \in S$. Suppose X has size $k+1$, and that X is disjoint from A. We must show that $A \cup X$ has size $m + (k+1)$.

Since $k + 1 > 0$, the set X is not empty. Let $x \in X$. Then $X - \{x\}$ has size k by Lemma 6. Since $k \in S$, we conclude that $A \cup (X - \{x\})$ has size $m + k$. Now since A and X are disjoint, x is not in $A \cup (X - \{x\})$. So $A \cup X = A \cup (X - \{x\}) \cup \{x\}$ has size $(m+k)+1$ by Theorem 12. By laws of Chapter 1 (before the associative and commutative laws)

$$(m+k)+1 = \sigma(m+k) = m + \sigma(k) = m + (k+1),$$

so $k+1 \in S_m$.

By the induction axiom, $S = \mathbb{N}$. Since $n \in \mathbb{N}$ we have $n \in S$. By definition of S, we have $|A \cup B| = m + n$. □

Remark 3. The above gives a second characterization of addition.

We now give new proofs of the commutative and associative laws.

Theorem 19 (Commutative Law)**.** *If $m, n \in \mathbb{N}$, then $m + n = n + m$.*

Proof. Let A and B be disjoint sets such that A has size m and B has size n (Lemma 16). Now $A \cup B$ has size $m + n$ by the above theorem, and $B \cup A$ has size $n + m$. Since $A \cup B = B \cup A$, we have $m + n = n + m$. □

Theorem 20 (Associative Law)**.** *If $x, y, z \in \mathbb{N}$, then $(x+y)+z = x+(y+z)$.*

Proof. (sketch). This follows from Lemma 17, and the identity

$$A \cup (B \cup C) = (A \cup B) \cup C.$$

□

Exercise 9. Write up the above proof. (You do not need to prove the identity $A \cup (B \cup C) = (A \cup B) \cup C$, since it is part of basic set theory.)

3.4 Subsets and functions in counting

In this section we investigate issues of cardinality for subsets and functions.

It is intuitively obvious that every subset of a finite set is also finite. This intuition is confirmed by the following theorem.

Theorem 21. *Every subset of a finite set is itself finite.*

Proof. This will be proved by induction on the size of the set. Let

$$S = \{x \in \mathbb{N} \mid \text{all sets } C \text{ of size } x \text{ have only finite subsets}\}.$$

Claim: $0 \in S$. To see this, observe that the only set C of size 0 is the empty set, and the only subset of the empty set is again the empty set. The empty set is finite. So $x = 0$ has the desired property.

Suppose $n \in S$. We must show $n + 1 \in S$. To do so, let C be a set of size $n + 1$. We claim that all subsets $B \subseteq C$ are finite. If $B = C$ we are done since by assumption C is finite. If B is a proper subset, let $a \in C$ be an element not in B. Then B is a subset of $C - \{a\}$. By Lemma 6, $C - \{a\}$ has size n. Since $n \in S$, it follows that subsets of $C - \{a\}$ are finite. Thus B is finite. We have established that $n + 1 \in S$.

By the induction axiom, $S = \mathbb{N}$. This establishes the theorem. □

Theorem 22. *Let C be a finite set of size c. If A is a subset of C of size a then $a \leq c$. If A is a proper subset then $a < c$.*

Proof. Consider the set B defined as follows:

$$B \stackrel{\text{def}}{=} C - A = \{x \in C \mid x \notin A\}.$$

Then B is a subset of C, so is finite by Theorem 21.

By basic set theory, A and B are disjoint and $A \cup B = C$. In particular, if b is the size of B then $c = a + b$ (Theorem 18). By a property of $\leq$ (Chapter 2) we get $a \leq c$.

Now if A is a proper subset of C, there is an element $w \in C$ that is not in A. So $w \in B$. Thus B is not empty, and the size b of B is not zero. Thus $a < c$ by definition of $<$ (Chapter 2). □

Theorem 23. *Let $f : A \to B$ be an injection where B is finite of size b. Then A is also finite and $a \leq b$ where a is the size of A. If, in addition, f is not surjective then $a < b$.*

Proof. Let $C = f[A]$ be the image of f. Since B is finite, the same is true of C (Theorem 21). Let $g : A \to C$ be the function obtained by restriction of codomain. In other words, $g(x)$ is defined to be $f(x)$ for all $x \in A$, and the functions f and g differ only in the choice of codomain. Observe that g is a bijection.

Since C is finite, A is finite and C and A have the same size (Theorem 10). Let a be the common size of A and C. Since C is a subset of B, $a \leq b$ (Theorem 22). The proof of the final statement is left to the reader. □

Exercise 10. Complete the above proof by proving the final statement: "If, in addition, f is not surjective then $a < b$".

Exercise 11. Prove the following two corollaries of Theorem 23.

Corollary 24 (Pigeonhole Principle)**.** *Let A be a finite set of size a and B a finite set of size b. If $f : A \to B$ is a function, and if $a > b$, then there are distinct elements of A mapping (via f) to the same element of B.*

Corollary 25. *Let $f : A \to B$ be an injection between two finite sets of the same size. Then f is a bijection.*

Remark 4. Informally, the Pigeonhole principles says that if there are more pigeons than pigeonholes, then there is a pigeonhole with more than one pigeon. Think of A as a set of pigeons, and B as a set of pigeonholes.

Theorem 26. *Let $g : A \to B$ be a surjection. If A is finite of size a then B is also finite, and the size b of B satisfies the inequality $a \geq b$. If, in addition, g is not injective then $a > b$.*

Proof. Let $h\colon \langle a \rangle \to A$ be a counting function (Definition 1). Then h is a bijection, so it is necessarily a surjection. Since g is a surjection, the composition $g \circ h : \langle a \rangle \to B$ is also a surjection.

Now $g \circ h$ might not be injective: there could be distinct integers $x, y \in \langle a \rangle$ with $g(h(x)) = g(h(y))$. We seek a subset $C \subseteq \langle a \rangle$ on which $g \circ h$ is injective. To form C, we will want to throw out either x or y whenever the equality $g(h(x)) = g(h(y))$ occurs. Let's agree to always throw out the larger integer. So officially C is defined to be the set of all $x \in \langle a \rangle$ with the property that

$$\forall\, z \in \langle a \rangle\,, \quad g(h(x)) = g(h(z)) \;\Rightarrow\; x \leq z.$$

Let f be the restriction of $g \circ h$ to C. We claim that $f : C \to B$ is injective. To see this, suppose that $f(x) = f(y)$ where $x, y \in C$. Since f is a restriction of $g \circ h$, we have $g(h(x)) = g(h(y))$. By the definition of C, this implies $x \leq y$ and $y \leq x$. So $x = y$. Thus f is injective.

We claim that $f : C \to B$ is surjective. Let $b \in B$. Since $g \circ h$ is surjective, there is an integer y with $g(h(y)) = b$. Let x be the smallest such integer (existence by well-ordered property). Then $x \in C$, and $f(x) = b$. Thus f is surjective.

So f is a bijection. Since $C \subseteq \langle a \rangle$, and since $\langle a \rangle$ has size a, we have that C is finite (Theorem 21) and $|C| \leq a$ (Theorem 22). Since f is a bijection, B is finite and $|B| = |C|$ (Theorem 10). So $|B| \leq a$ as desired.

To establish the second statement, observe that if g is not injective, then C is a proper subset of $\langle a \rangle$. So $|C| < a$ (Theorem 22), giving us $|B| < a$ as desired since $|B| = |C|$. □

Corollary 27. *Let $g : A \to B$ be a surjection between two finite sets of the same size. Then g is a bijection.*

Theorem 28. *If A is a finite set of size n, and if $m \leq n$, then there is a subset $B \subseteq A$ of size m.*

Proof. Let $f : \langle n \rangle \to A$ be a counting function. Let B be the image of $\langle m \rangle$ under f. Observe that f restricts to a bijection $\langle m \rangle \to B$, so B has cardinality m. □

In Chapter 2 we showed that $\mathbb{N}$ is well-ordered. In other words, every nonempty subset $S \subseteq \mathbb{N}$ is has a minimum. This leads to a question: which subsets have a maximum?

Theorem 29. *Let S be a nonempty subset of $\mathbb{N}$. Then S has a maximum if and only if S is finite.*

Proof. If S has a maximum n, then S is a subset of $T = \{0, \ldots, n\}$. Observe that $\{0, \ldots, n\} = \{0\} \cup \{1, \ldots, n\}$. So T is finite since it is the union of two finite sets. Since S is the subset of a finite set, it is finite.

Now we prove the converse: if $S \subseteq \mathbb{N}$ is finite and nonempty then it has a maximum. This is proved by induction on the size of S. Let A be the set consisting of all $x \in \mathbb{N}$ with the following property: *all nonempty subsets of $\mathbb{N}$ of size x have a maximum.*

Observe $0 \in A$ since there are no nonempty sets of size 0.

Suppose $n \in A$. We must show $n + 1 \in A$. In other words, if $S \subseteq \mathbb{N}$ has size $n + 1$ we must find a maximum M. Start with any $s \in S$. If s is a maximum then we are done: $M = s$. Otherwise, S contains an element larger than s, so $S - \{s\}$ is nonempty. In fact $S - \{s\}$ is a nonempty set of size n. Since $n \in A$, this implies that $S - \{s\}$ has a maximum M. Observer that M is a maximum of S.

We conclude that if $n \in A$ then $n + 1 \in A$. So, $A = \mathbb{N}$ (induction) as desired. □

Exercise 12. Let $S \subseteq \mathbb{N}$ be a nonempty subset. Suppose that S has an upper bound $B \in \mathbb{N}$. Show that S is finite.

3.5 New perspective on multiplication

In Chapter 1 multiplication was defined in terms of iteration of addition. However, there are other ways to characterize multiplication. For example, one can count the number of ordered pairs: if there are m choices for the

first coordinate of an ordered pair, and if there are n choices for the second coordinate of an ordered pair, then $m \cdot n$ gives the total number of ordered pairs. In other words, if A is finite of size m and B is finite of size n then the product $m \cdot n$ is the size of $A \times B$. We now prove that this alternative characterization of multiplication is valid.

Theorem 30. *Suppose that A and B are finite sets. Let m be the size of A, and let n be the size of B. Then $A \times B$ is finite with size $m \cdot n$.*

Proof. (Induction) Let

$$S = \{u \in \mathbb{N} \mid \text{the size of } A \times X \text{ is } m \cdot u \text{ for all } X \text{ of size } u\}.$$

First we show $0 \in S$. Let X have size 0, so X is empty. Observe that $A \times X$ is also empty since no ordered pair has second coordinate in the empty set. Thus $|A \times X| = 0 = m \cdot 0$. We conclude that $0 \in S$.

Suppose $k \in S$. We must show $k + 1 \in S$. In other words, for any X of size $k + 1$ we must show $|A \times X| = m(k + 1)$.

Since $|X| = k + 1$, the set X is not empty. Let $x \in X$. Then $X - \{x\}$ has size k by Lemma 6. Since $k \in S$, we conclude that $A \times (X - \{x\})$ has size $m \cdot k$. Observe that

$$A \times X \quad = \quad A \times (X - \{x\}) \quad \cup \quad A \times \{x\}$$

and that the union is disjoint. Also, there is a bijection $A \to A \times \{x\}$, so A and $A \times \{x\}$ have the same size. So, by Theorem 18, $A \times X$ has size $m \cdot k + m$. From Chapter 1, $m \cdot k + m = m \cdot \sigma k = m(k + 1)$. So $k + 1 \in S$.

By the induction axiom, $S = \mathbb{N}$. Since $n \in \mathbb{N}$ we have $n \in S$. By definition of S we have that $|A \times B| = mn$. □

Remark 5. This result is related to a fundamental counting principle: if you have m choices for one property, and n choices for a second property, then there are mn total combinations given by the two choices. For example, if your computer has five fonts and each font comes in plain, bold, and italic style, then there are 15 total combinations of font and style. To see the connection between this principle and the above theorem, think of the two choices as giving the coordinates of an ordered pair. So the number of combinations is the number of ordered pairs.

Remark 6. We can write the above theorem as

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

This explains why the symbol $\times$ is popular for cartesian product. Old set theory books sometimes use $+$ for union due to the connection between union and addition, but this notation lost out to $\cup$. If the $+$ notation had survived, we would have, for disjoint unions,

$$|A + B| = |A| + |B|.$$

Remark 7. The above theorem gives a new characterization of multiplication. Observe that the above theorem and proof are the first place where we have used multiplication in this chapter. For instance, we have used facts about addition and inequalities from Chapters 1 and 2, but the facts we used were not those dependent on multiplication. So everything so far has been "multiplication free".

The above proof uses just two properties about multiplication from Chapter 1: (i) $0 = m \cdot 0$ and (ii) $m \cdot n + m = m \cdot \sigma n$. These two facts occur in Chapter 1 before the commutative, associative, and distributive laws of multiplication.

We will now give new proofs of the commutative, associative and distributive laws of multiplication, which are independent of the old proofs. They give more insight into why these laws are true than the induction proof of Chapter 1. The commutative law is based on the following easy exercise.

Exercise 13. Let A and B be sets. Describe a very simple function between $A \times B$ and $B \times A$ involving switching coordinates that works no matter what A and B are. Prove that it is a bijection. This natural bijection is called a *canonical bijection*.

Theorem 31 (Commutative law)**.** *If $m, n \in \mathbb{N}$, then $m \cdot n = n \cdot m$.*

Proof. Let $A = \langle m \rangle$ and $B = \langle n \rangle$. By Theorem 30, the set $A \times B$ has size mn and $B \times A$ has size nm. By the previous exercise, there is a bijection

$$A \times B \to B \times A.$$

Thus $m \cdot n = n \cdot m$ by Theorem 10. □

Theorem 32 (Associative law)**.** *If $x, y, z \in \mathbb{N}$, then $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.*

Proof. This is similar to the previous proof. Observe that there is a bijection

$$(A \times B) \times C \to A \times (B \times C)$$

defined by the rule $\big((a, b), c\big) \mapsto \big(a, (b, c)\big)$. □

Theorem 33 (Distributive law)**.** *If $x, y, z \in \mathbb{N}$, then $(x + y)z = xz + yz$.*

Proof. Let A, B, C be finite sets of size x, y, z respectively, and choose A and B to be disjoint (Lemma 17). Then

$$(A \cup B) \times C = (A \times C) \cup (B \times C).$$

The result follows from Theorem 18 and Theorem 30. □

Exercise 14. In the above proof we used the fact that

$$(A \cup B) \times C = (A \times C) \cup (B \times C)$$

and the fact that $A \times C$ and $B \times C$ are disjoint. Show these two facts, and show how the statement of the theorem follows from Theorem 18 and Theorem 30.

Remark 8. There is another characterization of multiplication that is commonly used. Suppose A is a set of disjoint finite sets, where each member of A is a set of size n. Suppose A is finite of cardinality m. Then the union of all the sets in A has mn elements.

For example, if you have five apples, and each apple has three worms, then there are 15 worms total (here each member of A is the set of worms on one particular apple). This principle is closely related to the multiplication principle for cartesian products proved above. In fact, there is a bijection between the union of the sets in A and $A \times \langle n \rangle$.

3.6 New perspective on subtraction

Above we described set-theoretic characterizations of addition and multiplication. There is also a simple set-theoretic characterization of subtraction. For example, informally one might describe $5-2$ to be the number of apples you have when you start with a set of 5 apples, and remove a subset of 2 apples. The following theorem implements this idea.

Theorem 34. *Let A be a finite set of size n, and let B be a subset of size m. Then the set $A - B = \{a \in A \mid a \notin B\}$ has size $n - m$.*

Proof. (Sketch) Observe that $A = B \cup (A - B)$. □

Exercise 15. Prove the above theorem. Be sure to mention that the subsets on the right-hand side are disjoint. Also refer to Theorem 18 and theorems on subtraction in Chapter 2.

From Section 3.3 we know that addition gives the size of $A \cup B$ if A and B are finite disjoint sets. What if they are not disjoint? The answer is given by the inclusion-exclusion principle:

Theorem 35 (Inclusion-exclusion principle)**.** *Let A and B be finite sets that are not necessarily disjoint. Then $A \cup B$ is finite, and*

$$|A \cup B| = \big(|A| + |B|\big) - |A \cap B|.$$

In other words

$$|A \cup B| + |A \cap B| = |A| + |B|.$$

Exercise 16. Prove the above theorem. Hint: show

$$A \cup B = A \cup \big(B - (A \cap B)\big).$$

Also use the following from Chapter 2: given $x, y, z \in \mathbb{N}$ with $z \leq y$,

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

Remark 9. The above idea can be extended to three or more sets.[1] For example, if A, B, C are finite sets, then $A \cup B \cup C$ is finite and $|A \cup B \cup C|$ is given by

$$|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

3.7 New perspective on exponentiation

In Chapter 1 exponentiation was defined in terms of iteration of multiplication. However, there are other ways to characterize exponentiation. For instance, we will see that n^m is the number of functions $A \to B$ where A is finite of size m and where B is finite of size n. Because of this, the set of functions $A \to B$ is sometimes written B^A:

Definition 4. Let A and B be sets. Then define B^A to be the set of functions $A \to B$.

Theorem 36. *Suppose that A and B are finite sets. Let m be the size of A, and let n be the size of B. Then B^A is finite with size n^m.*

Proof. (Induction). Let

$$S = \{u \in \mathbb{N} \mid \ B^X \text{ has size } n^u \text{ for every } X \text{ of size } u\}$$

First we show $0 \in S$. So let X have size 0. In other words $X = \varnothing$. By Proposition 2 there is a unique function $\varnothing \to B$, so the size of $B^\varnothing$ is exactly 1. But $1 = n^0$, so B^X has size n^0. Hence, $0 \in S$.

Suppose $k \in S$. We must show $k + 1 \in S$. In other words, we must show that if $|X| = k + 1$, then $\left|B^X\right| = n^{k+1}$. Since X has size $k + 1$, it is not empty. Let $x \in X$. Then $X - \{x\}$ has size k by Lemma 6. Since $k \in S$ (inductive hypothesis), we conclude that $B^{X-\{x\}}$ has size n^k.

What is the relationship between functions from X to B and functions from $X - \{x\}$ to B? Observe that when you restrict a function $f : X \to B$ to $X - \{x\}$, you get a function $g : (X - \{x\}) \to B$. Note that g defines f at

[1] I do not recommend proving this now. Life is much easier when we have the identity $x - y = x + (-y)$, which is not developed until Chapter 4.

every element in X except x. So to describe f in terms of g you also need to know $f(x)$. Consider the function

$$\Phi\colon B^X \to B^{X-\{x\}} \times B$$

that takes a function $f : X \to B$ to $\big(g, f(x)\big)$ where g is the restriction of f described above. Now the function Φ has an inverse: given a function $g\colon (X - \{x\}) \to B$ and a value $b \in B$ there is a unique $f\colon X \to B$ that agrees with g on the set $X - \{x\}$ and at the same time has value $f(x) = b$. Since Φ has an inverse, it is a bijection.

Since $\Phi\colon B^X \to B^{X-\{x\}} \times B$ is a bijection, we can find the size B^X from the size of $B^{X-\{x\}} \times B$. But $B^{X-\{x\}} \times B$ has size $n^k \cdot n$ by Theorem 30 and the inductive hypothesis. By Theorem 10, B^X must also have size $n^k \cdot n$. From Chapter 1 we know $n^k \cdot n = n^{k+1}$. We have established that $k+1 \in S$.

By the induction axiom, $S = \mathbb{N}$. Since $m \in \mathbb{N}$ it must be in S. By the definition of S, the set B^A has size n^m. □

Remark 10. We can write the conclusion of the above theorem as

$$\big|B^A\big| = |B|^{|A|}.$$

Remark 11. Under this interpretation, $n^0 = 1$ reflects the fact from Proposition 2 that there is a unique functions $\varnothing \to B$ from the empty set to any given set B of size n. This works even if $n = 0$ where B is empty. So the equation $0^0 = 1$ makes sense. This justifies our decision to define 0^0 to be 1 in Chapter 1.

Remark 12. The only exponentiation identities from Chapter 1 used in the above proof are that $n^0 = 1$ and $n^k n = n^{k+1}$. The following give new, independent, proofs of other identities from Chapter 1.

Theorem 37. *If $x, y, n \in \mathbb{N}$ then*

$$(xy)^n = x^n y^n.$$

Proof. (sketch). Let A be a finite set of size x, let B be a finite set of size y, and let C be a finite set of size n. Choosing a function $f : C \to A \times B$ is the same as choosing two functions (f_1, f_2) with $f_1 : C \to A$ and $f_2 : C \to B$. In other words, there is bijection

$$(A \times B)^C \to A^C \times B^C.$$

The result follows from Theorem 30 and Theorem 36. □

Theorem 38. *If $x, m, n \in \mathbb{N}$ then*

$$x^{m+n} = x^m x^n.$$

Proof. (sketch). Let A be a finite set of size x, let B be a finite set of size m, and let C be a finite set of size n. Choose B and C to be disjoint. Choosing a function $f : B \cup C \to A$ is the same as choosing two functions (f_1, f_2) with $f_1 : B \to A$ and $f_2 : C \to A$. In other words, there is bijection

$$A^{B \cup C} \to A^B \times A^C.$$

The result follows from Theorem 18, Theorem 30, and Theorem 36. □

Theorem 39. *If $n \in \mathbb{N}$ is not 0 then*

$$0^n = 0.$$

Proof. (sketch). Let B be the empty set, and let A be a set of size $n > 0$. There are no functions from A into the empty set (Proposition 2). So B^A is empty. The result follows from Theorem 36. □

Theorem 40. *If $n \in \mathbb{N}$ then*

$$1^n = 1.$$

Proof. (sketch). Let $B = \{1\}$, and let A be a finite set of size n. Every function $f : A \to B$ is given by the formula $f(x) = 1$. Thus there is one function in B^A. The result follows from Theorem 36. □

Theorem 41. *If $x, n, m \in \mathbb{N}$ then*

$$\big(x^m\big)^n = x^{mn}.$$

Proof. (sketch). Let A be a finite set of size x, let B be a finite set of size m, and let C be a finite set of size n. Claim: there is a bijection

$$\varphi : (A^B)^C \to A^{B \times C}.$$

To see this, suppose $f : C \to A^B$ is given. Then define $\varphi(f) : B \times C \to A$ by the rule $(b, c) \mapsto \big(f(c)\big)(b)$. This rule makes sense since $f(c)$ is itself a function $B \to A$. It is an exercise to show that φ has an inverse. Thus φ is a bijection.

The result follows from Theorem 30 and Theorem 36. □

3.8 Laws of iteration — additional perspectives on addition and multiplication

We have two fundamentally different ways of viewing addition: (i) iterated successor (from Chapter 1), and (ii) the size of disjoint unions. In this section we give a third way of looking at addition: (iii) the order of iteration obtained by composing two iterations. We give a similar result for multiplication.

Theorem 42. *Let $f : S \to S$ be a function whose domain equals its codomain. If $m, n \in \mathbb{N}$ then*

$$f^m \circ f^n = f^{m+n}.$$

Proof. (Induction on n). Fix $m \in \mathbb{N}$. Let $A_m = \{x \in \mathbb{N} \mid f^{m+x} = f^m \circ f^x\}$. Observe that $0 \in A_m$ since f^0 is the identity (Chapter 1).

Now assume $n \in A_m$. We will show that $n + 1 \in A_m$.

$$\begin{aligned} f^m \circ f^{n+1} &= f^m \circ (f^n \circ f) && \text{(Lemma 43 below)} \\ &= (f^m \circ f^n) \circ f && \text{(Assoc. of } \circ) \\ &= f^{m+n} \circ f && (n \in A_m) \\ &= f^{m+n+1} && \text{(Lemma 43 below)} \end{aligned}$$

So $n + 1 \in A_m$.

By the induction axiom $A = \mathbb{N}$. The result follows. □

The above used the following lemma. Recall that $f^{n+1} = f \circ f^n$ by the iteration axiom of Chapter 1 (actually a theorem: see the optional sections).

Lemma 43. *Let $f : S \to S$ be a function whose domain equals its codomain. If $n \in \mathbb{N}$ then*

$$f^{n+1} = f^n \circ f.$$

Proof. Let $A = \{x \in \mathbb{N} \mid f^{x+1} = f^x \circ f\}$. Observe that $0 \in A$ since f^0 is the identity and $f^1 = f$ (see Chapter 1).

Now assume $n \in A$. We must show that $n + 1 \in A$.

$$\begin{aligned} f^{(n+1)+1} &= f \circ f^{n+1} && \text{(Iteration Axiom/Theorem)} \\ &= f \circ (f^n \circ f) && (n \in A) \\ &= (f \circ f^n) \circ f && \text{(Assoc. of } \circ) \\ &= f^{n+1} \circ f. && \text{(Iteration Axiom/Theorem)} \end{aligned}$$

So $n + 1 \in A$.

By the induction axiom $A = \mathbb{N}$. The result follows. □

Theorem 44. *Let $f : S \to S$ be a function whose domain equals its codomain. If $m, n \in \mathbb{N}$ then*

$$(f^m)^n = f^{mn}.$$

Exercise 17. Prove the above theorem.

Observe that we now have three fundamental ways to think of multiplication. (i) iterated addition, (ii) size of finite cartesian products, (iii) the index of iteration of an iteration of an iteration.

3.9 Infinite sets

Most of this chapter has been concerned with finite sets. In this chapter we consider briefly infinite sets.

Definition 5. A set is *infinite* if it is not finite.

Theorem 45. *Suppose that B is infinite and that A is a finite subset of B. Then $B - A$ is infinite.*

Exercise 18. Use Theorem 15 to prove the above theorem.

Exercise 19. Prove the following by induction:

Theorem 46. *If A is infinite, then A has subsets of every finite cardinality. In other words, given n there is a subset of A of cardinality n.*

There is a converse:

Theorem 47. *If A has subsets of every finite cardinality, then A is infinite.*

Proof. Suppose not. Then $|A| = a$ for some $a \in \mathbb{N}$. By assumption, A has a subset of size $a + 1$. This contradicts Theorem 22. □

Theorem 48. *If there is an injection $\mathbb{N} \to A$ then A is infinite.*

Proof. (sketch) Such an injection can be used to produce subsets of every finite cardinality. □

There is another axiom of mathematics, that we have not needed, called the *axiom of choice*. If we assume such an axiom, we can prove the following converse to the above theorem.

Theorem 49. *If A is infinite, then there is an injection $\mathbb{N} \to A$.*

The basic idea of the proof is to use the axiom of choice to give a function h that chooses an element $h(B)$ in $A - B$ for each finite subset B of A. Next recursively define $B_0 = \varnothing$ and $B_{n+1} = B_n \cup \{h(B_n)\}$. Now consider the function $n \mapsto h(B_n)$ and show it is injective.

Remark 13. If there is a bijection between $\mathbb{N}$ and a set A then we say that A is a *countably infinite set*. For example, $\mathbb{N}$ itself is countable (use the identity map). This makes sense since given an infinite amount of time (and infinite patience) you *can* count every element of $\mathbb{N}$. For a general set A, if there is a bijection $f : \mathbb{N} \to A$, you could use the bijection to do a similar counting of A.

The above theorem shows that every infinite set has a countable subset: just take the image of the injection given by the theorem. We will see later that there are infinite sets that are so big that they are not countable. In fact, the real numbers will turn out to be uncountable. Surprisingly, the rational numbers turn out to be countable.

Chapter 4

The Integers $\mathbb{Z}$

4.1 Introduction

The natural numbers are designed for measuring the size of finite sets, but what if you want to compare the sizes of two sets? For example, you might want to compare the number of chairs in a classroom with the number of students to determine the number of free chairs. If there are more students than chairs, you would use *negative integers* to indicate the absence of free chairs.

Again, natural numbers are good for indicating the number of times you want to iterate a function $f\colon S \to S$. But what if you want to allow iterations of the inverse function: these are indicated by writing f^a where a a negative integer.

What if you want to subtract $m, n \in \mathbb{N}$ and you want $m-n$ to make sense even if $m < n$? Then you also need negative integers.

Finally, you might need negative numbers when you solve even the most basic types of algebraic equations. For instance, if you allow negative numbers you can always solve $x + b = c$ for x regardless of the sizes of b and c.

For these reasons and others the negative integers were introduced into mathematics. In this chapter we will construct and study the set $\mathbb{Z}$ of all integers, including negative integers. The amazing thing is that most of our algebraic laws that we developed for the natural numbers $\mathbb{N}$ continue to hold, and many new ones besides. Three modern ways of saying this is to say that $\mathbb{Z}$ is an *abelian group* under addition, $\mathbb{Z}$ is a *commutative ring*, and $\mathbb{Z}$ is an *integral domain*.

Our method of constructing $\mathbb{Z}$ will be a bit more involved than you might at first expect. You might expect a construction where you make a copy $\mathbb{N}^-$

of the positive integers $\mathbb{N}^+$ that you distinguish from the original set $\mathbb{N}^+$ somehow. For example, you could indicate the negative copy of 7 by writing -7. Then you take the (disjoint) union of $\mathbb{N}^-$ and $\mathbb{N}$ and call that $\mathbb{Z}$. Call this approach the *naive approach*. This is *not* the approach we will take.

One problem with the naive approach is that it requires a highly non-unified approach. For example, to define $x + y$ you need to consider six cases. Case i: if $x \geq 0$ and $y \geq 0$ then $x + y$ is defined as in Chapter 1. Case ii: if $x \geq 0$ and $y < 0$ with $y = -n$, and if $x \geq n$ then $x + y$ is defined as $x - n$. Case iii: if $x \geq 0$ and $y < 0$ with $y = -n$, and $x \leq n$ then $x + y$ is defined as $-(n - x)$. And so on. If you want to prove a law, such as the associative law $x + (y + z) = (x + y) + z$, you must allow for a very large number of cases. Our approach, adopted in this chapter, will require many fewer separate cases by treating $\mathbb{Z}$ in a more unified manner.

Our approach is closely tied to the idea that elements of $\mathbb{Z}$ are used to measure the *net difference* between two finite sets. Just as elements of $\mathbb{N}$ have as one of their main applications the ability to describe the size of a finite set A, the integers $\mathbb{Z}$ can be used to describe the net difference of the size of A over the size of B, even if B is bigger than A. For example, A can represent how many dollars you have, and B how many dollars you owe. Or A can represent the number of protons in a charged particle, and B the number of electrons. There can be a temporal aspect: A can be the number of sheep that a farmer has this year, and B the number he or she had last year.

If we write (m, n) for the sizes of two sets, A and B respectively, we want to describe the net difference between m and n. By removing one from each side until we reach 0, the pair (m, n) can be reduced to either the form $(m', 0)$ or the form $(0, n')$ depending on whether A is larger or smaller than B. You can think of $(m', 0)$ as representing a positive net difference if $m' \neq 0$, and you can think of $(0, n')$ as representing a negative net difference if $n' > 0$. For example, the pair $(19, 15)$ corresponds with $(4, 0)$, which describes a positive net difference, and $(8, 10)$ corresponds with $(0, 2)$, which describes a negative net difference.[1]

Roughly speaking, -2 can be thought of as the pair $(0, 2)$, and positive 4 as the pair $(4, 0)$. This is not quite what we will do in this chapter: we will define -2 as a certain type of *equivalence class* containing $(0, 2)$, and positive 4 as a certain type of equivalence class containing $(4, 0)$. Each equivalence class will be composed of pairs with the same net difference.

There are two reasons for using equivalence classes. First, the proofs of many of the theorems are easier using such equivalence classes. Second, the

[1] Of course this is for the difference of A over B. If our focus was on the difference of B over A, then $(m', 0)$ would be regarded as negative and $(0, n')$ as positive. In this chapter we will consistently measure the difference of the first set over the second.

idea of equivalence class is used in many branches of modern mathematics to construct new objects, and it is good for you to get used to the idea in a relatively simple situation. We will use the equivalence class approach later when we define the integers modulo n, the rational numbers $\mathbb{Q}$, and the real numbers $\mathbb{R}$. In group theory, equivalence classes are used to construct quotient groups, and so on.

There is a pleasant symmetry between positive and negative when we study only addition.[2] Something strange happens when we introduce multiplication. For example, the product of positive integers is positive: positive integers are closed under multiplication. However, the product of negative integers is not negative, it is positive: the negative integers are not closed under multiplication. What is the source of this asymmetry. *Why is the product of two negative integers positive?* This is probably the most mysterious question arising with the introduction of negative numbers.

The simplest answer to this question is probably one that involves negative iterations and inverse functions. We will discuss this explanation and others, including those involving algebraic laws that we expect $\mathbb{Z}$ to possess.

4.2 The net-difference equivalence relation

Before introducing the integers $\mathbb{Z}$, we will study a motivating concept:

Definition 1. The *net-difference function* $\Delta\colon \mathbb{N}\times\mathbb{N}\to\mathbb{N}\times\mathbb{N}$ is designed to measure the difference in sizes between two finite sets (of size m and n respectively). It is defined by the rule

$$\Delta(m,n)=\begin{cases}(m-n,0) & \text{if } m>n\\ (0,n-m) & \text{if } n>m\\ (0,0) & \text{if } m=n\end{cases}$$

This is well-defined by the trichotomy law for $\mathbb{N}$.

Informal Exercise 1. What is $\Delta(19,15)$? What is $\Delta(14,16)$?

Remark 1. Roughly speaking, you can think of $(x,0)$ as a positive difference, and $(0,x)$ as a negative difference (if $x\neq 0$). Later we will use this idea to define positive and negative integers in $\mathbb{Z}$, but we will work with equivalence classes of pairs, not the pairs themselves.

Exercise 2. Use the basic law of subtraction (Chapter 2) to show $m-0=m$. Use this to show that $\Delta(m,0)=(m,0)$ for all $m\in\mathbb{N}$. Don't forget to separate the case $m>0$ from the case $m=0$. Show that $\Delta(0,n)=(0,n)$.

[2] A fancy way of saying this is that the change of sign function yields an automorphism of the additive group of $\mathbb{Z}$.

Exercise 3. Use the above exercise to show that the second iteration of Δ is itself. In other words,

$$\Delta^2 = \Delta.$$

Lemma 1. *Let $m, n \in \mathbb{N}$. Then at least one of the coordinates of $\Delta(m, n)$ is zero. Furthermore,*
if $\Delta(m, n) = (x, 0)$ with $x > 0$ then $m > n$,
if $\Delta(m, n) = (0, x)$ with $x > 0$ then $n > m$, and
if $\Delta(m, n) = (0, 0)$ then $n = m$.

Proof. All the claims but the last follow directly from the definition. Recall from Chapter 2 that $n - m = 0$ implies $n = m$. So $\Delta(m, n) = (0, 0)$ is impossible if $m > n$ or if $n > m$. Thus $\Delta(m, n) = (0, 0)$ implies $m = n$. □

Definition 2 (Net-difference equivalence)**.** If $\Delta(m, n) = \Delta(m', n')$ we say that (m, n) and (m', n') are *net-difference equivalent*. In that case we write $(m, n) \sim (m', n')$. The relation $\sim$ is called *net-difference equivalence*.

Warning. The notation $\sim$ for net-difference equivalence is only really used in this chapter. The symbol $\sim$ will be used for other equivalence relations in other chapters.

Theorem 2. *Net-difference equivalence is an equivalence relation: it is reflexive, symmetric, and transitive.*

Exercise 4. Prove the above theorem.

We now discuss several lemmas in order to develop a variety of criteria to show that pairs are net-difference equivalent. These are summarized in Theorem 6 below.

Lemma 3. *For all $m, n, x \in \mathbb{N}$, $\Delta(m, n) = \Delta(m + x, n + x)$*

Proof. We divide into cases: $m > n$, $n > m$, $m = n$. First suppose $m > n$. Then $m + x > n + x$ by a result of Chapter 2. So $\Delta(m, n) = (m - n, 0)$ and $\Delta(m + x, n + x) = \big((m + x) - (n + x), 0\big)$. The result follows from the equality $(m + x) - (n + x) = m - n$ established in Chapter 2.

The proof in the other cases is similar. □

Lemma 4. *If $\Delta(m, n) = \Delta(m', n')$ then $m + n' = n + m'$.*

Proof. We divide into cases: $m > n$, $n > m$, $m = n$. First suppose $m > n$. Then $\Delta(m, n) = (x, 0)$ with $x > 0$ (since $m - n \neq 0$ by a result of Chapter 2). Thus $\Delta(m', n') = (x, 0)$ with $x > 0$. By Lemma 1, $m' > n'$. Also, $m - n = x = m' - n'$.

By the Basic Law of Subtraction (Chapter 2), $m = n + x$ and $m' = n' + x$. Thus, using arithmetic laws of Chapter 1,

$$m + n' = (n + x) + n' = n + (x + n') = n + (n' + x) = n + m'.$$

The cases where $n > m$ and $m = n$ are similar. □

Lemma 5. *Suppose $(m, n), (m', n') \in \mathbb{N} \times \mathbb{N}$ are such that $m + n' = n + m'$ and $n' \geq n$. Then there is an $x \in \mathbb{N}$ such that $n' = n + x$ and $m' = m + x$.*

Proof. Since $n' \geq n$, there is an $x \in \mathbb{N}$ such that $n' = n + x$ (Chapter 2). Our goal is to show $m' = m + x$ as well. Observe

$$n + m' = m + n' = m + (n + x) = n + (m + x).$$

The first equality is an assumption, the second is a substitution, and the last is a combination of the associative and commutative laws. The desired equation $m' = x + m$ follows from the cancellation law. □

Theorem 6. *Let $(m, n), (m', n') \in \mathbb{N} \times \mathbb{N}$. Then the following are equivalent:*
(i) $\Delta(m, n) = \Delta(m', n')$.
(ii) $m + n' = n + m'$.
(iii) Either $n' = n + x$ and $m' = m + x$ for some $x \in \mathbb{N}$, or $n = n' + x$ and $m = m' + x$ for some $x \in \mathbb{N}$.

Proof. We use the above lemmas to make an "implication circle".

For example (i) implies (ii) by Lemma 4.

Now if (ii) holds then divide into two cases: $n' \geq n$ and $n \geq n'$. In either case, Lemma 5 gives the result (with role reversal in the second case). So (ii) implies (iii).

Finally, (iii) implies (i) by Lemma 3. □

Note. The above theorem gives us three different ways to check for net-difference equivalence. In other words, the above theorem gives three characterizations for net-difference equivalence. To prove theorems about net-difference equivalence, one strategically uses the condition that makes the proof the easiest.

4.3 The integers $\mathbb{Z}$

Informally, if $n \neq 0$ then positive n is the equivalence class containing $(n, 0)$ and negative n is the equivalence class containing $(0, n)$.

Definition 3. Let $[m, n]$ be the equivalence class of (m, n) under net-difference equivalence.

The following allows us to use the arithmetic of natural numbers to prove things about these equivalence classes.

Theorem 7. *For all $m, n, m', n' \in \mathbb{N}$,*

$$[m, n] = [m', n'] \quad \Leftrightarrow \quad \Delta(m, n) = \Delta(m', n') \quad \Leftrightarrow \quad m + n' = n + m'$$

and

$$[m, n] = [\Delta(m, n)].$$

Exercise 5. Prove the above theorem. Hint: you might need to review equivalence classes from your set theory course. Also, use Theorem 6 and Exercise 3.

Definition 4 (Set of integers)**.** The *set of integers* $\mathbb{Z}$ is defined to be the set of equivalence classes under net-difference equivalence. In other words,

$$\mathbb{Z} = \{a \mid a = [m, n] \text{ for some } (m, n) \in \mathbb{N} \times \mathbb{N}\}.$$

Theorem 8. *If $a \in \mathbb{Z}$ then exactly one of the following hold:*

(i) $a = [0, 0]$.

(ii) $a = [n, 0]$ *for some* $n \in \mathbb{N}^+$.

(iii) $a = [0, n]$ *for some* $n \in \mathbb{N}^+$.

In addition, the natural number n in (ii) or (iii) is unique.

Proof. By definition of $\mathbb{Z}$ we have $a = [m, n]$ for some $m, n \in \mathbb{N}$. By theorem 7 we have $a = [\Delta(m, n)]$. From the definition of Δ we have that $\Delta(m, n)$ has at least one coordinate equal to zero. Thus at least one of (i), (ii), (iii) holds since $a = [\Delta(m, n)]$.

We will show that (i) and (ii) cannot both hold. Suppose otherwise that

$$[0, 0] = a = [n, 0]$$

for some $n \in \mathbb{N}^+$. This implies $0 + 0 = 0 + n$ by Theorem 7. By results of Chapter 1, we can simplify $0 + 0 = 0 + n$ to the equation $0 = n$. This contradicts $n \in \mathbb{N}^+$.

The proof that (i) and (iii) cannot both hold, and that (ii) and (iii) cannot both hold is similar. The proof that n is unique in case (ii) or case (iii) is also similar. □

Exercise 6. Show that (ii) and (iii) cannot both be true (even with different choices of n for (ii) and (iii)). Show that n is unique in case (ii).

Definition 5. Let $a \in \mathbb{Z}$. If a is $[0, 0]$ then a is called the *zero integer*. If a is $[n, 0]$ for $n \in \mathbb{N}^+$ then a is said to be *positive*. If a is $[0, n]$ for $n \in \mathbb{N}^+$ then a is said to be *negative*. From the above theorem, exactly one of these applies to each $a \in \mathbb{Z}$.

Note. The symbol $\mathbb{Z}$ is based on the German word *Zahlen* meaning 'numbers'. Some books write $\mathbf{Z}$ instead of $\mathbb{Z}$. In fact, the variant $\mathbb{Z}$ originated as a way to write a bold $\mathbf{Z}$ on the blackboard (without having to smash the chalk into the board to make the letter look bold). The letters $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are in a font style called *blackboard bold*.

4.4 Addition in $\mathbb{Z}$

In Chapter 3, addition in $\mathbb{N}$ is characterized by its ability to measure the size of a disjoint union of finite sets. We want to define an addition for $\mathbb{Z}$ with a similar capability. Now, individual elements of $\mathbb{Z}$ are used to measure the net difference resulting from the comparision of *two* sets A and B. We want addition in $\mathbb{Z}$ to measure the result of taking the disjoint union (in some sense) of one comparison A and B with another comparison A' and B'.

More specifically, if A is a set with m elements and if B is a set of n elements, then the integer $a = [m, n] = [\Delta(m, n)]$ measures the net difference of the size of A over B. It is positive when A is larger than B, it is negative when B is larger than A.

What happens if we simultaneously add m' elements to A, and n' elements to B where the new elements are in sets A' and B' disjoint from A and B respectively? Then we have added a net difference of $b = [m', n']$ to the sets A and B. We want our definition of $a + b$ in $\mathbb{Z}$ to represent the net differences of the sizes *after* adding the new elements. Since the first set has $m + m'$ elements after the addition, and the second has $n + n'$ elements, the resulting net difference is represented by $[m + m', n + n']$. This suggests that we define $a + b$ to be $[m + m', n + n']$.

Definition 6 (Addition)**.** Suppose $a, b \in \mathbb{Z}$ are integers such that $a = [m, n]$ and $b = [m', n']$. Then $a + b$ is defined to be $[m + m', n + n']$. The following lemma assures us that this definition is well-defined.

Remark 2. This definition brings up the subtle issue of "well-definedness". *Whenever you define a function or relation for equivalence classes, you must always check that the result depends only on the equivalence classes and not how they are represented.* For example, if $a = [14, 13]$ then we can describe a as $[11, 10]$, or $[101, 100]$ or in an infinite number of different ways. You want to make sure that any definition involving a depends only on a and not on the arbitrary numbers, 14 and 13 say, used to describe it. The above formula $[m + m', n + n']$ for addition seems to depend on the particular numbers! The following lemma shows that it does not.

Lemma 9. *If* $[m_1, n_1] = [m_2, n_2]$ *and* $[m_1', n_1'] = [m_2', n_2']$ *then*

$$[m_1 + m_1', n_1 + n_1'] = [m_2 + m_2', n_2 + n_2'].$$

Proof. By Theorem 7, $m_1 + n_2 = n_1 + m_2$ and $m_1' + n_2' = n_1' + m_2'$. So

$$\begin{aligned}(m_1 + m_1') + (n_2 + n_2') &= (m_1 + n_2) + (m_1' + n_2') \\ &= (n_1 + m_2) + (n_1' + m_2') \\ &= (n_1 + n_1') + (m_2 + m_2').\end{aligned}$$

The result now follows from Theorem 7. □

4.5 $\mathbb{Z}$ as an abelian group

The first main result concerning $\mathbb{Z}$ is that it is an *abelian group* under addition. This is just a fancy way of saying that (i) $\mathbb{Z}$ has an addition $\mathbb{Z}\times\mathbb{Z}\to\mathbb{Z}$ that is associative, (ii) that there is an additive identity (zero) element in $\mathbb{Z}$, and (iii) that every element $a\in\mathbb{Z}$ has an additive inverse.

The combination of properties mentioned above (associativity, identity, and inverse) is so common in mathematics, that there is a name for something that possesses them, a *group*. Forget the informal meaning of the word *group* used in everyday life; from now on it will be a technical term referring to a set with the combination of properties mentioned above.[3] If we throw in the commutative law for the binary operation of the group, we call the group *abelian* (after the famous mathematician Abel).

Definition 7 (Group)**.** A group G is a set together with a binary operation

$$*: G\times G\to G$$

such that

(i) $*$ is associative: $a*(b*c)=(a*b)*c$ for all $a,b,c\in G$.

(ii) G has an identity element. In other words, there is an element $e\in G$ such that $e*a=a*e=a$ for all $a\in G$. (It is easy to prove that the identity is unique).

(iii) Every element of G has an inverse. In other words, if $a\in G$ then there is a $b\in G$ such that $a*b=b*a=e$ where e is an identity for G. (It is easy to prove that the inverse of a is unique).

Informal Exercise 7. In many examples, $*$ is written $+$. Is $\mathbb{N}$ a group under $+$? What about $\mathbb{R}$ and $\mathbb{Q}$?

Definition 8 (Abelian group)**.** A group G is said to be *abelian* if the commutative law holds: $a*b=b*a$ for all $a,b\in G$.

In the previous section, we defined a binary operation $+:\mathbb{Z}\times\mathbb{Z}\to\mathbb{Z}$. In an effort to show that $\mathbb{Z}$ is a group under $+$, we investigate some of the properties of $+$. The fact that the following proofs are so easy is due to our decision to define addition with equivalence classes. The naive definition of $\mathbb{Z}$ would result in lots of special cases.

Theorem 10 (Associative law)**.** *If* $a,b,c\in\mathbb{Z}$ *then* $(a+b)+c=a+(b+c)$.

Proof. Write $a=[m_1,n_1], b=[m_2,n_2]$, and $c=[m_3,n_3]$. Since each m_i and n_i is in $\mathbb{N}$, we can use the associative law of Chapter 1:

$$(m_1+m_2)+m_3=m_1+(m_2+m_3)\quad\text{and}\quad(n_1+n_2)+n_3=n_1+(n_2+n_3).$$

[3] Mathematics majors typically study groups in more detail in an upper-division (abstract) algebra course.

Thus, by substitution,

$$\big[(m_1+m_2)+m_3,(n_1+n_2)+n_3\big]=\big[m_1+(m_2+m_3),n_1+(n_2+n_3)\big].$$

So, using the above and applying the definition of addition several times,

$$\begin{aligned}(a+b)+c &= \big([m_1,n_1]+[m_2,n_2]\big)+\big[m_3,n_3\big]\\ &= [m_1+m_2,n_1+n_2]+\big[m_3,n_3\big]\\ &= \big[(m_1+m_2),(n_1+n_2)\big]+\big[m_3,n_3\big]\\ &= \big[(m_1+m_2)+m_3,(n_1+n_2)+n_3\big]\\ &= \big[m_1+(m_2+m_3),n_1+(n_2+n_3)\big]\\ &= [m_1,n_1]+\big[(m_2+m_3),(n_2+n_3)\big]\\ &= [m_1,n_1]+[m_2+m_3,n_2+n_3]\\ &= [m_1,n_1]+\big([m_2,n_2]+[m_3,n_3]\big)=a+(b+c).\end{aligned}$$

□

Theorem 11 (Commutative law)**.** *If $a,b\in\mathbb{Z}$ then $a+b=b+a$.*

Exercise 8. Prove the commutative law using the previous proof as a model.

Theorem 12 (Identity law)**.** *If $a\in\mathbb{Z}$ then $a+[0,0]=a$.*

This shows that $[0,0]$ is an identity for $\mathbb{Z}$. (Because of the commutative law, we don't have to show $[0,0]+a=a$).

Definition 9. If $a=[m,n]$ is an element of $\mathbb{Z}$ then $-a$ is defined as $[n,m]$.

Actually, we cannot use this definition until we prove that is is well-defined.

Lemma 13. *If $[m,n]=[m',n']$ then $[n,m]=[n',m']$.*

Proof. If $[m,n]=[m',n']$ then $m+n'=n+m'$. So $n+m'=m+n'$. Thus $[n,m]=[n',m']$. (Theorem 7). □

Theorem 14 (Inverse law)**.** *If $a\in\mathbb{Z}$ then $a+(-a)=[0,0]$.*

Exercise 9. Show that $[n,n]=[0,0]$ for all $n\in\mathbb{N}$. Prove the identity and the inverse law.

The above theorems combine to give us the following. (Because of the commutative law, we don't have show $a+(-a)=(-a)+a$.)

Theorem 15. *The set $\mathbb{Z}$ is an abelian group under addition $+$.*

Definition 10 (Subtraction)**.** If $a,b\in\mathbb{Z}$ then $a-b$ is shorthand for $a+(-b)$. In particular $a-a=[0,0]$. This $-$ should not be confused with subtraction defined in Chapter 2. We will discuss the compatibility in the two types of $-$ in the next section.

Note. If G is a group, then the notation g^{-1} and $-g$ are both used for the inverse of $g \in G$. If the operation for G written '+', then $-g$ is usually used to signify the inverse (as we did in $\mathbb{Z}$ above), and $a - b$ is shorthand for $a + (-b)$.

Every abelian group satisfies the cancellation law.

Theorem 16. *Suppose $a, b, c \in \mathbb{Z}$, or more generally suppose $a, b, c \in G$ where G is a group under addition $+$. Then, if $a + c = b + c$ then $a = b$.*

Exercise 10. Use the inverse of c to prove the above theorem.

Theorem 17. *Suppose $a \in \mathbb{Z}$, or more generally suppose $a \in G$ where G is a group under addition $+$. Then $-(-a) = a$.*

Proof. Observe that $a + (-a)$ and $(-(-a)) + (-a)$ are both equal to the identity by the definition of inverse. Hence they are equal to each other:

$$(-(-a)) + (-a) = a + (-a).$$

Now use the cancellation law. □

One advantage of using integers is that you can always solve equations of the form $x + a = b$.

Theorem 18. *Suppose $a, b \in G$ where G is a group under addition $+$ (for example $G = \mathbb{Z}$). Then, the equation $x + a = b$ has a unique solution in G. The solution is $x = b - a$.*

Exercise 11. Prove the above theorem.

Informal Exercise 12. If $m, n \in \mathbb{N}$, does the equation $x + m = n$ always have a solution for x in $\mathbb{N}$? Give necessary and sufficient conditions for the solution to exist.

Exercise 13. Suppose $a, b \in G$ where G is an abelian group under addition $+$. Show that $-(a + b) = (-a) + (-b)$. We often write this with the parentheses removed as

$$-(a + b) = -a - b.$$

(But be sure to use parentheses in your proof.)

Exercise 14. Show that if e is the identity element of a group, then e is its own inverse.

4.6 The canonical embedding of $\mathbb{N}$ in $\mathbb{Z}$

According to our formal definition, $\mathbb{N}$ is not a subset of $\mathbb{Z}$. However, we would like to think of $\mathbb{N}$ as a subset of $\mathbb{Z}$. What we do is define an injection of $\mathbb{N}$ into $\mathbb{Z}$, and then "identify" a natural number with its image in $\mathbb{Z}$.

Definition 11. Consider the function $\mathbb{N} \to \mathbb{Z}$ defined by the rule $n \mapsto [n, 0]$. Call this the *canonical embedding* of the natural numbers into the integers.

Theorem 19. *The canonical embedding $\mathbb{N} \to \mathbb{Z}$ is injective.*

Proof. If $[n_1, 0] = [n_2, 0]$ then $n_1 + 0 = 0 + n_2$ by Theorem 7. □

Whenever we have an injective function $f : A \to B$, the image $A' = f[A]$ is a subset of B and the function f gives a bijection $A \to A'$ (using restriction of codomain). We can use this bijection to match each element $a \in A$ with an element $a' \in A'$. We can go further, and *identify* the element a with its image a'. When we do this, we think of a' as a copy or "clone" of a. This allows us to think of A as being in some sense equal to A', and allows us to think of A as a subset of B.

We apply this idea to the canonical embedding $\mathbb{N} \to \mathbb{Z}$. We identify $n \in \mathbb{N}$ with its image $[n, 0]$ in $\mathbb{Z}$. We treat n and $[n, 0]$ as the same object, and we think of $\mathbb{N}$ as a subset of $\mathbb{Z}$.

Warning. When we identify the elements of $\mathbb{N}$ with elements of $\mathbb{Z}$ we risk ambiguous definitions and notation. For example, $m + n$ can mean two things if $m, n \in \mathbb{N}$. First, it can mean $m + n$ as defined in Chapter 1: we write $m +_{\mathbb{N}} n$ in this case. Second, it can mean $m + n$, or rather $[m, 0] + [n, 0]$ as defined in the current chapter: we write $m +_{\mathbb{Z}} n$ in this case.

It turns out that both types of addition are equal if $m, n \in \mathbb{N}$ and so we do not really need to make a distinction. This is shown by the following:

Theorem 20 (Extension of addition)**.** *If we consider $\mathbb{N}$ as a subset of $\mathbb{Z}$, and if $m, n \in \mathbb{N}$, then $m +_{\mathbb{N}} n$ and $m +_{\mathbb{Z}} n$ define the same element of $\mathbb{Z}$.*

Proof. The definition of addition for $\mathbb{Z}$ (Definition 6) implies

$$[m, 0] +_{\mathbb{Z}} [n, 0] = [m +_{\mathbb{N}} n,\ 0 +_{\mathbb{N}} 0] = [m +_{\mathbb{N}} n,\ 0].$$

Since m is identified with $[m, 0]$ and n with $[n, 0]$, and furthermore $m +_{\mathbb{N}} n$ with $[m +_{\mathbb{N}} n, 0]$, we can rewrite this equation as

$$m +_{\mathbb{Z}} n = m +_{\mathbb{N}} n.$$

□

When thinking of $\mathbb{N}$ as a subset of $\mathbb{Z}$, we call $+_{\mathbb{Z}}$ an *extension* of $+_{\mathbb{N}}$ since we originally only had an addition on $\mathbb{N}$ (Chapter 1), but we defined this addition to the larger set $\mathbb{Z}$ without changing the values on the subset $\mathbb{N}$. From now on we will not make a difference between the two types of additions since they agree whenever both are defined.

Theorem 21 (Extension of subtraction)**.** *The definition of subtraction for $\mathbb{Z}$ extends the definition of subtraction for $\mathbb{N}$. In other words, if $m, n \in \mathbb{N}$ with $m \leq n$, then $n - m$, as defined in Chapter 2, agrees with $n + (-m)$ as defined in the current chapter.*

Proof. By definition of subtraction in Chapter 2, $n-m = b$ where $n = m+b$. To show that the two definitions of subtraction agree, we will establish the equation $[n, 0] + (-[m, 0]) = [b, 0]$. Observe that

$$\begin{aligned} [n,0] + (-[m,0]) &= [n,0] + [0,m] && \text{(Def. of add. inverse)} \\ &= [n,m] && \text{(Def. of addition)} \\ &= [m+b,m] && (n = m+b) \\ &= [b,0] && \text{(Theorem 7)} \end{aligned}$$

The last step can be justified by the equation $(m + b) + 0 = m + b$. □

Multiplication for $\mathbb{Z}$ has not been defined yet, but when it is we will check that it extends the multiplication for $\mathbb{N}$.

Remark 3. Since we identify $0 \in \mathbb{N}$ with $[0, 0] \in \mathbb{Z}$, we can use the symbol '0' for both. In particular, by Theorem 12 (the identity law),

$$a + 0 = a \qquad \text{for all } a \in \mathbb{Z}.$$

Similarly, by Theorem 14 (inverse law),

$$a + (-a) = 0 \qquad \text{for all } a \in \mathbb{Z}.$$

Finally, by Exercise 14,

$$-0 = 0.$$

In the following two theorems we use the the canonical embedding to view $\mathbb{N}$ as a subset of $\mathbb{Z}$.

Theorem 22. *If $a \in \mathbb{Z}$ then exactly one of the following occurs: (i) $a = 0$, (ii) $a = n$ for $n \in \mathbb{N}^+$, or (iii) $a = -n$ for $n \in \mathbb{N}^+$. Furthermore, the n occurring in case (ii) or (iii) is unique.*

Proof. Let $a \in \mathbb{Z}$. By Theorem 8 either (i) $a = [0, 0]$, (ii) $a = [n, 0]$ with $n \in \mathbb{N}^+$, or (iii) $a = [0, n]$ with $n \in \mathbb{N}^+$. In case (i), $[0, 0]$ is identified with 0, so $a = 0$. In case (ii), $[n, 0]$ is identified with n, so $a = n$. In case (iii), $a = [0, n] = -[n, 0]$ and $[n, 0]$ is identified with n, so $a = -n$.

To show at most one case occurs, and to show uniqueness of n, use Theorem 8 together with the fact that $-n$ is identified with $-[n, 0] = [0, n]$. □

Note. We can use this theorem to rephrase Definition 5 as follows. Integers of the form $a = n$ with $n \in \mathbb{N}^+$ are called *positive integers*. Integers of the form $a = -n$ with $n \in \mathbb{N}^+$ are called *negative integers*. Every integer is either zero, positive, or negative.

Note. Theorem 22 shows that the integers, as we have formally constructed them, give the same integers as the informal definition mentioned in the introductory section of this chapter.

Corollary 23. *If $a \in \mathbb{Z}$ then exactly one of the following occurs:* (i) $a \in \mathbb{N}$, *or* (ii) $a = -n$ *for a unique* $n \in \mathbb{N}^+$.

Theorem 24. *Suppose $a, b \in \mathbb{Z}$. Then $a - b = 0$ if and only if $a = b$.*

Proof. Suppose that $a + (-b) = 0$. Then $\big(a + (-b)\big) + b = 0 + b$. Thus

$$\begin{aligned} b &= b + 0 && \text{(Identity Law)} \\ &= 0 + b && \text{(Commutative Law)} \\ &= \big(a + (-b)\big) + b && \text{(as above)} \\ &= a + \big((-b) + b\big) && \text{(Associative law)} \\ &= a + \big(b + (-b)\big) && \text{(Commutative law)} \\ &= a + 0 && \text{(Inverse Law)} \\ &= a && \text{(Identity law)} \end{aligned}$$

Conversely, if $a = b$, then $a + (-b) = b + (-b) = 0$. □

4.7 Order in $\mathbb{Z}$

From now on we will view $\mathbb{N}$ as a subset of $\mathbb{Z}$. Thus the positive integers $\mathbb{N}^+$ form a subset of $\mathbb{Z}$ as well. We can use positive integers to define the concept of *order* just as we did in Chapter 2.

Definition 12. Suppose $a, b \in \mathbb{Z}$. Then $a < b$ is defined to mean that there is an element $x \in \mathbb{N}^+$ such that $b = a + x$.

Note. Since this is essentially the same definition as in Chapter 2, we observe that the order relation $<$ on $\mathbb{Z}$ extends the order $<$ on $\mathbb{N}$.

Theorem 25. *Suppose $a, b \in \mathbb{Z}$. Then $a < b$ if and only if $b - a \in \mathbb{N}^+$.*

Exercise 15. Prove the above theorem. Do not use subtraction as presented in Chapter 2, but use additive inverses instead.

Definition 13. Suppose $a, b \in \mathbb{Z}$. Then $a \leq b$ means either $a < b$ or $a = b$.

Theorem 26. *Suppose $a, b \in \mathbb{Z}$. Then $a \leq b$ if and only if $b - a \in \mathbb{N}$.*

Theorem 27. *Suppose $a, b, c \in \mathbb{Z}$. If $a < b$ then $a + c < b + c$. If $a \leq b$ then $a + c \leq b + c$.*

Theorem 28. *Let $a \in \mathbb{Z}$. Then $a \in \mathbb{N}$ if and only if $a \geq 0$. Also, a is positive if and only if $a > 0$, and a is negative if and only if $a < 0$.*

Theorem 29. *Transitivity for $<$ holds.*

Theorem 30. *Transitivity for $\leq$ hold. Mixed transitivity for $<$ and $\leq$ hold.*

Theorem 31. *Trichotomy for $<$ holds.*

Note. Since trichotomy and transitivity hold for $<$, it is a linear order.

Exercise 16. Write up three proofs for the above six theorems. (You should be able to do all of them, but write up three of them.)

Theorem 32. *Let $a, b \in \mathbb{Z}$. If $a < b$ then $-b < -a$. If $a \leq b$ then $-b \leq -a$.*

Proof. (sketch) Suppose $a < b$. Then $a + \big((-a) + (-b)\big) < b + \big((-a) + (-b)\big)$ by Theorem 27. Using the laws proved so far, the left-hand side simplifies to $-b$ and the right-hand side simplifies to $-a$.

A similar argument holds for $\leq$. □

Corollary 33. *Let $c \in \mathbb{Z}$. If $c > 0$ then $-c < 0$. If $c < 0$ then $-c > 0$. If $c \geq 0$ then $-c \leq 0$. If $c \leq 0$ then $-c \geq 0$.*

Proof. This makes use of the fact that $-0 = 0$ (see Remark 3). □

Theorem 34. *Let $a \in \mathbb{Z}$. There is no integer x with $a < x < a + 1$.*

Proof. Suppose $a < x < a + 1$. Then

$$a + (-a) < x + (-a) < (a + 1) + (-a).$$

Thus $0 < x - a < 1$. In particular, $x - a \in \mathbb{N}$, but we showed in Chapter 2 that there is no natural number between 0 and 1. □

4.8 Iteration by $a \in \mathbb{Z}$

We now investigate the concept of negative iteration. An example where iteration is useful is in the definition of multiplication. Recall that in Chapter 1 multiplication is defined in terms of iteration of addition. If we can figure out a way to define negative iteration then we can explain what it means to multiply by a negative number.

We developed properties of f^n in Chapter 1 for $n \geq 0$. In this section and the next we will define f^a for all $a \in \mathbb{Z}$, and show that the properties of Chapter 1 extend to this more general situation.

We already know how to define f^{-1}. It is just the inverse function. How do we define f^a for other negative a? Informally, think of f^{-n} as the nth iterate of the inverse function f^{-1}. Recall that only bijective functions have inverses, so we will not try to define f^{-n} for functions that are not bijections.

Informal Definition 14. Suppose $f : S \to S$ is a bijection. If $a = n$ is a positive integer, then f^n is the nth iterate of f. If $a = -n$ is a negative integer, then $f^a = f^{-n}$ is the nth iterate of the inverse f^{-1}. If $a = 0$ then f^a is the identity function $id : S \to S$.

Warning. We use f^{-1} to refer to the inverse of f, *not* to $1/f$. Similarly, f^2 refers to $f \circ f$, *not* to the product $f \cdot f$ of the function with itself.

The above definition is informal. Our formal definition will use equivalence classes to give a common definition for all cases at once. It may seem more elaborate, but it will be more convenient for proving theorems. Before giving the formal definition, we give a preliminary theorem.

Theorem 35. *If $n \in \mathbb{N}$ and if $f : S \to S$ is a bijection, then f^n and $\left(f^{-1}\right)^n$ are also bijections. Furthermore, the inverse of f^n is $(f^{-1})^n$. So*

$$(f^n)^{-1} = \left(f^{-1}\right)^n.$$

Proof. First we will show that $f^n \circ \left(f^{-1}\right)^n = id$ where $id : S \to S$ is the identity function. Let A be the set of $n \in \mathbb{N}$ such that $f^n \circ \left(f^{-1}\right)^n = id$.

First we show $0 \in A$. Observe that f^0 and $\left(f^{-1}\right)^0$ are both the identity function, and the composition of the identity function with itself is just the identity function. So $0 \in A$.

Now suppose $u \in A$. We must show that $u + 1 \in A$. By a result of Chapter 3,

$$f^{u+1} = f^u \circ f^1 = f^u \circ f.$$

Similarly,

$$\left(f^{-1}\right)^{u+1} = \left(f^{-1}\right)^{1+u} = \left(f^{-1}\right)^1 \circ \left(f^{-1}\right)^u = f^{-1} \circ \left(f^{-1}\right)^u.$$

So

$$\begin{aligned} f^{u+1} \circ \left(f^{-1}\right)^{u+1} &= \left(f^u \circ f\right) \circ \left(f^{-1} \circ \left(f^{-1}\right)^u\right) \\ &= \left(\left(f^u \circ f\right) \circ f^{-1}\right) \circ \left(f^{-1}\right)^u \\ &= \left(f^u \circ \left(f \circ f^{-1}\right)\right) \circ \left(f^{-1}\right)^u \\ &= \left(f^u \circ id\right) \circ \left(f^{-1}\right)^u \\ &= f^u \circ \left(f^{-1}\right)^u. \end{aligned}$$

In the second and third equalities we used the fact that function composition is associative (Chapter 0). Since $u \in A$, we have $f^u \circ \left(f^{-1}\right)^u = id$. Combining this with the above gives

$$f^{u+1} \circ \left(f^{-1}\right)^{u+1} = id.$$

Thus $u+1 \in A$.

By the induction axiom, $A = \mathbb{N}$. So $f^n \circ \left(f^{-1}\right)^n = id$ for all $n \in \mathbb{N}$.

A similar argument show that $\left(f^{-1}\right)^n \circ f^n = id$ for all $n \in \mathbb{N}$. By the definition of inverse function, we have that f^n and $\left(f^{-1}\right)^n$ are inverse functions. Finally, functions that have inverses must be bijections. □

Definition 15 (General iteration)**.** Let $f : S \to S$ be a bijection. Suppose that $a \in \mathbb{Z}$, and that $a = [m, n]$. Then

$$f^a \stackrel{\text{def}}{=} f^m \circ \left(f^{-1}\right)^n .$$

The term on the left refers to the new type of iteration, and the terms on the right use the old (Chapter 1) type of iteration.

The following lemma shows that this definition is well-defined: it doesn't matter what pair (m, n) is used to write the same a. Such a lemma is essential whenever we use equivalence classes.

Lemma 36. *Let $m, n, m', n' \in \mathbb{N}$, and let $f : S \to S$ be a bijection. If $[m, n] = [m', n']$ then*

$$f^m \circ \left(f^{-1}\right)^n = f^{m'} \circ \left(f^{-1}\right)^{n'} .$$

Proof. (Sketch) Since $[m, n] = [m', n']$ we have $\Delta(m, n) = \Delta(m', n')$. Without loss of generality, assume $m' \geq m$. So $m' = m + x$ and $n' = n + x$ for some $x \in \mathbb{N}$ (Theorem 6). Thus

$$\begin{aligned} f^{m'} \circ \left(f^{-1}\right)^{n'} &= f^{m+x} \circ \left(f^{-1}\right)^{x+n} \\ &= f^m \circ f^x \circ \left(f^{-1}\right)^x \circ \left(f^{-1}\right)^n \quad (f^{m+x} = f^m \circ f^x \text{: see Ch. 2}) \\ &= f^m \circ \left(f^{-1}\right)^n \qquad \text{(Thm. 35)}. \end{aligned}$$

(We can leave out parentheses since functional composition is associative.) □

Exercise 17. If $f : S \to S$ is bijective, and if $a \in \mathbb{Z}$, show that f^a is also a bijective function $S \to S$ using Definition 15.

Warning. We now have two types of iteration: that from Chapter 1, and that defined in Definition 15 (which made use of the earlier type of iteration). We will show that the new type of iteration extends the earlier type.

Lemma 37. *The new type of iteration extends the earlier type of iteration. In other words, if $f : S \to S$ is a bijection and $n \in \mathbb{N}$ then both definitions give the same result for f^n.*

In addition, if we use the new definition for f^{-1}, the result agrees with the old definition. In other words, f^{-1} according to the new definition is just the inverse function.

Proof. Let $n \in \mathbb{N}$. In $\mathbb{Z}$, the integer n is identified with the equivalence class $[n, 0]$. Observe

$$f^{[n,0]} = f^n \circ \left(f^{-1}\right)^0 = f^n \circ id = f^n$$

where the right hand side is as in Chapter 1. Thus both types of iteration agree for $n \in \mathbb{N}$.

In $\mathbb{Z}$, the number -1 is identified with the equivalence class $-[1, 0]$ which is $[0, 1]$. Observe

$$f^{[0,1]} = f^0 \circ \left(f^{-1}\right)^1 = id \circ \left(f^{-1}\right)^1 = \left(f^{-1}\right)^1 = f^{-1}.$$

So the new definition of f^{-1} gives the inverse. □

The following shows that f^{-n} is what we expect.

Theorem 38. *Suppose $f : S \to S$ is bijective. If $n \in \mathbb{N}$ then*

$$f^{-n} = \left(f^{-1}\right)^n = (f^n)^{-1}.$$

Proof. If $n \in \mathbb{N}$ then $-n$ is $[0, n]$. So

$$f^{-n} = f^{[0,n]} = f^0 \circ \left(f^{-1}\right)^n = id \circ \left(f^{-1}\right)^n = \left(f^{-1}\right)^n.$$

Note that $\left(f^{-1}\right)^n = (f^n)^{-1}$ by Theorem 35. □

We will generalize the above result in the next section (Corollary 44 and Corollary 47).

Iteration and the commutativity of composition

We now derive several properties concerning iteration, especially those associated with the idea of commutativity of composition. The most significant identity concerning iteration is the additive identity:

$$f^{a+b} = f^a \circ f^b$$

where $a, b \in \mathbb{Z}$ (already proved for $a, b \in \mathbb{N}$ in Chapter 3). This gives an important application for addition in $\mathbb{Z}$: it describes the resulting iteration associated with the composition of two iterations. It gives evidence that our definition of $+$ was "the right one".

Definition 16. Let $f : S \to S$ and $g : S \to S$ be functions. We say that f *and* g *commute* if $f \circ g = g \circ f$. We also say f *commutes with* g and that g *commutes with* f.

Exercise 18. Suppose $f : S \to S$ is a bijection. Verify that f and $g = f^{-1}$ commute. Also, verify that the identity function $id : S \to S$ commutes with all functions $f : S \to S$. (Hint: your proofs should be very short.)

Informal Exercise 19. Find two functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ that do not commute. Hint: try polynomial functions (that are not monomials).

If you know about linear transformations and their matrices you can use matrixes to find examples of non-commuting functions $\mathbb{R}^2 \to \mathbb{R}^2$.

Lemma 39. *Suppose $f : S \to S$ and $g : S \to S$ are functions that commute. Then f^n and g commute for all $n \in \mathbb{N}$.*

Proof. Let A be the set of all $n \in \mathbb{N}$ such that f^n and g commute. We know that $0 \in A$ since f^0 is just the identity function.

Suppose $u \in A$. We must show that $u + 1 \in A$. Observe that

$$\begin{aligned}
f^{u+1} \circ g &= \big(f \circ f^u\big) \circ g && \text{(From Chapter 1)} \\
&= f \circ \big(f^u \circ g\big) && \text{(Set theory: associativity of composition)} \\
&= f \circ \big(g \circ f^u\big) && \text{(Since } u \in A\text{)} \\
&= \big(f \circ g\big) \circ f^u && \text{(Set theory: associativity of composition)} \\
&= \big(g \circ f\big) \circ f^u && \text{(Since } f \text{ and } g \text{ commute)} \\
&= g \circ \big(f \circ f^u\big) && \text{(Set theory: associativity of composition)} \\
&= g \circ f^{u+1} && \text{(From Chapter 1).}
\end{aligned}$$

Thus $u + 1 \in A$.

By the induction axiom, $A = \mathbb{N}$. The result follows. □

Theorem 40. *Suppose $f : S \to S$ and $g : S \to S$ are functions that commute. Then f^m and g^n commute for all $m, n \in \mathbb{N}$.*

Proof. By the above lemma, f^m and g commute. By the above lemma applied to g and f^m, we get that g^n and f^m commute. □

We can extend the above to negative iterations. This is done in the following lemma and theorem.

Lemma 41. *Suppose that $f : S \to S$ and $g : S \to S$ are functions that commute. If f is bijective, then f^{-m} and g^n commute for all $m, n \in \mathbb{N}$.*

Proof. By Theorem 40, $g^n \circ f^m = f^m \circ g^n$. Thus

$$f^{-m} \circ g^n \circ f^m \circ f^{-m} = f^{-m} \circ f^m \circ g^n \circ f^{-m}$$

(parentheses can be left off since function composition is associative). We know that f^m and f^{-m} are inverses by Theorem 38. Thus the above equation simplifies to $f^{-m} \circ g^n = g^n \circ f^{-m}$. □

Theorem 42. *Suppose that $f : S \to S$ and $g : S \to S$ are bijections that commute. Then f^a and g^b commute for all $a, b \in \mathbb{Z}$.*

Proof. If neither a, b are negative use Theorem 40. If one of a, b is negative use Lemma 41.

Suppose that a and b are negative where $a = -m$ and $b = -n$. Lemma 41 shows that f and g^{-1} commute (switching the roles of f and g). Thus, by Lemma 41 again, f^{-m} and $\left(g^{-1}\right)^n$ commute. However, $\left(g^{-1}\right)^n = g^{-n}$ by Theorem 38. □

The following is the key theorem of this section.

Theorem 43. *Suppose $f : S \to S$ is bijective and that $a, b \in \mathbb{Z}$. Then*

$$f^{a+b} = f^a \circ f^b.$$

Proof. Write $a = [m, n]$ and $b = [m', n']$. Thus $a + b = [m + m', n + n']$. Let g be the inverse of f. So

$$\begin{aligned}
f^{a+b} &= f^{[m+m',n+n']} && \text{(Def. 6)} \\
&= f^{m+m'} \circ g^{n+n'} && \text{(Def. 15)} \\
&= \left(f^m \circ f^{m'}\right) \circ \left(g^n \circ g^{n'}\right) && \text{(Chapter 3)} \\
&= \left(f^m \circ \left(f^{m'} \circ g^n\right)\right) \circ g^{n'} && \text{(Assoc. of } \circ\text{.: twice)} \\
&= \left(f^m \circ \left(g^n \circ f^{m'}\right)\right) \circ g^{n'} && \text{(Thm. 40)} \\
&= \left(f^m \circ g^n\right) \circ \left(f^{m'} \circ g^{n'}\right) && \text{(Assoc. of } \circ\text{: twice)} \\
&= f^{[m,n]} \circ f^{[m',n']} = f^a \circ f^b && \text{(Def. 15)}
\end{aligned}$$

□

The following generalizes part of Theorem 38.

Corollary 44. *Suppose $f : S \to S$ is bijective and that $a \in \mathbb{Z}$. Then*

$$f^{-a} = \left(f^a\right)^{-1}.$$

Proof. We will prove this by showing that f^a and f^{-a} are inverse functions, in other words, we show that $f^{-a} \circ f^a$ and $f^a \circ f^{-a}$ are both the identity function. By Theorem 43,

$$f^{-a} \circ f^a = f^{-a+a} = f^0 \qquad \text{and} \qquad f^a \circ f^{-a} = f^{a-a} = f^0.$$

Since f^0 is the identity function, we see that f^a and f^{-a} are inverse functions. In other words, $\left(f^a\right)^{-1} = f^{-a}$. □

Suppose $f : S \to S$ and $g : S \to S$ are bijective functions. We usually do not expect that $(f \circ g)^a = f^a \circ g^a$. However, there is a case where this does indeed happen. We begin (Lemma 45) with the non-negative case.

Informal Exercise 20. Find polynomial functions

$$f : \mathbb{R} \to \mathbb{R} \text{ and } g : \mathbb{R} \to \mathbb{R}$$

such that $(f \circ g)^2$ does not equal $f^2 \circ g^2$.

Lemma 45. *Suppose $f : S \to S$ and $g : S \to S$ commute. Then, for all $n \in \mathbb{N}$,*

$$(f \circ g)^n = f^n \circ g^n.$$

Exercise 21. Prove the above using induction. Where did you use the hypothesis that f and g commute?

Theorem 46. *Suppose $f : S \to S$ and $g : S \to S$ are bijective functions, and suppose $a \in \mathbb{Z}$. If f and g commute, then*

$$(f \circ g)^a = f^a \circ g^a.$$

Proof. If $a \geq 0$ then use Lemma 45. If $a < 0$ then $a = -n$ where $n \in \mathbb{N}$, and

$$\begin{aligned}
(f \circ g)^n \circ (f^a \circ g^a) &= (f^n \circ g^n) \circ (f^a \circ g^a) && \text{(Lemma 45)} \\
&= \left(f^n \circ (g^n \circ f^a)\right) \circ g^a && \text{(by assoc. of } \circ\text{, twice)} \\
&= \left(f^n \circ (f^a \circ g^n)\right) \circ g^a && \text{(Thm. 42)} \\
&= (f^n \circ f^a) \circ (g^n \circ g^a) && \text{(by assoc. of } \circ\text{, twice)} \\
&= f^{n+a} \circ g^{n+a} && \text{(by Thm. 43)} \\
&= f^0 \circ g^0 = id && \text{(since } a = -n\text{)}
\end{aligned}$$

A similar argument show that $(f^a \circ g^a) \circ (f \circ g)^n$ is the identity. So $(f \circ g)^n$ and $(f^a \circ g^a)$ are inverse functions, and

$$(f \circ g)^a = (f \circ g)^{-n} = \left((f \circ g)^n\right)^{-1} = (f^a \circ g^a).$$

□

The following generalizes part of Theorem 38.

Corollary 47. *Suppose $f : S \to S$ is bijective and that $a \in \mathbb{Z}$. Then*

$$f^{-a} = \left(f^{-1}\right)^a.$$

Proof. Since f and f^{-1} commute, Theorem 46 gives

$$f^a \circ \left(f^{-1}\right)^a = \left(f \circ f^{-1}\right)^a = (id)^a = id.$$

The last step uses Lemma 48 (below). Similarly, $\left(f^{-1}\right)^a \circ f^a$ is the identity. Thus f^a and $\left(f^{-1}\right)^a$ are inverses. In other words,

$$(f^a)^{-1} = \left(f^{-1}\right)^a.$$

But $f^{-a} = (f^a)^{-1}$ (Corollary 44), so $f^{-a} = \left(f^{-1}\right)^a$. □

Lemma 48. *If $id : S \to S$ is the identity, then $id^a = id$ for all $a \in \mathbb{Z}$.*

Exercise 22. Prove the above. First show it by induction for all $a \geq 0$. Next show it for $a < 0$ by appealing to the fact that $a = -n$ for some $n \in \mathbb{N}$, so $id^a = id^{-n} = (id^n)^{-1}$.

4.9 Multiplication in $\mathbb{Z}$

Recall that in Chapter 1, multiplication was defined as iterated addition. More precisely, if $\alpha_m : \mathbb{N} \to \mathbb{N}$ is the function $x \mapsto x + m$, then $m \cdot n$ was defined to be $\alpha_m^n(0)$. In Section 4.8 of the current chapter we extended iteration to negative iteration. We can use this idea to define multiplication by a negative integer.

In particular, let $a, b \in \mathbb{Z}$. Define the function $A_a : \mathbb{Z} \to \mathbb{Z}$ by the rule $x \mapsto x + a$. (We use 'A' instead of 'α' to indicate that the domain is $\mathbb{Z}$ and not just $\mathbb{N}$, and that a is allowed to be negative). We propose

$$a \cdot b \stackrel{\text{def}}{=} A_a^b(0).$$

For this definition to make sense we must show that A_a is bijective.

Since the theory of multiplication is so heavily dependent on A_a, we will take a break and study A_a itself, also called a *translation function.*

Translation functions

Definition 17. Let $a \in \mathbb{Z}$. Then the *translation function* by a is defined to be the function $A_a : \mathbb{Z} \to \mathbb{Z}$ defined by the rule $x \mapsto x + a$.

Theorem 49. *The function $A_0 : \mathbb{Z} \to \mathbb{Z}$ is the identity function.*

Theorem 50. *If $a, b \in \mathbb{Z}$ then $A_a \circ A_b = A_{a+b}$.*

Exercise 23. Prove the above two theorems, and the following corollary.

Corollary 51. *If $a \in \mathbb{Z}$ then $A_a \circ A_{-a}$ and $A_{-a} \circ A_a$ are the identity function. Furthermore, A_a is bijective with inverse $(A_a)^{-1} = A_{-a}$.*

Exercise 24. Show that A_a and A_b commute.

We now discuss some applications of the translation functions. First recall the following definition from Chapter 2, which we extend to all the integers.

Definition 18. If $a, b \in \mathbb{Z}$ then

$$\{a, \ldots, b\} \stackrel{\text{def}}{=} \{x \in \mathbb{Z} \mid a \leq x \leq b\}.$$

Theorem 52. *Let $a, b, c \in \mathbb{Z}$. There is a bijection*

$$\{a, \ldots, b\} \to \{a + c, \ldots, b + c\}.$$

Proof. By Theorem 27, if $a \leq x \leq b$ then $a + c \leq x + c \leq b + c$. In other words, if $x \in \{a, \ldots, b\}$, then $A_c(x) \in \{a+c, \ldots, b+c\}$. So we can restrict the function A_c to produce a function $\{a, \ldots, b\} \to \{a + c, \ldots, b + c\}$. Similarly, the restriction of A_{-c} gives an inverse. Since these functions are inverses, they are bijective. □

Corollary 53. *Let $a, b \in \mathbb{Z}$. Then the set $\{a, \ldots, b\}$ has $(b-a)+1$ elements. So this set is finite.*

Proof. Apply the previous theorem with $c = -a + 1$. □

Note. This corollary tells us that if you work on a project from the 3rd of November to the 10th of November (inclusive), you have worked on it for $(10 - 3) + 1 = 8$ days. It is a bit counter-intuitive that you don't just subtract: $10 - 3 = 7$ is the wrong answer.

Theorem 54. *Let S be a non-empty subset of $\mathbb{Z}$ that is bounded from above, then S has a maximum.*

Proof. First consider the case where S intersects $\mathbb{N}$. Let $T = \mathbb{N} \cap S$. Now T is bounded by the same upper bound as S, and T is not empty. Thus, by a result of Chapter 2, T has a maximum M. Since $M \in \mathbb{N}$, we have $M \geq 0$. Since every negative element of S is less than 0, and since $M \geq 0$, we have that M is a maximum for all of S (transitivity, definition of maximum).

Now suppose that S does not intersect $\mathbb{N}$. Let $-n \in S$, and let S' be the image of S under the translation map A_n. Now S' contains $A_n(-n)$ which is just 0. So S' has a maximum M' since it intersects $\mathbb{N}$ (see the first case above). Let $M = M' - n$. Claim: M is a maximum of S. To see that M is in S first observe that $A_n(M) = M'$, and that $A_n(x) = M'$ for some $x \in S$ since $M' \in S'$ (def. of image). So $M = x$ by injectivity of A_n. Thus $M \in S$. Now we show that M is the maximum of S. Suppose that $y \in S$. Then we have $A_n(y) \leq M'$ (def. of max.), so $y + n \leq M'$. Hence $y \leq M' - n$. So $y \leq M$ as desired. □

Theorem 55. *Let S be a non-empty subset of $\mathbb{Z}$ that is bounded from below, then S has a minimum.*

Proof. (sketch) Let c be a lower bound of S. Let S' be the image of S under the translation map A_{-c}. Then $S' \subseteq \mathbb{N}$. Thus S' has a minimum n. This means that $n + c$ is the minimum of S. □

Exercise 25. Fill in the details of the proof of Theorem 55. Justify all steps.

Now we return to the study of translations. One important fact about translations is that the iteration of a translation is a translation. This follows from the fact that the composition of translations is a translation.

Lemma 56. *If $a \in \mathbb{Z}$ and $n \in \mathbb{N}$ then A_a^n is a translation.*

Proof. Fix $a \in \mathbb{Z}$, and let $S_a = \{x \in \mathbb{N} \mid A_a^x \text{ is a translation}\}$. Observe that $0 \in S_a$ since $A_a^0 = id = A_0$ (Theorem 49) and A_0 is a translation.

Suppose that $k \in S_a$. This means $A_a^k = A_b$ for some $b \in \mathbb{Z}$. Thus

$$A_a^{k+1} = A_a \circ A_a^k = A_a \circ A_b = A_{a+b}. \qquad \text{(Thm. 50)}$$

Since A_{a+b} is a translation, we have that $k + 1 \in S_a$.

By induction, $S_a = \mathbb{N}$. Since $n \in \mathbb{N}$, it is in S_a. The result follows. □

This lemma can be generalized:

Theorem 57. *If $a, b \in \mathbb{Z}$ then the iteration A_a^b of A_a is also a translation.*

Proof. If $b \geq 0$ then use Lemma 56. If $b < 0$ then $b = -n$ for some $n \in \mathbb{N}$. So, by Theorem 38 and Corollary 51,

$$A_a^b = A_a^{-n} = \left(A_a^{-1}\right)^n = (A_{-a})^n .$$

But $(A_{-a})^n$ is a translation by Lemma 56. □

The following will be useful in future sections:

Lemma 58. *Let $a, b \in \mathbb{Z}$. If $A_a = A_b$ then $a = b$.*

Exercise 26. Prove the above lemma. Hint, apply translations to 0.

Note. A fancy way to say some of the above is is that the set of translations forms an abelian group under composition, and the map $x \mapsto A_x$ is an isomorphism between the additive group $\mathbb{Z}$ and this group of translations.

Definition of multiplication in $\mathbb{Z}$

Above we proposed that multiplication be defined using iterated addition. In other words, if A_a is the addition by a map (also called the translation by a), then the proposal was that $a \cdot b$ be defined as $A_a^b(0)$. Since A_a is a bijection (Corollary 51), iteration A_a^b is defined even if b is negative.

Now we are officially ready to carry out this proposal. Properties of the map A_a will help us prove theorems about multiplication.

Definition 19 (Multiplication)**.** Let $a, b \in \mathbb{Z}$. Then

$$a \cdot b \stackrel{\text{def}}{=} A_a^b(0).$$

The bth iterate is defined since A_a is a bijection $\mathbb{Z} \to \mathbb{Z}$. By definition of iteration, A_a^b is a function $\mathbb{Z} \to \mathbb{Z}$. Thus $a \cdot b = A_a^b(0)$ is in $\mathbb{Z}$.

In particular, multiplication takes pairs of integers a, b to an integer. In other words, multiplication is a binary operation $\mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$.

Now we develop a few easy consequences of this definition. (None require induction in their proof).

Theorem 59. *If $a \in \mathbb{Z}$ then*

$$a \cdot 0 = 0 \qquad \textit{and} \qquad 0 \cdot a = 0.$$

Exercise 27. Prove Theorem 59. Hint: What is the zeroth iteration A_a^0? You will also need to use Lemma 48 and Theorem 49.

Lemma 60. *If $a \in \mathbb{Z}$ then*

$$a \cdot 1 = a.$$

Exercise 28. Prove Lemma 60.

Theorem 61. *If $a \in \mathbb{Z}$ then*

$$a \cdot (-1) = -a.$$

Exercise 29. Prove Theorem 61. Hint: see Corollary 51. Now apply the function $A_a^{-1} = A_{-a}$ to 0.

Lemma 62. *This definition extends the definition of multiplication given in Chapter 1. In other words, if $m, n \in \mathbb{N}$, then $m \cdot n$ is the same whether you use the definition in Chapter 1 or Definition 19.*

Proof. (Sketch) Let α_m be as in Chapter 1. Observe that $\alpha_m(x) = A_m(x)$ for all $x \in \mathbb{N}$. By induction, one can show that $\alpha_m^n(x) = A_m^n(x)$ for all $x, n \in \mathbb{N}$. In particular, $\alpha_m^n(0) = A_m^n(0)$. □

Remark 4 (Negative times a negative: one answer). Now we are in a position to explain a major puzzle of elementary mathematics: why is a negative times a negative equal to a positive? First we give an informal explanation: a formal proof will follow. For simplicity, this informal explanation will focus on the case $(-n)(-1)$ where $n \in \mathbb{N}$.

First, consider the two functions A_n and A_{-n}. The first of these translates all the integers n units in the positive direction, and the second translates all the integers n units in the negative direction. Clearly these two processes are inverse to each other (see Corollary 51).

Now when we multiply $-n$ by -1 we need to iterate the translation A_{-n} a total of -1 times. But we know that iterating by -1 is the same as taking the inverse. The inverse of A_{-n} is A_n. In short, $(-n)(-1)$ is n.

The same argument applies to multiplying $-n$ by $-m$. Multiplying by $-m$ involves inverting and iterating A_{-n}. After inverting we get A_n, which we then iterate a total of m times. This yields $m \cdot n$.

The above informal discussion is incorporated into the proof of the following.

Theorem 63. *If $m, n \in \mathbb{N}$ then $(-n)(-m) = n \cdot m$. Thus a negative integer times a negative integer is a positive integer.*

Proof.

$$\begin{aligned} A_{-n}^{-m} &= (A_{-n})^{-m} && \text{(rewriting)} \\ &= \left((A_{-n})^{-1}\right)^{m} && \text{(Thm. 38)} \\ &= (A_n)^m && \text{(Corollary 51)} \\ &= A_n^m && \text{(rewriting)} \end{aligned}$$

Thus, by definition of multiplication (Definition 19),

$$(-n)(-m) = A_{-n}^{-m}(0) = A_n^m(0) = n \cdot m.$$

□

Note. When we discuss rings, we will discuss other reasons why a negative times a negative is a positive. We will see that the axioms for a ring imply that $(-x)(-y) = x \cdot y$ for all elements x, y in the ring.

However, this answer might beg the question: why should $\mathbb{Z}$ be a ring? The nice thing about the above answer leading to Theorem 63 is that it is a natural consequence of thinking about multiplication as iterated addition.

Note (Discussion on Symmetry). When we study $\mathbb{Z}$ with no multiplication but only addition ($\mathbb{Z}$ as an abelian group) we get the sense that the positive and the negative integers are analogous. For example, the sum of two positive numbers is positive, and the sum of two negative numbers is negative.

The sum of a positive and a negative number depends on the sizes of the two numbers involved in a completely symmetric manner.

The terms "positive" and"negative" seemed like arbitrary labels in a similar manner to what happens with electrical charge. Here one type of charge is called "positive" and the other is "negative", where these terms are purely conventional: there is no inherent reason why a proton's charge should be called "positive" and an electron's charged called "negative".

However, when we move to multiplication ($\mathbb{Z}$ as a ring), we see a stark difference between positive and negative integers. For example, the product of positive integers is positive, but the product of two negative numbers is not negative. Where does this difference between positive and negative emerge? Our point of view is that the difference emerges with the idea of iteration f^a. When $a \in \mathbb{Z}$ is negative there is an element of inversion that is not present when $a \geq 0$. It is this that leads to the loss of symmetry between positive and negative numbers.[4]

Note. We defined $\mathbb{Z}$-multiplication in terms of iteration, but there are a few other common approaches. One way is to simply define it by cases:

$$m \cdot n \stackrel{\text{def}}{=} mn, \qquad m \cdot (-n) \stackrel{\text{def}}{=} -(mn)$$

$$(-m) \cdot n \stackrel{\text{def}}{=} -(mn), \qquad (-m) \cdot (-n) \stackrel{\text{def}}{=} mn$$

where multiplication on the right-hand side of each equation is as in Chapter 1, and where $m, n \in \mathbb{N}$. Proofs have to be done in cases as well. Under this definition, the answer to the question "why is a negative times an negative equal to a positive?" is that we defined it in this way!

A second alternative is to define multiplication by the formula

$$[m, n] \cdot [m', n'] = [mm' + nn', mn' + nm']$$

and show that the formula is well defined. Proving the laws of arithmetic is straightforward, but somewhat messy.

We believe our iteration approach is better motivated than these alternatives since it grows out of the definition of multiplication in Chapter 1. It also gives a reasonable answer to the question about a negative times a negative. The downside is it requires more study of negative iteration and translation maps. However, these topics are worth their own study, and so the downside is not too bad.

Laws of multiplication

We will prove the laws of multiplication. We begin with another law concerning translation.

[4] A fancy way of saying the above is that $a \mapsto -a$ is an isomorphism of $\mathbb{Z}$ as a group, but not as a ring.

Theorem 64. *If $a, b \in \mathbb{Z}$ then*

$$A_a^b = A_{a \cdot b}.$$

Proof. By Theorem 57, A_a^b is a translation. So we can write $A_a^b = A_c$ for some $c \in \mathbb{Z}$. What is c? By the definition of $a \cdot b$ (Definition 19) and the definition of translation (Definition 17),

$$a \cdot b = A_a^b(0) = A_c(0) = 0 + c = c.$$

So $c = ab$. Thus $A_a^b = A_c = A_{ab}$ as desired. □

Theorem 65 (Left distributive law)**.** *If $a, b, c \in \mathbb{Z}$ then*

$$a(b + c) = ab + ac.$$

Proof. Observe that

$$\begin{aligned} A_{a(b+c)} &= A_a^{b+c} && \text{(Thm. 64)} \\ &= A_a^b \circ A_a^c && \text{(Thm. 43)} \\ &= A_{ab} \circ A_{ac} && \text{(Thm. 64)} \\ &= A_{ab+ac} && \text{(Thm 50).} \end{aligned}$$

The result follows from Lemma 58. □

Theorem 66 (Right distributive law)**.** *If $a, b, c \in \mathbb{Z}$ then*

$$(a + b)c = ac + bc.$$

Proof. Recall that A_a and A_b commute (Exercise 24), so we can apply Theorem 46. By this and other results,

$$\begin{aligned} A_{(a+b)c} &= A_{a+b}^c && \text{(Thm. 64)} \\ &= (A_{a+b})^c && \text{(Rewrite)} \\ &= \big(A_a \circ A_b\big)^c && \text{(Thm. 50)} \\ &= A_a^c \circ A_b^c && \text{(Thm. 46)} \\ &= A_{ac} \circ A_{bc} && \text{(Thm. 64)} \\ &= A_{ac+bc} && \text{(Thm. 50).} \end{aligned}$$

The result follows from Lemma 58. □

Theorem 67. *If $a, b \in \mathbb{Z}$ then*

$$(-a)b = a(-b) = -(ab).$$

Proof. Observe

$$\begin{aligned} A_{(-a)b} &= A_{-a}^{b} && \text{(Thm. 64)} \\ &= (A_{-a})^{b} && \text{(Rewrite)} \\ &= \left(A_a^{-1}\right)^{b} && \text{(Cor. 51)} \\ &= A_a^{-b} && \text{(Cor. 47)} \\ &= A_{a(-b)} && \text{(Thm. 64).} \end{aligned}$$

So $(-a)b = a(-b)$ by Lemma 58.

Next observe,

$$\begin{aligned} A_{a(-b)} &= A_a^{-b} && \text{(Thm. 64)} \\ &= \left(A_a^{b}\right)^{-1} && \text{(Cor. 44)} \\ &= (A_{ab})^{-1} && \text{(Thm. 64)} \\ &= A_{-(ab)} && \text{(Cor. 51)} \end{aligned}$$

So $a(-b) = -(ab)$ by Lemma 58. □

Lemma 68. *If $a \in \mathbb{Z}$ and $n \in \mathbb{N}$ then $an = na$.*

Proof. (Using Corollary 23). If $a \in \mathbb{N}$, the result follows from the Commutative law of Chapter 1. If $a = -m$ with $m \in \mathbb{N}^+$, then

$$\begin{aligned} (-m)n &= -(mn) && \text{(Theorem 67)} \\ &= -(nm) && \text{(Chapter 1)} \\ &= n(-m) && \text{(Theorem 67)} \end{aligned}$$

□

Theorem 69 (Commutative law)**.** *If $a, b \in \mathbb{Z}$ then*

$$a \cdot b = b \cdot a.$$

Proof. (Using Corollary 23). If $b \in \mathbb{N}$, the result follows from Lemma 68. If $b = -n$ with $n \in \mathbb{N}^+$, then

$$\begin{aligned} a(-n) &= -(an) && \text{(Theorem 67)} \\ &= -(na) && \text{Lemma 68} \\ &= (-n)a && \text{(Theorem 67)} \end{aligned}$$

□

Here is an application of the commutative law.

Theorem 70. *If $a \in \mathbb{Z}$ then*

$$1 \cdot a = a \cdot 1 = a.$$

Proof. Combine Lemma 60 with the commutative law (Theorem 69). □

Before proving the associative law, we prove an important result concerning the interaction between iteration and multiplication. It generalizes a result from Chapter 3.

Lemma 71. *Let $a \in \mathbb{Z}$, $n \in \mathbb{N}$. If $f : S \to S$ is bijection then $(f^a)^n = f^{an}$.*

Proof. (Using Corollary 23). If $a \in \mathbb{N}$, the result follows from a law in Chapter 3. If $a = -m$ with $m \in \mathbb{N}^+$, then

$$\begin{aligned}
\left(f^{-m}\right)^n &= \left(\left(f^{-1}\right)^m\right)^n && \text{(Theorem 38)} \\
&= \left(f^{-1}\right)^{mn} && \text{(Law of Ch. 2)} \\
&= f^{-(mn)} && \text{(Theorem 38)} \\
&= f^{(-m)n} && \text{(Theorem 67)}
\end{aligned}$$

□

Theorem 72. *Let $f : S \to S$ be a bijection, and $a, b \in \mathbb{Z}$. Then*

$$(f^a)^b = f^{ab}.$$

Proof. (Using Corollary 23). If $b \in \mathbb{N}$, the result follows from a law in Chapter 3. If $b = -n$ with $n \in \mathbb{N}^+$, then

$$\begin{aligned}
(f^a)^{-n} &= ((f^a)^n)^{-1} && \text{(Theorem 38)} \\
&= (f^{an})^{-1} && \text{(Lemma 71)} \\
&= f^{-(an)} && \text{(Corollary 44)} \\
&= f^{a(-n)} && \text{(Theorem 67)}
\end{aligned}$$

□

Note. This gives an important application for multiplication in $\mathbb{Z}$: multiplication describes the resulting iteration associated with the iteration of an iteration. It gives evidence that our definition was "the right one".

Theorem 73 (Associative law)**.** *If $a, b, c \in \mathbb{Z}$ then*

$$a(bc) = (ab)c.$$

Proof. Observe

$$\begin{aligned} A_{a(bc)} &= A_a^{bc} && \text{(Thm. 64)} \\ &= (A_a)^{bc} && \text{(by rewriting)} \\ &= \Big((A_a)^b\Big)^c && \text{(Thm. 72)} \\ &= \Big(A_a^b\Big)^c && \text{(by rewriting)} \\ &= (A_{ab})^c && \text{(Thm. 64)} \\ &= A_{(ab)c} && \text{(Thm. 64).} \end{aligned}$$

The result follows from Lemma 58. □

4.10 The ring of integers $\mathbb{Z}$

There is an type of algebraic structure called a *ring*. A typical example is $\mathbb{Z}$.

Definition 20 (Ring)**.** A *ring* is a set R equipped with *two* binary operations $R \times R \to R$ satisfying the properties listed below. The binary operations are called *addition* and *multiplication*, and are typically written $+ : R \times R \to R$ and $\cdot : R \times R \to R$. Multiplication is also indicated by juxtaposition, and the usual conventions for parentheses are employed. A ring R must satisfy the following:

(i) The set R is an abelian group under addition. In other words,
  (i.1) addition is associative,
  (i.2) addition has an identity, typically written 0,
  (i.3) every element $x \in R$ has an additive inverse, typically written $-x$, and
  (i.4) addition is commutative.

(ii) Multiplication is associative: for all $x, y, z \in R$,

$$(xy)z = x(yz).$$

(iii) There is a multiplicative identity[5], typically written 1: for all $x \in R$,

$$x \cdot 1 = 1 \cdot x = x.$$

(iv) The distributive law holds: for all $x, y, z \in R$

$$x(y + z) = xy + xz,$$

$$(y + z)x = yx + zx.$$

[5]Some algebra textbooks do not require the multiplicative identity, but many do. Most rings that one considers, however, have a multiplicative identity.

Definition 21. Suppose that R is a ring such that

$$xy = yx$$

for all $x, y \in R$. Then we say that the *commutative law* holds for R, and we call R a *commutative ring*.

Note (Informal). The set of 2 by 2 matrices with entries in $\mathbb{R}$ forms a non-commutative ring. Thus not all rings are commutative.

Informal Exercise 30. Explain why $\mathbb{N}$ is not a ring.

Theorem 74. *The integers $\mathbb{Z}$ form a commutative ring.*

Proof. This follows from earlier results. □

Exercise 31. List the results needed to prove the above theorem. Hint: start with Theorem 15.

Many results that hold for $\mathbb{Z}$ actually extends to other rings as well. We give four examples.

Theorem 75. *If R is a ring, then*

$$x \cdot 0 = 0 \cdot x = 0$$

for all $x \in R$.

Proof. Recall the law $y + 0 = y$ for any additive group. So $0 + 0 = 0$. Thus

$$x \cdot 0 = x(0 + 0) = x \cdot 0 + x \cdot 0$$

by the distributive law. By adding the inverse $-(x \cdot 0)$ to both sides, we get

$$0 = x \cdot 0.$$

A similar argument shows $0 \cdot x = 0$. □

Theorem 76. *If $x, y \in R$ where R is a ring, then*

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

Proof. Observe that

$$xy + (-x)y = \big(x + (-x)\big)y = 0 \cdot y = 0$$

by the previous theorem. Thus xy and $(-x)y$ are additive inverses. So

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

A similar argument shows $x(-y) = -(xy)$. □

Theorem 77. *If $y \in R$ where R is a ring, then*

$$(-1)y = y(-1) = -y.$$

Proof. This follows from the previous theorem. For example, if $x = 1$ then $(-1)y = -(1y)$ by the previous theorem. □

Theorem 78. *If $x, y \in R$ where R is a ring, then*

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

Proof. By Theorem 76, used twice,

$$(-x)(-y) = -\big(x(-y)\big) = -\big(-(xy)\big).$$

However, $-(-z) = z$ for all z in an additive group (and R is an additive group under addition by the definition of ring). Thus

$$(-x)(-y) = -\big(-(xy)\big) = xy.$$

□

Note. The above theorem gives another explanation to our big question of why a negative times a negative yields a positive: it is just a special case of the above theorem. In other words, once you know that $\mathbb{Z}$ satisfies the properties of a ring, such as the distributive law, then from these properties or laws you can deduce that the product $(-x)(-y)$ is xy. In other words, the equality $(-x)(-y) = xy$ is just a consequence of basic laws of arithmetic, and is so general that it holds for any ring.

This gives a different explanation than that given in Section 4.9. There we appealed to special properties of $\mathbb{Z}$, but the present explanation works far more generally. The weakness of the present explanation is that it is only compelling once you know that the traditional laws of arithmetic, such that the distributive law, holds of $\mathbb{Z}$. The explanation of Section 4.9, on the other hand, gives an explanation directly from the definition of multiplication in $\mathbb{Z}$.

We end the section with the so-called "FOIL" rule.[6]

Theorem 79. *Let $a, b, c, d \in R$ where R is a ring then*

$$(a + b)(c + d) = ac + ad + bc + bd.$$

Note. Parentheses are not needed because $+$ is associative.

Exercise 32. Prove Theorem 79 using the distributive law.

[6] A favorite mnemonic in HS algebra: "first, outside, inside, last"

4.11 ℤ as an integral domain

We end with some important properties of ℤ.[7]

Theorem 80. *The product of two positive integers is positive, the product of two negative integers is positive, the product of a positive and a negative integer is negative.*

Exercise 33. Prove this theorem. Hint: the first was already proved in Chapter 1, so does not require an additional proof.

Corollary 81. *Suppose $x, y \in \mathbb{Z}$. If $x \neq 0$ and $y \neq 0$, then $xy \neq 0$.*

Corollary 82. *If $x, y \in \mathbb{Z}$ and if $xy = 0$, then $x = 0$ or $y = 0$.*

Exercise 34. Show how the above corollaries follow from Theorem 80.

We will see later in the course that there are rings with strange properties. There are rings where $0 = 1$. There are also rings where two nonzero elements have a zero product. Of course these do not happen with the integers. In fact, if these strange things do not happen, we call the ring an integral domain since it has many properties in common with the integers.

Definition 22. An *integral domain* is a commutative ring R with the additional properties that (i) $0 \neq 1$, and (ii) for $x, y \in R$,

$$xy = 0 \;\Rightarrow\; x = 0 \;\vee\; y = 0.$$

The name *integral* domain suggests it was inspired by the integers. Needless to say, there are other interesting integral domains that mathematicians study.

Corollary 83. *The ring ℤ is an integral domain.*

The cancellation law for ℕ can be found in Chapter 2. This law also holds in ℤ. In fact, it doesn't just hold in ℤ, but it holds in any integral domain. (Warning: it does not hold in every ring though).

Theorem 84 (Cancellation Law for Multiplication)**.** *Let R be ℤ or any integral domain. If $a, b, c \in R$ and if $c \neq 0$ then*

$$ac = bc \implies a = b.$$

Proof. Add $-ac$ to both sides of the equation $ac = bc$:

$$0 = ac + (-ac) = bc + (-ac) = bc + (-a)c = (b - a)c$$

(using Theorem 76, and the Distributive Law). Since R is an integral domain, $b - a = 0$ or $c = 0$. However, $c \neq 0$ by assumption. Thus $b + (-a) = 0$. By adding a to both sides, and using the identity, associative, and inverse laws (valid since R is a ring), we get $b = a$. □

[7] These are properties that will extend to ℚ and ℝ, but not to rings in general.

We end with standard laws concerning multiplication and inequalities.

Theorem 85. *Suppose that $x, y, z \in \mathbb{Z}$. Then*

$$x < y \ \wedge \ z > 0 \ \Rightarrow \ xz < yz,$$

$$x < y \ \wedge \ z < 0 \ \Rightarrow \ xz > yz,$$

$$x \leq y \ \wedge \ z \geq 0 \ \Rightarrow \ xz \leq yz,$$

and

$$x \leq y \ \wedge \ z \leq 0 \ \Rightarrow \ xz \geq yz.$$

Proof. Suppose $x < y$ and $z > 0$. Then $y - x$ is positive (Theorem 25). Thus $(y - x)z$ is positive by Theorem 80. But, by Theorem 66 and Theorem 67,

$$(y - x)z = \big(y + (-x)\big)z = yz + (-x)(z) = yz + \big(-(xz)\big).$$

Since $yz - xz$ is positive, we get $xz < yz$ by Theorem 25.

Suppose $x < y$ and $z < 0$. Then $y - x$ is positive (Theorem 25). Thus the integer $(y - x)z$ is negative by Theorem 80. As before,

$$(y - x)z = yz + \big(-(xz)\big).$$

So $yz + \big(-(xz)\big) < 0$. Add xz to both sides (Theorem 27). The result follows.

Suppose $x \leq y$ and $z \geq 0$. If both inequalities are strict, we have proved the result already. If $x = y$ the result $xz = yz$ follows by multiplying both sides of the equation by z. If $z = 0$, then $xz = yz = 0$ (Theorem 59). The result follows.

Suppose $x \leq y$ and $z \leq 0$. If both inequalities are strict, the result follows what we have done. If $x = y$ or $z = 0$ then $xz = yz$ as before. □

Chapter 5

Exploring $\mathbb{Z}$

5.1 Introduction

In this chapter we continue the study of the ring $\mathbb{Z}$. We begin with absolute values. The absolute value function $\mathbb{Z} \to \mathbb{N}$ is the identity when restricted to $\mathbb{N}$. The fundamental law $|ab| = |a| \cdot |b|$ shows that this function is compatible with products. Equally important is the fact that it is not always compatible with sums.

Next we consider induction. In previous chapters we used only a limited form of induction where the base case is zero and where we have to prove a statement n when assuming it for $n-1$. In practice we sometimes want the base case to start at another integer (positive or negative). Also, sometimes we want to be able to prove the case n not from the assumption that it holds for $n-1$, but under the stronger assumption that it holds for all suitable integers less than n. These variants are developed in this chapter.[1] Unlike the earlier principle of induction, these new forms of induction will not be the basis of new axioms, but will be proved to be valid from previous results.

A major theme of this chapter is divisibility. We consider division b/a, but at first only in the case where $a \mid b$ (and where $a \neq 0$). This is followed by a more general conception of division captured by the important Quotient-Remainder Theorem, which introduces the basic concepts of quotient and remainder. We use the Quotient-Remainder Theorem to prove a few things about least common multiples (LCMs). We also briefly discuss the analogous idea of greatest common divisors (GCDs). We then consider

[1] We won't consider all forms of induction. For example, *transfinite induction* will not be covered in these notes. This is a type of induction concerning collections of transfinite ordinals instead of just $\mathbb{N}$ or well-ordered subsets of $\mathbb{Z}$.

prime numbers and relatively prime pairs, and prove a few basic results including the principle, valid for prime p, that

$$p \mid ab \implies p \mid a \text{ or } p \mid b.$$

From these topics, there are three other topics that naturally follow (i) the fact that every $n > 1$ is the product of primes (part of the Fundamental Theorem of Arithmetic), (ii) the fact that the set of prime numbers is infinite, and (iii) the fact that, for any fixed base $B > 1$, every integer has a unique base B representation. The only difficulty with these topics is that they involve finite products $\prod a_i$ and sums $\sum a_i$. We have yet to consider such products and sums and justify their basic laws. These concepts also require the concept of a finite sequence $a_1, \ldots, a_k$.

A large part of the chapter will be used to justify the basic laws for finite sums and products, but before this is done there will be a section where we discuss informal proofs of the three facts (i), (ii), (iii) mentioned above (concerning primes and base B representations), and where we discuss what properties of products $\prod a_i$ and sums $\sum a_i$ are required for their proofs. In subsequent sections the necessary theory of finite sequences, finite sums, and finite products is developed. The official proofs of (i), (ii), and (iii) are given in later sections. Optional sections follow which discuss $\prod a_i$ and $\sum a_i$ further.

5.2 Absolute values in $\mathbb{Z}$

Definition 1 (Absolute Value)**.** The *absolute value* $|a|$ of $a \in \mathbb{Z}$ is defined as follows.

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0 \end{cases}$$

The following is an easy consequence of the definition and the fact, from Chapter 4, that $a < 0$ if and only if $-a > 0$.

Theorem 1. *If $a \in \mathbb{Z}$ then $|a| \geq 0$. Furthermore, for $n \in \mathbb{N}$,*

$$|a| = n \iff a = n \text{ or } a = -n.$$

In particular (since $-0 = 0$), $|a| = 0$ if and only if $a = 0$.

Remark 1. Since $|a| \geq 0$, the rule $x \mapsto |x|$ defines a function $\mathbb{Z} \to \mathbb{N}$. If $a \in \mathbb{N}$ then $|a| = a$ so the restriction from $\mathbb{Z}$ to $\mathbb{N}$ of $x \mapsto |x|$ is the identity function.

Exercise 1. Use the last statement of Theorem 1 to show that $|a| \geq 1$ if and only if $a \neq 0$.

Next we establish that absolute value is compatible with multiplication.

Theorem 2. *If $a, b \in \mathbb{Z}$ then*

$$|ab| = |a| \cdot |b|.$$

Exercise 2. Prove this theorem. Hint: Divide the proof into four cases. In the cases where $a < 0$, write $a = -m$ for $m \in \mathbb{N}$. In the cases where $b < 0$, write $b = -n$ for $n \in \mathbb{N}$.

Informal Exercise 3. Absolute value is less compatible with addition. Give examples where $|a + b| = |a| + |b|$ holds, and give examples where it fails.

Theorem 3. *Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. Then $|a| \leq n$ if and only if $-n \leq a \leq n$. Similarly, $|a| < n$ if and only if $-n < a < n$.*

Proof. Suppose $|a| \leq n$. Let $m = |a|$. So $0 \leq m \leq n$. By Theorem 1, either $a = m$ or $a = -m$. In the first case $0 \leq a \leq n$. In the second case $0 \leq -a$ and $-a \leq n$, which implies $0 \geq a$ and $a \geq -n$ by results of Chapter 4. In either case $-n \leq a \leq n$.

Now suppose $-n \leq a \leq n$. If $a \geq 0$ then $|a| = a$ so $|a| \leq n$. If $a < 0$ then observe that $-n \leq a$ implies $-a \leq n$ by a result of Chapter 4. Thus $|a| \leq n$ in this case as well.

The proof for $<$ is similar. □

Theorem 4. *Let $x, y, n \in \mathbb{Z}$. If $0 \leq x < n$ and $0 \leq y < n$ then $|x - y| < n$.*

Proof. We have $-y \leq 0$ (see Chapter 4), so $x + (-y) \leq x + 0 < n + 0$. Thus $x - y < n$ by mixed transitivity.

We also have $-n < -y$ (Chapter 4). So $0 + (-n) < 0 + (-y) \leq x + (-y)$. Thus $-n < x - y$ by mixed transitivity.

By Theorem 3, $|x - y| < n$. □

5.3 Induction and recursion variants

In Chapter 1, the axiom of induction was introduced. This axiom allows us to prove a statement for all natural numbers provided we know the statement is true for 0, and provided we have an argument that its truth for n implies its truth for $n + 1$. Obviously this is not the only valid form of induction. For example, one can choose to start at integers other than 0, and adjust the conclusion accordingly. There is also a variant called "strong induction" that is easier to use when the n and $n + 1$ cases are not clearly connected. Here, in the inductive step, you get to assume that the statement holds of *all* integers from the base to $n - 1$, and then you try to prove that it holds for n. This allows you to use a stronger hypothesis than regular induction, which in turn can make it easier to prove desired results.

Since these variant forms of induction were not included in the axioms, we need to prove they are valid before we can use them. This is the purpose of this section. We also consider a variant of the definition of recursion, a version especially useful for defining summations and general finite products.

In Chapter 4 translation functions were used to show the following: *Let S be a nonempty subset of $\mathbb{Z}$. If S has a lower bound then it has a minimum, and if S has an upper bound than it has a maximum.* We use this result to to justify the variant forms of induction.[2]

Theorem 5 (Base b induction)**.** *Let b be an integer, and S a subset of $\mathbb{Z}$ such that (i) $b \in S$ and (ii) $n \in S \Rightarrow n+1 \in S$ for arbitrary integers $n \geq b$. Then*

$$\{x \in \mathbb{Z} \mid x \geq b\} \subseteq S.$$

Proof. Consider the set E of exceptions. In other words, let E be the set of all integers $x \geq b$ not in S. We wish to show that E is empty. So suppose that E is not empty.

Observe that E has lower bound b, but $b \notin E$ (by assumption (i)). So, by the above mentioned property (from Chapter 4), E must have a minimum $m > b$. Let $a = m - 1$. Since $a + 1 = m$, we know $b < a + 1$. Note $b \leq a$, otherwise we would have $a < b < a + 1$, but we know there are no integers strictly between a and $a + 1$. Also note that a cannot be an exception since m is the minimum of E and a is less than m. In other words, $a \in S$. By assumption, $a \in S \Rightarrow a + 1 \in S$. Thus $m = a + 1$ is in S, a contradiction. □

Here is a finite version of the above:

Theorem 6. *Let b and c be integers with $b \leq c$, and let S be a subset of integers such that (i) $b \in S$ and (ii) $n \in S \Rightarrow n + 1 \in S$ for all $b \leq n < c$. Then*

$$\{b, \ldots, c\} \subseteq S.$$

Exercise 4. Modify the proof of Theorem 5 to prove the above. Keep in mind that you are not doing a proof by induction; you are proving that this method of induction is valid.

The next theorem shows that the more traditional form of induction, without using an induction set, is also a valid proof form. This means that any induction proofs following this theorem can be done in the traditional way.

Theorem 7 (Traditional induction)**.** *Let b be an integer, and for any $x \in \mathbb{Z}$, let $P(x)$ be a proposition. Assume (i) $P(b)$ is true (ii) $P(n)$ is true implies*

[2] Warning: "base" here is used in a different sense then at the end of this chapter where we discuss base B representations of an integer.

$P(n+1)$ is true for arbitrary integers $n \geq b$. Then $P(x)$ is true for all integers $x \geq b$.

Proof. Define $S = \{x \in \mathbb{Z} \mid P(x)\}$.

First observe that by assumption (i) and definition of S, $b \in S$. Next assume that for some $n \geq b$, $n \in S$. Thus by definition of the set S, $P(n)$ is true. So by assumption (ii), $P(n+1)$ is also true, hence $n+1 \in S$. By Theorem 5,

$$\{x \in \mathbb{Z} \mid x \geq b\} \subseteq S.$$

Therefore for all $x \geq b$, $P(x)$ is true. □

Finally, one sometimes needs the following form of induction:

Theorem 8 (Strong induction)**.** *Let $b \in \mathbb{Z}$ and $S \subseteq \mathbb{Z}$. Suppose $b \in S$ and*

$$\{b, \ldots, n-1\} \subseteq S \implies n \in S \qquad \forall\, n > b.$$

Then

$$\{x \in \mathbb{Z} \mid x \geq b\} \subseteq S.$$

Proof. Let E be the set of all integers $x \geq b$ not in S. We wish to show that E is empty. So suppose that E is not empty.

Observe that E has lower bound b. So E must have a minimum m. Since $b \notin E$ we have $m > b$. Since m is the minimum, $\{b, \ldots, m-1\} \subseteq S$. By assumption, however, $\{b, \ldots, m-1\} \subseteq S \Rightarrow m \in S$. This means $m \in S$, a contradiction. □

Remark 2. Recall that if $c < b$ we defined $\{b, \ldots, c\}$ to be the empty set. With that in mind, observe that the hypothesis in the above theorem can be restated as

$$\{b, \ldots, n-1\} \subseteq S \implies n \in S \qquad \forall\, n \geq b$$

with $n \geq b$ replacing $n > b$. If $n = b$ this becomes $\varnothing \subseteq S \implies b \in S$. Since $\varnothing \subseteq S$ is always true, this is logically equivalent to $b \in S$. So there is no reason to explicitly require $b \in S$ if we require the implication for all $n \geq b$ and not just $n > b$.

Remark 3. Note that using a proof similar to our proof of Theorem 7, we could also show that finite base b induction (Theorem 6) and strong induction (Theorem 8) could also be done in the traditional way, without using sets.

In Chapter 2 we introduced the idea of *recursive definitions* which is very similar to induction. (We use *induction* in the context of a proof, and *recursion* in the context of a definition). One common version of recursion is captured by the following theorem whose proof was given in the last (optional) section of Chapter 2.

Theorem 9. *Let S be a set, c an element of S, and $g : \mathbb{N} \times S \to S$ a function. Then there is a unique function $f : \mathbb{N} \to S$ satisfying the two equations*

$$f(0) = c$$
$$f(n+1) = g\big(n, f(n)\big)$$

for all $n \in \mathbb{N}$.

This theorem allows us to define a function just by giving a base case equation, and a recursive equation defining $f(n+1)$ in terms of n and $f(n)$. The function g gives the dependency of $f(n+1)$ on n and $f(n)$. For example, the factorial function $f(n) = n!$ is the unique function satisfying the two equations

$$f(0) = 1$$
$$f(n+1) = (n+1)f(n)$$

In this case $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is the function $g(n, m) = (n+1)m$. The recursive definition leads to the two laws

$$0! = 1$$
$$(n+1)! = (n+1)n!$$

which we can then use to prove theorems about the factorial function.

We now give a few variants of this recursive definition theorem.

Theorem 10 (Base b recursion)**.** *Let b be an integer, and consider*

$$I \;=\; \{b, b+1, \ldots\} \subseteq \mathbb{Z}.$$

Let S be a set, c an element of S, and $g : I \times S \to S$ a function. Then there is a unique function $f : I \to S$ satisfying the two equations

$$f(b) = c$$
$$f(n+1) = g\big(n, f(n)\big)$$

where the second equation holds for all $n \in I$.

Proof. To show uniqueness, suppose f_1 and f_2 both satisfy the desired conditions. Let $S \subseteq I$ the subset of $n \in I$ such $f_1(n) = f_2(n)$. Use an induction argument (via Theorem 5) to show that $S = I$. Thus $f_1 = f_2$ as functions.

To show existence, shift from I to $\mathbb{N}$ and use Theorem 9. More precisely, let $\tilde{g}\colon \mathbb{N} \times S \to S$ be defined by the equation

$$\tilde{g}(k, m) = g(b + k, m).$$

Now use Theorem 9 to define a function $\tilde{f}\colon \mathbb{N} \to S$ by the equations

$$\begin{aligned} \tilde{f}(0) &= c \\ \tilde{f}(n+1) &= \tilde{g}\left(n, \tilde{f}(n)\right). \end{aligned}$$

Finally define $f\colon I \to S$ by

$$f(n) = \tilde{f}(n-b).$$

Observe that

$$f(b) = \tilde{f}(b-b) = \tilde{f}(0) = c$$

and

$$\begin{aligned} f(n+1) &= \tilde{f}(n+1-b) \\ &= \tilde{g}\big(n-b, \tilde{f}(n-b)\big) \\ &= \tilde{g}\big(n-b, f(n)\big) \\ &= g\big(n, f(n)\big). \end{aligned}$$

□

A finite version of the above can also be given:

Theorem 11. *Let b and d be integers such that $b \leq d$, and consider*

$$I \;=\; \{b, \ldots, d\} \subseteq \mathbb{Z}.$$

Let S be a set, c an element of S, and $g : (I - \{d\}) \times S \to S$ a function. Then there is a unique function $f : I \to S$ satisfying the two equations

$$\begin{aligned} f(b) &= c \\ f(n+1) &= g\big(n, f(n)\big) \end{aligned}$$

where the second equation holds for all $b \leq n < d$.

Proof. To show uniqueness, suppose f_1 and f_2 both satisfy the desired conditions. Let $S \subseteq I$ the subset of $n \in I$ such $f_1(n) = f_2(n)$. Use an induction argument (via Theorem 6) to show that $S = I$. Thus $f_1 = f_2$ as functions.

To show existence, extend the function g from the finite set I to the infinite set $\tilde{I} = \{b, b+1, \ldots\}$ and use Theorem 10. Finally, restrict the result $\tilde{f}$ of Theorem 10 to the set I.

More precisely, let $\tilde{g}\colon \tilde{I} \times S \to S$ be defined by the equation

$$\tilde{g}(n,m) = \begin{cases} g(n,m) & \text{if } n < d \\ m & \text{if } n \geq d. \end{cases}$$

(The definition for the case $n \geq d$ doesn't matter for the proof, but the value m is a convenient choice.) Now use Theorem 10 to define a function $\tilde{f}\colon \tilde{I} \to S$ by the equations

$$\tilde{f}(b) = c$$
$$\tilde{f}(n+1) = \tilde{g}\left(n, \tilde{f}(n)\right).$$

Finally define $f\colon I \to S$ to be the restriction of $\tilde{f}$ to the subset I. Observe

$$f(b) = \tilde{f}(b) = c$$

and if $b \leq n < d$ then $n, n+1 \in I$ and

$$\begin{aligned} f(n+1) &= \tilde{f}(n+1) \\ &= \tilde{g}\big(n, \tilde{f}(n)\big) \\ &= g\big(n, f(n)\big). \end{aligned}$$

□

5.4 Divisibility and division

In previous chapters we have discussed addition, subtraction, multiplication, and even exponentiation. We have covered almost all of basic arithmetic, except we have avoided the delicate topic of division. We start with divisibility

Definition 2. [Divisibility and divisor] Let $d \in \mathbb{Z}$. An integer of the form cd with $c \in \mathbb{Z}$, is called a *multiple* of d. If $b = cd$ is a multiple of d, then we also say that d *divides* b. In this case we call d a *divisor* of b, and we write $d \mid b$.

In other words, given $b, d \in \mathbb{Z}$, the statement $d \mid b$ holds if and only if there exists a $c \in \mathbb{Z}$ such that $b = cd$.

Warning. The term *divides* refers to a relation: it is either true or false when applied to two integers. It does not produce a number.

The relation $\mid$ is written with a vertical stroke, and should not be confused with / (Definition 3) which produces a number. There is a relationship between these two ideas. In fact, $a|b$ if and only if b/a is an integer. Note that the order is reversed! (Here we assume $a \neq 0$).

Exercise 5. Prove the following simple consequences of the definition.

Theorem 12. *Suppose $a, b \in \mathbb{Z}$.*
(i) $a \mid a$.
(ii) $1 \mid a$.
(iii) $a \mid ab$.
(iv) $a \mid 0$.

Exercise 6. So $a \mid 0$ for all $a \in \mathbb{Z}$. Show, however, that $0 \nmid a$ for all $a \neq 0$.

Exercise 7. Show that the divisibility relation is reflexive but not symmetric.

Exercise 8. Prove the following theorem.

Theorem 13. *The divisibility relation is transitive: for all $a, b, c \in \mathbb{Z}$, if $a \mid b$ and $b \mid c$ then $a \mid c$.*

Exercise 9. Prove the following.

Theorem 14. *Suppose $a, b, d \in \mathbb{Z}$ and $a \neq 0$. Then $d \mid b$ if and only if $ad \mid ab$.*

Exercise 10. Prove the following theorem and its corollary.

Theorem 15. *Suppose $a, b, c, u, v \in \mathbb{Z}$. If $c \mid a$ and $c \mid b$ then $c \mid ua + vb$.*

Corollary 16. *Suppose that $c \mid a$ and $c \mid b$ where $a, b, c \in \mathbb{Z}$. Then c divides the sum and difference of a and b.*

Theorem 17. *Let $d, a \in \mathbb{Z}$. If $d \mid a$ where $a \neq 0$, then $|d| \leq |a|$.*

Proof. By definition, $a = cd$ for some $c \in \mathbb{Z}$. Claim: $c \neq 0$. To see this, observe that if $c = 0$ then $a = 0$, a contradiction.

By Exercise 1, $|c| \geq 1$. Now multiply both sides of the inequality $1 \leq |c|$ by $|d|$. By Theorem 2, we get

$$|d| \leq |c||d| = |cd| = |a|.$$

□

Exercise 11. Show that the only divisors of 1 are ± 1. Show that the only divisors of 2 are ± 1 and ± 2. Hint: see Theorem 1.

Exercise 12. Show that the set of divisors of a non-zero integer a is finite. Hint: apply Theorem 3 to $n = |a|$. Is $\{-n, \ldots, n\}$ finite?

Exercise 13. Prove the following. (Treat zero cases separately).

Corollary 18. *If $a \mid b$ and $b \mid a$ then $|a| = |b|$.*

Remark 4. The above results hint at the fact that the sign of the integers does not affect divisibility. The following lemma and corollaries illustrates this. Thus it is traditional to focus on the positive divisors only.[3]

Lemma 19. *Let $a, b \in \mathbb{Z}$. If $a \mid b$ then $-a \mid b$, $a \mid -b$, and $-a \mid -b$.*

[3]The proofs of the next few results are left to you the reader. In general, some of the easier results will not be proved. You the reader, should supply the proofs. It is fine to do this in your head if the proof is simple enough.

Corollary 20. *Let $a, b \in \mathbb{Z}$. Then*

$$a \mid b \iff |a| \mid b \iff a \mid |b| \iff |a| \mid |b|.$$

In particular $a \mid |a|$ and $|a| \mid a$ (since $a \mid a$).

Corollary 21. *Let $b \in \mathbb{Z}$. Then b and $-b$ have the same divisors.*

Corollary 22. *Let $b \in \mathbb{Z}$. Then $d \mid b$ if and only if $-d \mid b$.*

We now define division, but only in the case where $a \mid b$. The general case must wait until we introduce the rational numbers $\mathbb{Q}$.

Definition 3 (Division)**.** Suppose $a, b \in \mathbb{Z}$ are such that $a \mid b$ and $a \neq 0$. Then b/a is defined to be the integer $c \in \mathbb{Z}$ such that $b = ca$. (This integer exists since $a \mid b$. You will show it is unique.)

Exercise 14. For the above definition to be valid, the element c must be unique. Show the uniqueness.

Remark 5. Division is analogous to subtraction. Subtraction, which is defined in terms of addition, is only partially defined in $\mathbb{N}$, but becomes totally defined in the ring $\mathbb{Z}$. Similarly division, which is defined in terms of multiplication, is only partially defined in $\mathbb{Z}$, but becomes almost totally defined in the field $\mathbb{Q}$. Division is never totally defined: you cannot divide by zero.

It is sometimes handy to restate a definition as a theorem. Obviously for such theorems the proof is a simple appeal to the definition, and does not usually need to be written out. We now restate the definition of division:

Theorem 23 (Basic law of division)**.** *Suppose $a, b, c \in \mathbb{Z}$ are such that $a \mid b$ and $a \neq 0$. Then $b/a = c$ if and only if $b = ca$.*

Exercise 15. Prove the following four theorems with the basic law of division.

Theorem 24. *Let $a \in \mathbb{Z}$ be non-zero. Then $a/a = 1$ and $0/a = 0$.*

Theorem 25. *Suppose $a, b \in \mathbb{Z}$ are such that $a \neq 0$ and $a \mid b$. Then*

$$b = (b/a) \cdot a.$$

Theorem 26. *Suppose $a, b, c \in \mathbb{Z}$ are non-zero integers such that a and b divide c. Then $c/a = b$ if and only if $c/b = a$.*

Theorem 27. *Suppose $a, b \in \mathbb{Z}$ where $b \neq 0$ Then $ab/b = a$.*

5.5 The quotient-remainder theorem

Suppose $a \neq 0$ and a possibly does not divide b. Then we do not consider a simple quotient b/a. Instead we get a both quotient and a *remainder*. If $a \mid b$ then the remainder is 0. These ideas are based on the following:

Theorem 28 (Quotient-Remainder theorem)**.** *Let $a, b \in \mathbb{Z}$ where $a \neq 0$. Then there are unique $q, r \in \mathbb{Z}$ such that*

$$b = qa + r \qquad \text{and} \qquad 0 \leq r < |a|.$$

Definition 4 (Quotient and remainder)**.** Let $b, a \in \mathbb{Z}$ where $a \neq 0$. The integers q and r above are called the *quotient* and *remainder* of dividing b by a. Also, $\text{Rem}(b, a)$ is defined to be the remainder when dividing b by a.

The strategy of the proof is to define q to be such that qa is the largest multiple of a that is less than b. We need a lemma that shows that there is a largest multiple.

Lemma 29. *Suppose $a, b \in \mathbb{Z}$ are such that $a \neq 0$. Then there is a largest multiple of a that is less than or equal to b.*

Proof. Let S be the set of multiples of a that are less than or equal to b. By a result of Chapter 4, if S is non-empty and has an upper bound, then it has a maximum. Obviously b is an upper bound. So we only need to show that S is non-empty.

If $b \geq 0$ then $0 \in S$, and we are done. So assume that $b < 0$. By assumption $a \neq 0$, and we have $|a| \geq 1$ (Exercise 1). Multiplying both sides of $|a| \geq 1$ by b gives $|a| \cdot b \leq b$. Since a divides $|a|$ we have a divides $|a| \cdot b$ (transitivity). In particular, $|a| \cdot b \in S$. □

We now prove the existence of q and r in Theorem 28.

Proof of existence. Let qa be the largest multiple of a such that $qa \leq b$. This exists by the previous lemma. Let $r \stackrel{\text{def}}{=} b + (-qa)$. By adding qa to both sides we get $qa + r = b$ as desired.

We still need to show that $0 \leq r < |a|$. Since $qa \leq b$, we have

$$qa + (-qa) \leq b + (-qa).$$

In other words, $0 \leq r$. So we must only show $r < |a|$.

Suppose otherwise that $r \geq |a|$. Then $r + (-|a|) \geq |a| + (-|a|)$. In other words $r - |a| \geq 0$. So

$$b = qa + r = qa + |a| + (r - |a|) \geq qa + |a|.$$

However, $qa + |a| > qa$, and a divides $qa + |a|$ by Theorem 15. This contradicts the choice of qa as the maximum multiple of a less than or equal to b. □

Proof of uniqueness. Suppose that $b = qa + r = q'a + r'$ where $0 \leq r < |a|$ and $0 \leq r' < |a|$. Then, using laws from Chapter 4,

$$r - r' = (b + (-qa)) - (b + (-q'a)) = (q' - q)a.$$

By Theorem 4, $|r - r'| < |a|$. By Theorem 2,

$$|r - r'| = |q' - q| \cdot |a|.$$

So $|q' - q| \cdot |a| < |a|$. This means $|q' - q| < 1$ (Chapter 2). Since $|q' - q|$ is an integer, we have $|q' - q| = 0$. So $q' - q = 0$ (Theorem 1). Thus $q' = q$. Also, since $r - r' = (q' - q)a$, we have $r - r' = 0$. So $r' = r$. □

Exercise 16. Prove the following:

Theorem 30. *Let $a, b \in \mathbb{Z}$ where $a \neq 0$. Then*

$$\text{Rem}(b, a) = 0 \iff a \mid b.$$

If $\text{Rem}(b, a) = 0$ then b/a is the quotient (as defined in Definition 4).

Informal Exercise 17. Find the quotient and remainder of dividing 20 by 9. Find the quotient and remainder of dividing -30 by 7.

Informal Exercise 18. What is $\text{Rem}(109, 7)$, $\text{Rem}(-109, 7)$, $\text{Rem}(-70, 7)$?

5.6 GCDs and LCMs

Definition 5. Suppose that $a, b \in \mathbb{Z}$. Then a *common divisor* is an integer d such that $d \mid a$ and $d \mid b$. A *common multiple* is an integer m that is both a multiple of a and a multiple of b. In other words, $a \mid m$ and $b \mid m$.

Informal Exercise 19. Find all the common divisors of -8 and 12 (even the negative divisors). Find four common multiples of -8 and 12.

Theorem 31. *Let a, b be integers, not both zero. Then a and b have a greatest common divisor. This divisor is also called the GCD of a and b, and is written $\gcd(a, b)$.*

Proof. Let S be the set of common divisors. We know that $1 \in S$, so S is not empty. Without loss of generality, suppose $a \neq 0$. By Theorem 17, all elements $x \in S$ satisfy $x \leq |a|$. Thus S has an upper bound. By a result of Chapter 4, S has a maximum. □

Informal Exercise 20. Find two positive integers a and b whose GCD is 1. Find two distinct positive integers a and b, both greater than 1, whose GCD is just a.

Theorem 32. *Let a, b be non-zero integers. Then a and b have a least common positive multiple. This multiple is usually called the least common multiple, or the LCM, of a and b.*

Proof. Let S be the set of positive common multiples. Since $|ab| \in S$, S is not empty. By the well-ordering property of $\mathbb{N}$, the set S has a minimum. □

Informal Exercise 21. Find two positive integers a and b whose LCM is ab. Find two distinct positive integers a and b whose LCM is not ab.

Exercise 22. Prove the following.

Theorem 33 (Linear Combination)**.** *Let $a, b, u, v \in \mathbb{Z}$. Every common divisor of a and b divides $ua + vb$. In particular,* $\gcd(a, b) \mid ua + vb$.

The following is sometimes handy:

Lemma 34. *Let b, a be integers where $a \neq 0$. Then any common divisor of b and a also divides* $\mathrm{Rem}(b, a)$. *In particular,* $\gcd(b, a) \mid \mathrm{Rem}(b, a)$.

Proof. We have that $b = qa+r$ where q is the quotient and r is the remainder. Thus $\mathrm{Rem}(b, a) = (1)b + (-q)a$. By Theorem 33, any common divisor of b and a divides $\mathrm{Rem}(b, a)$. □

It is easy to see that any multiple of the LCM is a common multiple, the following gives a converse.

Theorem 35. *Let a, b be non-zero integers, and let m be the LCM. Then any common multiple of a and b is a multiple of m.*

Proof. Let c be a common multiple of a and b. Observe that a is a common divisor of c and m. Thus a is a divisor of $\mathrm{Rem}(c, m)$ by Lemma 34. Likewise, b is a divisor of $\mathrm{Rem}(c, m)$. Thus $\mathrm{Rem}(c, m)$ is a common multiple of a and b. But $\mathrm{Rem}(c, m) < m$ (Quotient-Remainder theorem), and m is the least common positive multiple. Thus $\mathrm{Rem}(c, m) = 0$ which implies that c is a multiple of m. □

5.7 Prime numbers and relatively prime pairs

Definition 6 (Prime number)**.** A *prime number* (or a *prime*) is an integer p such that (i) $p > 1$, and (ii) the only positive divisors of p are 1 and p.

Exercise 23. Show that 2 and 3 are prime, but that 4 is not. You may use the facts $\{1, \ldots, 2\} = \{1, 2\}$ and $\{1, \ldots, 3\} = \{1, 2, 3\}$. You may also use the facts $3 = 2 + 1$, $4 = 2 \cdot 2$, and $1 < 2 < 4$. (These facts are all easily provable using the results of Chapters 1 and 2). Hint: use Theorem 17.

The following is a great illustration of the usefulness of strong induction. Regular induction is not as easy to use here since knowing that n has a prime divisor does not help us to show that $n+1$ has a prime divisor.

Theorem 36. *Let $n \geq 2$ be an integer. Then n has at least one prime divisor.*

Proof. Let S be the set of all integers $x \geq 2$ such that x has a prime divisor. Observe that S contains all prime numbers since $p \mid p$ for all such p. In particular $2 \in S$. Now suppose that $n > 2$ and that we have established that $\{2, \ldots, n-1\} \subseteq S$. If n is prime we have $n \in S$, so consider the case where n is not prime. Then n has a positive divisor d where $d \neq 1$ and $d \neq n$. By Theorem 17 this implies that $1 < d < n$. So $d \in S$. Thus d has a prime divisor p. Since $p \mid d$ and $d \mid n$, we have $p \mid n$ by transitivity. So $n \in S$.

By the principle of strong induction (Theorem 8), all integers $n \geq 2$ are in S. So any such n has a prime divisor. □

Definition 7 (Relatively prime)**.** Let $a, b \in \mathbb{Z}$. We say that a and b are *relatively prime* if 1 is the only positive common divisor of a and b. In other words, a and b are relatively prime if and only if $\gcd(a, b) = 1$.

Remark 6. Observe that being prime is a property of one integer, while being relatively prime is a property of a pair of integers.

Theorem 37. *If $p, q \in \mathbb{N}$ are distinct prime numbers, then p and q are relatively prime. More generally, if p is a prime and $p \nmid a$ where $a \in \mathbb{Z}$ then p and a are relatively prime.*

Exercise 24. Prove the above theorem.

Exercise 25. Show that 3 and 4 are relatively prime. Hint: $4 = 3 + 1$, so what is $\text{Rem}(4, 3)$?

Theorem 38. *Suppose that $a, b \in \mathbb{Z}$ are non-zero and relatively prime. Then the LCM of a and b is $|ab|$.*

Proof. Let m be the LCM of a and b. Since $|ab|$ is a common multiple of a and b, we have $mq = |ab|$ for some $q \in \mathbb{Z}$ (Theorem 35). We will show that $m = |ab|$ by showing that $q = 1$.

Since a and b are non-zero, the same is true of ab. Thus $|ab|$ is positive. Also m is positive by definition of LCM. Thus q must be positive (the other cases lead to contradictions). Also $mq' = ab$ where $q' = q$ or $q' = -q$.

Claim: $q \mid a$. To see this, write $m = kb$ (m is a multiple of b). So

$$ab = q'm = q'(kb) = (q'k)b.$$

By the cancellation law for multiplication (Chapter 4), $a = q'k$. Thus $q' \mid a$. Hence $q \mid a$ (Lemma 19).

Likewise, $q \mid b$. Thus q is a common positive divisor of a and b. Since a and b are relatively prime, $q = 1$. So $m = |ab|$. □

Theorem 39. *Suppose that $a, b, c \in \mathbb{Z}$, and that a and b are relatively prime. If $a \mid c$ and $b \mid c$ then $ab \mid c$.*

Proof. If $a = 0$ then we must have $c = 0$ since c is a multiple of a. Since $0 \mid 0$ we are done. Likewise if $b = 0$ then $c = 0$, and we are done. So we can now assume a and b are non-zero.

By Theorem 38 the LCM of a and b is $|ab|$. In particular $|ab| \mid c$ by Theorem 35. The result follows from Corollary 20. □

Informal Exercise 26. Give two examples of the above theorem for specific values of a, b, c. Now give two counter-examples if we drop the requirement that a and b be relatively prime.

Here is an important fact about prime numbers:

Theorem 40. *Let $a, b \in \mathbb{Z}$, and p a prime. If $p \mid ab$ then $p \mid a$ or $p \mid b$.*

Proof. If $p \mid a$ we are done, so we will assume that $p \nmid a$. By Theorem 37, p and a are relatively prime. Observe that $a \mid ab$ (def. of divisibility) and $p \mid ab$ (assumption), so $ap \mid ab$ by Theorem 39. By Theorem 14, $p \mid b$ as desired (note $a \neq 0$ since $p \nmid a$). □

Informal Exercise 27. Give two examples of the above theorem for specific values of a, b, p. Now give two counter-examples if we drop the requirement that p be prime.

Definition 8. A *composite number* is a positive integer n such that $n = ab$ for some $a, b \in \mathbb{N}$ with $1 < a < n$ and $1 < b < n$.

Remark 7. Some sources may allow some negative integers to be classified as composite as well, but our definition is adequate for most situations.

Theorem 41. *Let $n \in \mathbb{Z}$. Suppose $n > 1$. Then exactly one of the following occurs: (i) n is prime, or (ii) n is composite.*

Proof. It is clear that both cannot occur: if $n = ab$ with $1 < a < n$ then a is a divisor of n not equal to 1 or n, so n is not a prime.

Now we will show that at least one of the two cases occurs. If n is prime we are done. Otherwise, by the negating the definition of prime, we see that there is a positive divisor a of n such that $a \neq 1$ and $a \neq n$. So $1 < a < n$ (Theorem 17). Since $a \mid n$, there is a $b \in \mathbb{Z}$ such that $ab = n$. Since a and n are positive, b must be as well (the other possibilities lead to contradictions).

The assumptions $b = 1$ or $b = n$ lead to contradictions. Also b is a positive divisor of n. Thus $1 < b < n$ (Theorem 17). Thus n is composite as desired. □

Exercise 28. Show that, in the above proof, the assumption $b = 1$ leads to a contradiction. Show that $b = n$ also leads to a contradiction.

5.8 Three key theorems (informal)

The results developed in the above sections give us a powerful set of tools to explore the integers. We will conclude the chapter by using these tools to prove three important results: (i) every integer $n \geq 2$ can be factored into prime numbers, (ii) the set of prime numbers is an infinite set, and (iii) given a base $B > 1$ every integer has a unique base B representation. We have the tools to prove many more results about $\mathbb{Z}$, but the exploration of advanced properties of $\mathbb{Z}$ is in the purview of a branch of mathematics called *number theory* and goes beyond the scope of this chapter.

In this section we give sketches of the proofs of the three results (i), (ii), and (iii) mentioned above. The sketches are informal and utilize facts about finite sums and products that have not been developed yet. In the next few sections, we will formally develop the needed background on sequences, summation, and products. The chapter ends with the formal proofs of the three featured results (followed by few optional sections).

We begin with the statement that *every $n \geq 2$ can be written as the product of primes*. In other words,

$$n = \prod_{i=1}^{k} p_i$$

for some finite sequence $p_1, \ldots, p_k$ of prime numbers.

This result has a quick proof using strong induction. Let S be the set of integers $x \geq 2$ which can be written as a product of primes. Obviously all primes are in S (use $k = 1$ and $p_1 = n$). In particular, we have the base case: $2 \in S$ since 2 is a prime. Now we wish to show $n \in S$ assuming that $\{2, \ldots, n-1\} \subseteq S$. Given such an $n > 2$, let p be a prime divisor of n, and write $n = pm$. We know such p exists by Theorem 36. If $m = 1$ we are done: n is prime. If $m > 1$ then $m \in \{2, \ldots, n-1\}$. Thus $m \in S$, and so m is the product of primes (inductive hypothesis and definition of S):

$$m = \prod_{i=1}^{k} p_i.$$

So, if p_{k+1} is defined to be p then

$$n = \prod_{i=1}^{k} p_i \cdot p = \prod_{i=1}^{k+1} p_i$$

as desired. By the principle of strong induction, S contains all $n \geq 2$.

The second result is that *the set of prime numbers is infinite*. The proof very old: it can be found in Euclid's *Elements of Geometry*. It proceeds by contradiction: suppose the set S of primes is finite. Then we can list all the primes in a finite sequence $p_1, \ldots, p_k$. Let

$$n = 1 + \prod_{i=1}^{k} p_i.$$

By Theorem 36, there is a prime p dividing n. Since $p_1, \ldots, p_k$ lists all primes, $p = p_j$ for some j. Observe that

$$n - 1 = \prod_{i=1}^{k} p_i,$$

so $p = p_j$ divides $n - 1$. Since p divides n and $n - 1$, it must divide the difference. The difference is 1, a contradiction since 1 has no prime divisors.

The third result is that, given a base $B > 1$, *every positive integer n can be written uniquely in the form*

$$n = \sum_{i=0}^{k} d_i B^i$$

where k is a non-negative integer, and where $d_0, \ldots, d_k$ is a finite sequence with each $d_i \in \{0, \ldots, B-1\}$ where $d_k \neq 0$. The sequence is called the *base B representation of n*.

The proof is by strong induction. Fix $B > 1$ and let S be the set of all positive integers with unique base B representation. First we observe that S contains all integers n where $1 \leq n < B$. To see existence of the base B expansion for n, let $k = 0$ and $d_0 = n$. To see uniqueness of the base B expansion for n, observe that k must be zero, otherwise the sum has value B or more. Since $k = 0$, we must have $d_0 = n$. In particular, we have the base case $1 \in S$.

Now we wish to show $n \in S$ assuming that $\{1, \ldots, n-1\} \subseteq S$. By the above argument, we can assume $n \geq B$. By the Quotient-Remainder Theorem, $n = qB + r$ for some q and r with $r \in \{0, \ldots, B-1\}$. Since $1 \leq q < n$, we have $q \in S$. So

$$q = \sum_{i=0}^{l} e_i B^i$$

for unique l and unique $e_i \in \{0, \ldots, B-1\}$ with $e_l \neq 0$. Thus

$$n = qB + r = B \sum_{i=0}^{l} e_i B^i + r = \sum_{i=0}^{l} e_i B^{i+1} + r = \sum_{i=1}^{l+1} e_{i-1} B^i + r.$$

Let $k = l + 1$, let $d_i = e_{i-1}$ if $1 \leq i \leq k$, and let $d_0 = r$. This choice gives existence. Uniqueness of the base B expansion of n can be proved by (1) using the fact that q and r are unique (using the Quotient-Remainder Theorem), (2) showing that r is d_0 in any base B expansion, and (3) using the fact that the base B expansion of q is unique (by the inductive hypothesis) to show that d_i is unique for $i > 0$.

Observe that the above proofs make use of three key concepts that have not been developed yet: finite sequences $(a_1, \ldots, a_k)$, summations (using $\sum$), and general finite products (using $\prod$). So before we can give formal proofs for the above results, we need to develop these three concepts. This is the purpose of the next three sections.

5.9 Sequences

Functions provide a common framework for much of mathematics, and a surprising number of mathematical objects and concepts are actually just functions. For example, addition is thought of as a binary operator. In other words, addition is a function $\mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$. Other binary operations, such as multiplication and subtraction are also thought of as functions. Successor is such a basic function that it was incorporated into our axioms. We interpreted iteration as composing a function with itself a certain number of times. In Chapter 3 we even interpreted the counting process as a bijective function $\{1, \ldots, n\} \to S$.

In this section we use functions to interpret another basic concept: the sequence. We begin with finite sequences.

Definition 9. Let S be a set. A *finite sequence* with values in S is a function

$$\{m, \ldots, n\} \to S$$

where m and n are integers with $m \leq n$.

Remark 8. If c is such a finite sequence, we usually write $c(i)$ as c_i. As with other kinds of functions, we often define a sequence by giving a rule expressed in terms of a generic element i in the domain. For instance, we might say something like "$c_i = 2^i + 1 \in \mathbb{N}$ where $i = 0, \ldots, 5$" which means that we are defining the sequence c as the function $\{0, \ldots, 5\} \to \mathbb{N}$ given by the rule $i \mapsto 2^i + 1$. The variable i used here to define our sequence can be replaced by any other unused variable. So the above sequence could just as well be defined as $c_j = 2^j + 1 \in \mathbb{N}$ where $j = 0, \ldots, 5$. A variable (such as i above) that can be replaced by any other undeclared variable (such as j above) is called a *bound* or *dummy variable*.

The image of a particular i is called the *ith term of the sequence*, or the *ith value*, and i is called the *index*. The domain $\{m, \ldots, n\}$ is called the *index set*.

Remark 9. We often denote a finite sequence with the notation $(c_i)_{i=m,\ldots n}$, or just (c_i) if there is no need to describe the domain. In other words we just indicate the generic ith term inside parentheses. For instance, the sequence in the previous remark can be written $(2^i + 1)_{i=0,\ldots,5}$. Again, i here is a bound or dummy variable. So

$$(2^i + 1)_{i=m,\ldots,n} = (2^k + 1)_{k=m,\ldots,n}.$$

There are actually a variety of ways to denote sequences. For instance, we could write $(c_m, \ldots, c_n)$ or even $c_m, \ldots, c_n$. Other notation may be employed, but always keep in mind that, regardless of how we denote it, the sequence is just a function.

We often define sequences in terms of other sequences. So the sequence $(3a_i + 4)_{i=m,\ldots,n}$ denotes a sequence $(c_i)_{i=m,\ldots,n}$ where $c_i = 3a_i + 4$ and where (a_i) is a previously given sequence. Similarly $(3B_{2i+1})$ denotes a sequence defined in terms of another sequence (B_i). Here, only some of the terms (the odd terms) of (B_i) are used. The ith term of the new sequence uses the $2i + 1$st term of another sequence (B_i).

The sequence $(b_{i+k})_{i=m-k,\ldots,n-k}$ is not equal to $(b_i)_{i=m,\ldots,n}$ even though they have the same values in the same order. The reason is that sequences are functions, and two functions with differing domains are not equal. The one sequence is called a *shift* of the other.

Now we describe infinite sequences. These are important in analysis, and will be considered later in connection with the real numbers.

Definition 10. If $n \in \mathbb{Z}$ then let $\{n, n+1, \ldots\}$ denote $\{x \in \mathbb{Z} \mid x \geq n\}$.

Definition 11. An *infinite sequence* in S is a function $\{n, n+1, \ldots\} \to S$ where $n \in \mathbb{Z}$.

Remark 10. Notational conventions described in the definition of finite sequences will be extended to infinite sequences in the obvious way. For example, $(a_i)_{i=1,2,\ldots}$ denotes a sequence with domain or index set $\{1, 2, \ldots\}$. We can also write $(a_i)_{i\geq 1}$ or $(a_1, a_2, \ldots)$ to denote such an infinite sequence.

Informal Exercise 29. Consider the values $5, 10, 17, 26, 37$. Define a sequence with these values. What is the domain? What is the index set? What is the value set? Write a shifted sequence with the same values. Extend the original sequence to an infinite sequence.

5.10 Summation

Informally, the expression

$$\sum_{i=m}^{n} b_i$$

denotes the sum $b_m + b_{m+1} + \ldots + b_n$. For example, if we have a sequence with terms $c_1, c_2, c_3, c_4 \in \mathbb{Z}$, then

$$\sum_{i=1}^{4} c_i = c_1 + c_2 + c_3 + c_4.$$

The summation notation can be used for any sequence (b_i) with values b_i in a ring or additive group U. More generally, it can be used for sequences with values in any set U possessing a binary operation called $+$. The formal definition takes the following recursive form:

Definition 12 (Summation)**.** Let (b_i) be a finite or infinite sequence. Suppose each $b_i \in \mathbb{Z}$ or, more generally, each $b_i \in U$ where U is a set possessing a binary operation called $+$. Suppose m is in the index set of (b_i) then we define

$$\sum_{i=m}^{n} b_i$$

recursively for all $n \geq m$ in the index set of (b_i) as follows:

In the base case:

$$\sum_{i=m}^{m} b_i \stackrel{\text{def}}{=} b_m.$$

If $n > m$ then write $n = k + 1$ where $k \geq m$ and use the following:

$$\sum_{i=m}^{k+1} b_i \stackrel{\text{def}}{=} \left(\sum_{i=m}^{k} b_i \right) + b_{k+1}.$$

This recursive definition yields elements of U.

Lemma 42. *The above definition is well-defined.*

Proof. Let $J \subseteq \mathbb{Z}$ be the index set of (b_i). We use Theorem 10 or Theorem 11 depending on whether J is infinite or finite. We describe in detail the finite case; from this it will be clear on how the handle the infinite case.

Write $J = \{e, \ldots, d\}$. Fix m where $e \leq m \leq d$. Let $I = \{m, \ldots, d\}$. Now define a function $g\colon (I - \{d\}) \times U \to U$ by the rule $g(k, x) = x + b_{k+1}$. By Theorem 11, there is a unique function $f : I \to U$ satisfying the two equations

$$\begin{aligned} f(m) &= b_m \\ f(k+1) &= g\big(k, f(k)\big) = f(k) + b_{k+1} \end{aligned}$$

where the second equation holds for all $m \leq k < d$.

We get the desired definition if we write $f(n)$ as $\sum_{i=1}^{n} b_i$. □

Remark 11. The variable i in the above definition is a dummy variable and is not an essential part of the definition. It can be replaced by any other variable not currently in use. So, for instance,

$$\sum_{i=m}^{n} b_i = \sum_{u=m}^{n} b_u.$$

Remark 12. It is important to allow U to be any set with an additive binary operation. This allows us to use the summation idea for all the number systems of the course (including $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}_m$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$) without redeveloping it for each special case. Most of the number systems are rings, but $\mathbb{N}$ is not, so we do not want to restrict ourselves only to rings.

One consequence of the definition is that the summation will be an element of U if all the terms are in U (assuming $+$ is a binary operation on U which means that U is closed under addition). For example, if each $b_i \in \mathbb{N}$ then $\sum b_i \in \mathbb{N}$. Likewise, if each b_i is a positive integer then so is $\sum b_i$, and if each $b_i \in \mathbb{Z}$ is divisible by $d \in \mathbb{Z}$ then so is $\sum b_i$. This is because the set of positive integers and the set of multiples of d are sets closed under addition, and so can be chosen as U.

Exercise 30. Use the above definition to show that

$$\sum_{i=1}^{1} a_i = a_1, \quad \sum_{i=1}^{2} a_i = a_1 + a_2, \quad \sum_{i=1}^{3} a_i = (a_1 + a_2) + a_3,$$

and

$$\sum_{i=1}^{4} a_i = \big((a_1 + a_2) + a_3\big) + a_4$$

Theorem 43 (General distributive law)**.** *Let R be a ring.*[4] *Let c be a constant and let (b_i) be a sequence such that $c \in R$ and each $b_i \in R$. Suppose that the domain of (b_i) contains $\{m, \ldots, n\}$ where $m \leq n$. Then*

$$c \sum_{i=m}^{n} b_i \;=\; \sum_{i=m}^{n} c\, b_i.$$

Proof. We will use the form of induction for a finite set of integers (see Theorem 6). Let S be the set of integers $k \in \{m, \ldots, n\}$ such that

$$c \sum_{i=m}^{k} b_i \;=\; \sum_{i=m}^{k} c\, b_i$$

holds.

[4]Or at least assume R is a set with additive and multiplicative binary operations for which the left distributive law holds.

First we need to show that $m \in S$ (base case). In this case

$$c\left(\sum_{i=m}^{m} b_i\right) = c(b_m)$$

and

$$\sum_{i=m}^{m} (c\, b_i) = (c\, b_m)$$

by Definition 12 (case where $n = m$). Thus $m \in S$.

Now assume that $k \in S$ with $m \leq k < n$. We wish to show $k+1 \in S$. Observe that

$$\begin{aligned} c\left(\sum_{i=m}^{k+1} b_i\right) &= c\left(\left(\sum_{i=m}^{k} b_i\right) + b_{k+1}\right) \quad \text{(Def. 12)} \\ &= c\left(\sum_{i=m}^{k} b_i\right) + c\, b_{k+1} \quad \text{(Distr. Law)} \\ &= \left(\sum_{i=m}^{k} c\, b_i\right) + c\, b_{k+1} \quad \text{(Since } k \in S\text{)} \\ &= \sum_{i=m}^{k+1} c\, b_i \qquad \text{(Def. 12)} \end{aligned}$$

Thus $k+1 \in S$.

By Induction (Theorem 6) we have $\{m, \ldots, n\} \subseteq S$. In particular $n \in S$. The theorem follows. □

Note. If R is a non-commutative ring, then we would want a general right distributive law. Its proof is essentially the same.

Exercise 31. Show that the usual distributive law is just the special case of the above theorem where $m = 1$ and $n = 2$.

For $R = \mathbb{Z}$, or for R any commutative ring, we have a general commutative law. Since the proof and statement are a bit complicated, and since we do not need it in what follows, this is discussed in an optional section below. Informally it states that $\sum a_i$ is preserved when we permute the terms of the sequence (a_i). The following is also a type of commutative law:

Theorem 44. *Let R be a ring.*[5] *Suppose that (b_i) and (c_i) are two sequences in R, and suppose the domains of (b_i) and (c_i) both contain $\{m, \ldots, n\}$ where $m \leq n$. Then*

$$\sum_{i=m}^{n} (b_i + c_i) = \sum_{i=m}^{n} b_i + \sum_{i=m}^{n} c_i.$$

[5] Or at least assume R is a set with an additive binary operation that is commutative and associative.

Exercise 32. Prove the above theorem.

Theorem 45. *Suppose $m \leq n$ where $m, n \in \mathbb{Z}$. Then*

$$\sum_{i=m}^{n} 0 = 0.$$

(Here the summation is in $\mathbb{Z}$ or in any set U with an addition operation that possesses an additive identity 0.)

Exercise 33. Prove the above theorem.

Theorem 46. *Suppose (b_i) is a sequence in $\mathbb{Z}$ (or in any additive abelian group). Suppose the domain of (b_i) contains $\{m, \ldots, n\}$ where $m \leq n$. Then*

$$-\sum_{i=m}^{n} b_i \;=\; \sum_{i=m}^{n} (-b_i)$$

Proof. By Theorem 44 and Theorem 45

$$\sum_{i=m}^{n} b_i + \sum_{i=m}^{n} (-b_i) = \sum_{i=m}^{n} \big(b_i + (-b_i)\big) = \sum_{i=m}^{n} 0 = 0.$$

Now add $-\sum b_i$ to both sides. □

Theorem 47 (General associative law)**.** *Let (b_i) be a sequence in a ring R, or more generally in a set U with an associative binary operation called $+$. Suppose $l, m, n \in \mathbb{Z}$ satisfy $l \leq m - 1$ and $m \leq n$. If the domain of (b_i) contains $\{l, \ldots, n\}$, then*

$$\sum_{i=l}^{n} b_i \;=\; \sum_{i=l}^{m-1} b_i \;+\; \sum_{i=m}^{n} b_i.$$

Proof. We will use the form of induction of Theorem 6 designed for finite sets of integers. Let S be the set of all integers k such that $m \leq k \leq n$ and

$$\sum_{i=l}^{k} b_i \;=\; \sum_{i=l}^{m-1} b_i \;+\; \sum_{i=m}^{k} b_i.$$

First we need to show that $m \in S$ (base case). In this case

$$\sum_{i=l}^{m} b_i \;=\; \sum_{i=l}^{m-1} b_i + b_m \;=\; \sum_{i=l}^{m-1} b_i + \sum_{i=m}^{m} b_i$$

using Definition 12. Thus $m \in S$.

Now assume that $k \in S$ with $m \leq k < n$. We must show $k+1 \in S$. Observe that

$$\begin{aligned}
\sum_{i=l}^{k+1} b_i &= \left(\sum_{i=l}^{k} b_i\right) + b_{k+1} && \text{(Def. 12)} \\
&= \left(\sum_{i=l}^{m-1} b_i + \sum_{i=m}^{k} b_i\right) + b_{k+1} && \text{(Since } k \in S\text{. Ind. Hyp.)} \\
&= \sum_{i=l}^{m-1} b_i + \left(\sum_{i=m}^{k} b_i + b_{k+1}\right) && \text{(Assoc. Law)} \\
&= \sum_{i=l}^{m-1} b_i + \sum_{i=m}^{k+1} b_i && \text{(Def. 12).}
\end{aligned}$$

Thus $k+1 \in S$.

By Induction (Theorem 6) we have $\{m, \ldots, n\} \subseteq S$. In particular $n \in S$. The theorem follows. □

Exercise 34. Show that the general associative law for the case where $l = 1$, $m = 2$, and $n = 4$ is

$$\big((a_1 + a_2) + a_3\big) + a_4 = a_1 + \big((a_2 + a_3) + a_4\big)$$

Exercise 35. Show that the usual associative law is given by the case where $l = 1, m = 2, n = 3$.

Remark 13. As the above exercises illustrate, this theorem is called the *general associative law* because it allows you to move parentheses. For instance, the default placement of (outer) parentheses of $a_1 + a_2 + a_3 + a_4 + a_5$ is $(a_1 + a_2 + a_3 + a_4) + a_5$, but the above theorem allows you to equate this with, for instance, $(a_1 + a_2) + (a_3 + a_4 + a_5)$ or even $a_1 + (a_2 + a_3 + a_4 + a_5)$. Of course, one can use this law to regroup the subgroupings as well. So, for instance, it implies

$$(((a_1 + a_2) + a_3) + a_4) + a_5 = a_1 + \big(a_2 + ((a_3 + a_4) + a_5)\big).$$

To see this, step by step, observe

$$\begin{aligned}
(((a_1 + a_2) + a_3) + a_4) + a_5 &= \sum_{i=1}^{5} a_i && \text{(Def. 12)} \\
&= \sum_{i=1}^{1} a_i + \sum_{i=2}^{5} a_i && \text{(Thm. 47)} \\
&= a_1 + \left(\sum_{i=2}^{2} a_i + \sum_{i=3}^{5} a_i\right) && \text{(Thm 47)} \\
&= a_1 + (a_2 + ((a_3 + a_4) + a_5)) && \text{(Def. 12)}
\end{aligned}$$

Remark 14. The general associative law allows you to move the parentheses around, but it does not allow you to reorder the a_i. For that you need general commutative law discussed in the later (optional) Section 5.15.

There is one more summation law that we will need:

Theorem 48 (Shift of index)**.** *Suppose that (b_i) is a sequence with values in a set U with binary operation $+$. Suppose $m, n, k \in \mathbb{Z}$ are such that $m \leq n$ and such that the domain of (b_i) contains $\{m, \ldots, n\}$. Then*

$$\sum_{i=m}^{n} b_i = \sum_{i=m+k}^{n+k} b_{i-k}.$$

Proof. First observe that the function $i \mapsto b_{i-k}$ has domain containing the set $\{m+k, \ldots, n+k\}$, so $\sum_{i=m+k}^{n+k} b_{i-k}$ is defined. This is due to the fact that $m + k \leq i \leq n + k$ implies $m \leq i - k \leq n$.

The proof will use the form of induction in Theorem 6 for the finite set $\{m, \ldots, n\}$. Let S be the set of all integers $l \in \{m, \ldots, n\}$ such that

$$\sum_{i=m}^{l} b_i = \sum_{i=m+k}^{l+k} b_{i-k}.$$

First we need to show that $m \in S$ (base case). In this case

$$\sum_{i=m}^{m} b_i \;=\; b_m \;=\; b_{(m+k)-k} \;=\; \sum_{i=m+k}^{m+k} b_{i-k}$$

by Definition 12. Thus $m \in S$.

Now assume that $l \in S$ with $m \leq l < n$. We need to show $l + 1 \in S$. Observe that

$$\begin{aligned}
\sum_{i=m}^{l+1} b_i &= \left(\sum_{i=m}^{l} b_i\right) + b_{l+1} && \text{(Def. 12)} \\
&= \left(\sum_{i=m+k}^{l+k} b_{i-k}\right) + b_{((l+1)+k)-k} && \text{(Since } l \in S\text{)} \\
&= \sum_{i=m+k}^{(l+1)+k} b_{i-k} && \text{(Def. 12)}
\end{aligned}$$

Thus $l + 1 \in S$.

By Induction (Theorem 6) we have $\{m, \ldots, n\} \subseteq S$. In particular $n \in S$. The theorem follows. □

5.11 General finite products

Informally, the general finite product

$$\prod_{i=m}^{n} b_i$$

denotes the product $b_m \cdot b_{m+1} \cdots b_n$. For example, if we have a sequence with terms $b_1, b_2, b_3 \in \mathbb{Z}$ then

$$\prod_{i=1}^{3} b_i = b_1 \cdot b_2 \cdot b_3.$$

The finite product can be defined for sequences (b_i) with values in any set U that possesses a binary operation that is written multiplicatively.

The concept of a general finite product is similar to that of a finite sum discussed in the previous section. In fact, the difference in some of the proofs is purely notational, and the definition is identical except for notational changes:

Definition 13 (General products)**.** Let (b_i) be a sequence with values in a ring U or, more generally, a set possessing a binary operation written with multiplicative notation. Suppose m is in the index set of (b_i) then we define

$$\prod_{i=m}^{n} b_i$$

recursively for all $n \geq m$ in the index set of (b_i) as follows:

In the base case:

$$\prod_{i=m}^{m} b_i \stackrel{\text{def}}{=} b_m.$$

If $n > m$ then write $n = k + 1$ where $k \geq m$ and use the following:

$$\prod_{i=m}^{k+1} b_i \stackrel{\text{def}}{=} \left(\prod_{i=m}^{k} b_i \right) \cdot b_{k+1}.$$

This recursive definition yields elements of U.

Lemma 49. *The above definition is well-defined.*

Proof. The proof is a notational variant of the proof of Lemma 42. □

Remark 15. As with summation, it is important to allow U to be any set with a binary operation that is written multiplicatively. In particular, if U is any set that is closed under a multiplication operation, then the general finite

product will have a value in U if all the terms are in U. For example, if all the terms b_i are positive integers, then so is the general finite product $\prod b_i$. This is seen by choosing, in this example, the set U to be the set of positive integers.

Remark 16. The variable i in the above definition is a dummy variable, and can be replaced by any variable not currently in use. So, for instance,

$$\prod_{i=m}^{n} b_i = \prod_{w=m}^{n} b_w.$$

Exercise 36. Suppose $(a_i)_{i=1,\ldots,4}$ is a sequence in a ring R. Use the above definition to show that

$$\prod_{i=1}^{1} a_i = a_1, \quad \prod_{i=1}^{2} a_i = a_1 \cdot a_2, \quad \prod_{i=1}^{3} a_i = (a_1 \cdot a_2) \cdot a_3,$$

and

$$\prod_{i=1}^{4} a_i = \big((a_1 \cdot a_2) \cdot a_3\big) \cdot a_4$$

Theorem 50 (General associative law)**.** *Suppose U is a ring, or at least a set with an associative binary operation written in multiplicative notation. Let (b_i) be a sequence in U. Suppose $l, m, n \in \mathbb{Z}$ satisfy $l \leq m-1$ and $m \leq n$. If the domain of (b_i) contains $\{l, \ldots, n\}$, then*

$$\prod_{i=l}^{n} b_i \;=\; \prod_{i=l}^{m-1} b_i \;\cdot\; \prod_{i=m}^{n} b_i.$$

Proof. Adapt the proof of Theorem 47. □

Exercise 37. Show that the usual associative law is given by the case where $l = 1, m = 2, n = 3$.

Exercise 38. Describe the associative law for the case where $l = 1$, $m = 3$, and $n = 5$.

Now we consider divisibility properties of general finite products. First we prove a special case as a lemma:

Lemma 51. *Suppose (a_j) is a sequence with values in a ring R and with domain containing $\{m, \ldots, n\}$ where $m \leq n$. Then, for some $b \in R$,*

$$\prod_{j=m}^{n} a_j = a_m \cdot b.$$

Proof. If $n = m$ then $\prod_{j=m}^{m} a_j = a_m$ (Def. 13), so let $b = 1$.
If $m < n$ then

$$\begin{aligned}\prod_{j=m}^{n} a_j &= \prod_{j=m}^{m} a_j \cdot \prod_{j=m+1}^{n} a_j \qquad \text{(Thm. 50, Gen. Assoc.)} \\ &= a_m \cdot \prod_{j=m+1}^{n} a_j \qquad \text{(Definition 13).}\end{aligned}$$

□

Theorem 52. *Suppose (a_j) is a sequence with values in $\mathbb{Z}$ and with domain containing $\{m, \ldots, n\}$ where $m \leq n$. Then, for each $i \in \{m, \ldots, n\}$,*

$$a_i \quad \textit{divides} \quad \prod_{j=m}^{n} a_j.$$

Proof. The case $i = m$ is covered by the previous lemma. So assume that $m < i \leq n$. Then

$$\begin{aligned}\prod_{j=m}^{n} a_j &= \prod_{j=m}^{i-1} a_j \cdot \prod_{j=i}^{n} a_j \qquad \text{(Thm. 50, Gen. Assoc.)} \\ &= \prod_{j=m}^{i-1} a_j \cdot (a_i \cdot b) \quad \text{for some } b \in \mathbb{Z} \qquad \text{(Lem. 51)} \\ &= (a_i \cdot b) \cdot \prod_{j=m}^{i-1} a_j \qquad \text{(Comm. Law)} \\ &= a_i \left(b \cdot \prod_{j=m}^{i-1} a_j \right) \qquad \text{(Assoc. Law).}\end{aligned}$$

So a_i divides the product. □

Theorem 53. *Let R be a commutative ring.*[6] *Suppose that (b_i) and (c_i) are two sequences in R. Suppose the domains of (b_i) and (c_i) both contain $\{m, \ldots, n\}$ where $m \leq n$. Then*

$$\prod_{i=m}^{n} (b_i \, c_i) = \prod_{i=m}^{n} b_i \cdot \prod_{i=m}^{n} c_i.$$

Proof. Adapt the proof of Theorem 44. □

[6] Or at least assume R is a set with a multiplicative binary operation that is commutative and associative.

Theorem 54. *Suppose $m \leq n$ where $m, n \in \mathbb{Z}$. Then*

$$\prod_{i=m}^{n} 1 = 1.$$

(Here the product is in a ring U or in any set U with a multiplication operation that possesses an multiplicative identity 1.)

Theorem 55 (Shift of index)**.** *Suppose that (b_i) is a sequence with values in a set U with binary operation written multiplicatively. Suppose $m, n, k \in \mathbb{Z}$ are such that $m \leq n$ and such that the domain of (b_i) contains $\{m, \ldots, n\}$. Then*

$$\prod_{i=m}^{n} b_i = \prod_{i=m+k}^{n+k} b_{i-k}.$$

Proof. Adapt the proof of Theorem 48. □

We end this section with a lemma we will need later.

Lemma 56. *Let (a_i) and (b_i) be sequences with values in a set U and with respective index sets I_a and I_b containing $\{m, \ldots, n\}$. Suppose $a_i = b_i$ for all i with $m \leq i \leq n$ (they can differ for other i). If U has an additive binary operation then*

$$\sum_{i=m}^{n} a_i = \sum_{i=m}^{n} b_i \ ,$$

and if U has a multiplicative binary operation then

$$\prod_{i=m}^{n} a_i = \prod_{i=m}^{n} b_i \ .$$

Proof. Use the form of induction of Theorem 6 for the set $\{m, \ldots, n\}$. □

5.12 Prime factorization

Now that we have developed the concepts of summation and general finite products, we can return to the topics mentioned in Section 5.8. We begin with prime factorizations.

Theorem 57. *Let n be an integer with $n \geq 2$. Then there is a sequence of primes numbers $(p_i)_{i=1,\ldots,k}$ such that*

$$n = \prod_{i=1}^{k} p_i.$$

Proof. Let S be the set of integers $n \geq 2$ with such a prime sequence. Observe that if p is a prime then $p \in S$. To see this, let $p_1 = p$ and $k = 1$. then $\prod_{i=1}^{1} p_i = p$ by Definition 13.

We will use strong induction to show that every n with $n \geq 2$ is in S. The base case $2 \in S$ has already been shown since 2 is a prime. Now we assume $\{2, \ldots, n-1\} \subseteq S$ with the goal of showing $n \in S$.

By Theorem 36, there is a prime p with $p \mid n$. So write $n = pm$ for some $m \in \mathbb{Z}$. Since n and p are positive, m cannot be zero or negative. Thus m is positive. If $m = 1$ then n is a prime, and $n \in S$ as observed above.

Now suppose $m > 1$. Thus $m \geq 2$. Since $p > 1$ we get $mp > m$, so $m < n$. This means $m \in S$ by the inductive hypothesis. So there is a sequence $(p_i)_{i=1,\ldots,k}$ of primes with $m = \prod_{i=1}^{k} p_i$. Define a new sequence $(p'_i)_{i=1,\ldots,k+1}$ of primes by the rule $p'_i = p_i$ if $1 \leq i \leq k$, and $p'_{k+1} = p$. Thus, using Lemma 56 and Definition 13,

$$n = mp = \left(\prod_{i=1}^{k} p_i\right) \cdot p = \left(\prod_{i=1}^{k} p'_i\right) \cdot p'_{k+1} = \prod_{i=1}^{k+1} p'_i.$$

Hence $n \in S$.

By the principle of strong induction (Theorem 8), S contains all $n \geq 2$. □

Informal Exercise 39. Illustrate Theorem 57 for the integers 12, 20, 5, and 84. For $n = 12$ find three different sequences that work.

Remark 17. The above theorem is part of what is know as the *fundamental theorem of arithmetic*. The full version of this theorem also asserts that the sequence of primes for a given n is essentially unique. More precisely, that two sequences for the same n have the same prime values and every prime value occurs the same number of times in both sequences. Another way to say this is that the terms of once sequence can be obtained by permuting the terms of the other. We will wait until a future chapter to prove the full theorem where we prove it for both integers and for polynomials.

5.13 Infinitude of primes

Now we give a formal proof of Euclid's classic theorem.

Theorem 58. *Let $\mathcal{P}$ be the set of prime numbers. Then $\mathcal{P}$ is infinite.*

Proof. Suppose otherwise, suppose that $\mathcal{P}$ is finite. Then there is a bijection $\{1, \ldots, k\} \to \mathcal{P}$ for some k. This bijection can be thought of as a sequence (p_i) with domain $\{1, \ldots, k\}$ and with values that give all the primes (by definition of sequence, Def 9). Let

$$n = 1 + \prod_{j=1}^{k} p_j.$$

By Theorem 36, there is a prime p dividing n. Since p is a prime, $p = p_i$ for some $1 \leq i \leq k$. Observe that

$$n - 1 = \prod_{j=1}^{k} p_j,$$

so $p = p_i$ divides $n - 1$ (Theorem 52). Since p divides n and $n - 1$, it must divide the difference (Corollary 16). The difference is 1. Thus $p \leq 1$ (Theorem 17). This gives a contradiction. □

Informal Exercise 40. Modify the above proof as follows. Suppose $\mathcal{P}$ is finite. Then $\mathcal{P}$ must have a maximum (Chapter 3). Let q be the maximum prime. Let $n = 1 + q!$, and derive a contradiction.

5.14 Base B representations of integers

Let $B > 1$ be a fixed integer called the *base*. The standard in most of the world today is $B = 10$ (decimal). However, $B = 2$ (binary), $B = 8$ (octal), and $B = 16$ (hexadecimal) are common in computer science. In their scientific work the Babylonians used $B = 60$ (sexagesimal), a choice that still survives in our use of minutes and seconds.

To develop base B in general is no harder than to develop a specific base such as 10, so we will develop the general theory.[7]

Remark 18. In a sense, base systems are a luxury, not a necessity. With 0 and σ we can, with enough patience, denote any natural number. Likewise, with $0, 1$ and $+$ we can denote any natural number. For instance, ignoring parentheses, we can denote sixteen as

$$1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1,$$

but it is much more efficient to write '16'.

Definition 14. Let $B > 1$ be a fixed base. A *base B representation* of an integer $n > 0$ is a sequence $(d_i)_{i=1,\ldots,k}$ in $\mathbb{Z}$ with each $0 \leq d_i < B$ such that

$$n = \sum_{i=0}^{k} d_i B^i$$

and such that $d_k \neq 0$. The number d_i is called the *i-th digit* of n.

Remark 19. We often choose specific symbols (numerals) for the numbers in the set $\{0, \ldots, B-1\}$. These symbols are often also called *digits*.[8] We

[7] The above discussion is informal: we have not defined $10, 16$, or 60 yet.

[8] The word *digit* comes from Latin *digitus* meaning 'finger'. So really the term *digit* is most appropriate to base ten. We use it for other bases where we imagine an alien or mythical being with B fingers. Homer Simpson would be be a good choice for $B = 8$.

abbreviate $\sum_{i=0}^{k} d_i B^i$ by listing such symbols for the d_i in order (decreasing the index i as you go from left to right). For example, in base 8 we we can use $[7, 4, 4, 0]$, or simply $[7440]$, to denote

$$7 \cdot 8^3 + 4 \cdot 8^2 + 4 \cdot 8^1 + 0 \cdot 8^0.$$

Exercise 41. Show that a base B representation of B itself is given by the sequence $(d_i)_{i=0,\ldots,1}$ where $d_0 = 0$ and $d_1 = 1$. In other words, we can write B as $[1, 0]$ or $[10]$.

Definition 15. Let *ten* be $9 + 1$ (recall, we have defined 9 in Chapter 1). By the above exercise, ten can be written $[10]$ in base ten.

Definition 16. We will use square brackets around the digits in any base except base ten. In base ten we will usually write the digits without brackets in the usual way. Thus, if $B = 5$ we write $[4, 3, 1]$ or $[431]$ for $4B^2 + 3B + 1$, but if B is ten, we write $4B^2 + 3B + 1$ simply as 431.

In particular, 10 refers to ten. So $10 = 9 + 1$. In general, however, $[10]$ or $[1, 0]$ refers to B where B is the base being used.

In base ten, or in any base, we can separate digits by commas for readability. Thus 20138 or $20,138$ both represent

$$2 \cdot 10^4 + 0 \cdot 10^3 + 1 \cdot 10^2 + 3 \cdot 10 + 8.$$

Definition 17. If B is not clear from context, we can write B in base 10 as a subscript following the base B representation. Thus

$$[24]_5 = 2 \cdot 5 + 4, \qquad [24]_{16} = 2 \cdot 16 + 4$$

where the right-hand sides are in base 10.

Definition 18. When working in base $B > 10$, one needs symbols for digits up to B. Define $a = 9 + 1, b = a + 1, c = b + 1, d = c + 1, e = d + 1$, and $f = e + 1$. Other digits can be defined if needed.[9]

Informal Exercise 42. Using symbols $0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c, d, e, f$ write 1219 in base $B = 16$. Write 1219 in base 4. Write 512 in base 12. Write your answers out in summation and short form. Example:

$$603 = 2B^2 + 5B + b \qquad \text{and} \qquad 603 = [25b]_{16}$$

in base $B = 16$.

[9] In base 60 the custom is to use base 10 to denote the digits up to 59 instead of making up new symbols. So $[34, 2, 17]_{60}$ is the number $34 \cdot 60^2 + 2 \cdot 60 + 17$. Observe that commas must be used to separate digits to avoid confusion.

The key theorem of this section is that every positive integer has a unique base B representation. This shows that base B representation gives a way of representing positive integers, and gives a way of determining if two such integers are distinct. By using '0' and negation '$-$' we can denote all integers uniquely.

Theorem 59. *Let $B > 1$. Then every positive integer has a unique base B representation.*

Before the proof, we give a few lemmas.

Lemma 60. *Suppose (d_i) be a sequence of natural numbers whose domain contains $\{0, \ldots, k\}$ where $k \geq 0$. If $d_k \neq 0$ and B is a positive integer then*

$$\sum_{i=0}^{k} d_i B^i > 0.$$

Proof. If $k = 0$ the result is clear since $d_0 B^0 = d_0$, so assume $k > 0$. By Definition 12,

$$\sum_{i=0}^{k} d_i B^i = \sum_{i=0}^{k-1} d_i B^i + d_k B^k.$$

By the results of Chapter 1, each term $d_i B^i$ is in $\mathbb{N}$. Let $U = \mathbb{N}$. By Definition 12, $\sum_{i=0}^{k-1} d_i B^i$ is in $U = \mathbb{N}$ since $+$ is a binary operation on $\mathbb{N}$ (See Remark 12). Since d_k and B^k are positive by results of Chapter 2, $d_k B^k$ is positive. Thus

$$\sum_{i=0}^{k} d_i B^i = \sum_{i=0}^{k-1} d_i B^i + d_k B^k \geq 0 + d_k B^k > 0.$$

□

Lemma 61. *Let (d_i) be a sequence of integers whose domain contains $\{0, \ldots, k\}$ with $k > 0$. Let B be a positive integer. Then*

$$\sum_{i=0}^{k} d_i B^i = B\left(\sum_{i=1}^{k} d_i B^{i-1}\right) + d_0 = B\left(\sum_{i=0}^{k-1} d_{i+1} B^i\right) + d_0.$$

If $0 \leq d_0 \leq B-1$, then d_0 is the remainder and $\sum_{i=1}^{k} d_i B^{i-1} = \sum_{i=0}^{k-1} d_{i+1} B^i$ is the quotient when we divide $\sum_{i=0}^{k} d_i B^i$ by B.

Proof. (Sketch) Use the general assoc. law (Thm. 47) to separate d_0 from the rest of the terms. Use the general distributive law (Thm. 43) to factor out a B. Use the theorem on shifting the sequence (Thm. 48) to justify the equation $\sum_{i=1}^{k} d_i B^{i-1} = \sum_{i=0}^{k-1} d_{i+1} B^i$. Also use basic laws of Ch. 1. For the final statement, use the Quotient-Remainder Theorem (Thm. 28). □

Lemma 62. *Suppose B and n are positive integers such that $n < B$. Let $d_0 = n$. Then the sequence (d_i) with domain $\{0\}$ is the unique base B representation of n.*

Proof. By definition of summation (Def. 12),

$$\sum_{i=0}^{0} d_i B^i = d_0 B^0 = d_0 = n.$$

We conclude that (d_i) is indeed a base B representation.

Suppose (d'_i) is another base B representation with domain $\{0, \ldots, k'\}$. When we divide n by B we have remainder $r = n$ and quotient $q = 0$ since $n = 0 \cdot B + n$ and $0 \leq n < B$. By Lemma 61, applied to (d'_i) we get that the remainder is d'_0. Thus $d'_0 = n = d_0$. If $k' > 0$ then the quotient is $\sum_{i=0}^{k'-1} d'_{i+1} B^i$ by Lemma 61, and is positive by Lemma 60. This contradicts that the quotient is 0. So $k' = 0$. Thus (d_i) and (d'_i) are the same sequence, giving uniqueness. □

Proof of Main Theorem (Thm. 59). (Strong Induction). Let S be the set of all positive integers with unique base B expansions. Our goal is to show S includes all positive integers. The base case, $1 \in S$, is covered by Lemma 62.

Now we wish to show $n \in S$ assuming that $\{1, \ldots, n-1\} \subseteq S$. If $n < B$ then $n \in S$ by Lemma 62. So we can assume $n \geq B$. By the Quotient-Remainder Theorem, $n = qB + r$ for some q and r with $r \in \{0, \ldots, B-1\}$.

Claim: $1 \leq q$. Suppose otherwise. If $q = 0$, then $n = qB + r = r$, contrarily to the assumption $n \geq B$. If $q < 0$, then $qB < 0$. This implies that $qB + r < r$ contrary to the assumption $n \geq B$. Thus $q \geq 1$.

Since $B > 1$ we get $q < qB \leq qB + r$. Thus $q < n$. Since $1 \leq q < n$ we have $q \in S$ (inductive hypothesis). So q has a unique base B representation (e_i). Thus (e_i) has domain $\{0, \ldots, l\}$ for some l, and

$$q = \sum_{i=0}^{l} e_i B^i$$

where $e_l \neq 0$ and $0 \leq e_i < B$ for all $i \in \{0, \ldots, l\}$. This gives

$$\begin{aligned} n &= qB + r \\ &= B \sum_{i=0}^{l} e_i B^i + r \\ &= \sum_{i=0}^{l} e_i B^{i+1} + r \qquad \text{(General Distr.)} \\ &= r + \sum_{i=1}^{l+1} e_{i-1} B^i \qquad \text{(Thm. 48)} \\ &= \sum_{i=0}^{k} d_i B^i. \qquad \text{(General Assoc., see below for } d_i) \end{aligned}$$

In the last equation, we set $k = l + 1$, and we define (d_i) by the rule $d_0 = r$ and $d_i = e_{i-1}$ for $i \in \{1, \ldots, l + 1\}$. So (d_i) is a sequence with domain $\{0, \ldots, k\}$. Since $e_l \neq 0$ we have $d_k \neq 0$. Likewise, $0 \leq d_i < B$. Thus (d_i) is a base B representation of n. So existence holds for n.

We need to show uniqueness. Suppose (d'_i) is another base B representation with domain $\{0, \ldots, k'\}$. If $k' = 0$ then $n = d'_0$, but $n \geq B$, a contradiction. So we can assume $k' > 0$. By Lemma 61, the remainder is d'_0. But d_0 was defined to be the remainder. Thus $d_0 = d'_0$. By Lemma 61, the quotient q is $\sum_{i=0}^{k'-1} d'_{i+1} B^i$. Earlier we determined that $q \in S$, and defined (e_i) as the unique base B representation of q. So

$$q = \sum_{i=0}^{k'-1} d'_{i+1} B^i = \sum_{i=0}^{l} e_i B^i.$$

By uniqueness (since $q \in S$), $l = k' - 1$ and $d'_{i+1} = e_i$ for $i \in \{0, \ldots, l\}$. This means that $k' = l + 1$ and $d'_i = e_{i-1}$ for all $i \in \{1, \ldots, k'\}$. But $k = l + 1$ and $d_i = e_{i-1}$ for all $i \in \{1, \ldots, k\}$. So $k = k'$ and $d_i = d'_i$ for all i in $\{1, \ldots, k\}$. Uniqueness follows for n. We conclude that $n \in S$.

By the principle of strong induction, S contains all positive integers. The result follows. □

Remark 20. Suppose that you have made an addition table for $n + m$ for all $n, m \in \{0, \ldots, B - 1\}$ (and proved the table is valid). Then you can perform any base B addition using the laws we have proved. Likewise, if you have made a multiplication table for $n \cdot m$ for all $n, m \in \{0, \ldots, B - 1\}$ (and proved it is valid), then you can perform any base B multiplication using the basic laws of arithmetic. The algorithms you were taught in grade school are just a notational short-cut for the more rigorous use of the laws of arithmetic.

For example, suppose you want to add 108 and 17 in a rigorous manner. Suppose your table, which you suppose was derived earlier using rigorous methods from Chapter 1, shows that $8 + 7 = 15$ and $1 + 1 = 2$. Then, using only results from Chapter 1 and this information from the table (and combining several steps into some of the steps),

$$\begin{aligned} 108 + 17 &= ((1 \cdot 10^2 + 0 \cdot 10^1) + 8 \cdot 10^0) + (1 \cdot 10 + 7 \cdot 10^0) \\ &= (10^2 + 8) + (10 + 7) = (10^2 + 10) + (8 + 7) \\ &= (10^2 + 10) + 15 = (10^2 + 1 \cdot 10) + (1 \cdot 10 + 5) \\ &= (10^2 + (1 \cdot 10 + 1 \cdot 10)) + 5 = (10^2 + (1 + 1) \cdot 10^1) + 5 \\ &= (1 \cdot 10^2 + 2 \cdot 10^1) + 5 \cdot 10^0 = 125. \end{aligned}$$

Observe how we "carried the 1". This is a simple example, but in principle any addition and multiplication can be carried out rigorously by a careful use of the laws from Chapter 1. One could attempt to prove (at least informally) that the algorithms that are taught in grade school will always work, and can always be translated to a formal proof.

Informal Exercise 43. Make an addition table for $n + m$ for all n, m in $\{0, \ldots, 7\}$ for base 8. Use the basic laws of Chapter 1 (distributive, etc.) to find a base 8 representation of $[2, 7, 3]_8 + [7, 3, 1]_8$. (You can use the associative and commutative laws freely, drop parentheses, and you can skip steps using these laws. Be sure to write $[2, 7, 3]_8$ as $2 \cdot 8^2 + 7 \cdot 8 + 3$, and so on.)

Informal Exercise 44. Using the techniques of Chapter 1, one could easily make an addition table for $n + m$ for all $n, m \in \{0, \ldots, 9\}$, and prove it to be valid. Do not do this, but assume that someone has done this. Suppose also that you do not yet know how to add larger numbers. Use the basic laws of Chapter 1 to determine $1094 + 329$. Do this by expanding 329 as $3 \cdot 10^2 + 2 \cdot 10 + 9$, and expanding 1094 in a similar manner. (You can use the associative and commutative laws freely, drop parentheses, and skip steps using these laws.)

Informal Exercise 45. Using the techniques of Chapter 1, one could easily make an multiplication table for nm for all $n, m \in \{0, \ldots, 9\}$, and prove it to be valid. Do not do this, but assume that someone has done this. Suppose also that you do not yet know how to multiply larger numbers. Use the basic laws of Chapter 1 to determine $17 \cdot 15$. Do this by expanding 17 as $1 \cdot 10 + 7$, and expanding 15 in a similar manner. (You can use the associative and commutative laws freely, drop parentheses, and skip steps using these laws.)

You should be able to see the connection between the above exercises and the traditional algorithms for addition and multiplication.

5.15 Summation and product conventions

Informally, the general associative law allows us to move parentheses. See Remark 13 for an illustration. This means that in situations where this law applies, there is no real need to use parentheses, and in everyday mathematics parentheses are dropped. We will do so in future chapters, but for definiteness let us agree to the following convention.

Definition 19. If $a, b, c \in U$ where U is a set with a binary operation $+$, then $a + b + c$ is defined to be $\sum_{i=1}^{3} u_i$ where (u_i) is the sequence defined by $u_1 = a$, $u_2 = b$, $u_3 = c$. As we saw above, this means $a+b+c$ is officially defined to be $(a + b) + c$.

If $a, b, c \in U$ where U is a set with a binary operation written multliplicatively, then abc is defined to be $\prod_{i=1}^{3} u_i$ where (u_i) is the sequence defined by $u_1 = a$, $u_2 = b$, $u_3 = c$.

We extend this to more than three terms. So, $a+b+c+d$ is $\sum_{i=1}^{4} u_i$ for the corresponding sequence (u_i). So, officially

$$a + b + c + d = ((a + b) + c) + d.$$

In essence, we are adopting a left association convention for addition and multiplication: the two leftmost terms are bound together first. In most situations, for example in rings such as $\mathbb{Z}$, addition and multiplication are associative, so a right association convention gives the same result, but we choose the left association convention as our official default.

Remark 21. We will adopt a left association convention for functional composition as well. For example, $f \circ g \circ h$ is officially $(f \circ g) \circ h$. Since function composition is associative, so one can prove general associativity laws for composition as well.

Earlier, we defined $n!$ in terms of recursion. However, we can define it in terms of our new notation. If n is a positive integer, define $n!$ as

$$n! \stackrel{\text{def}}{=} \prod_{k=1}^{n} k.$$

You can use methods from this chapter to prove various facts about factorial including $1! = 1$, $2! = 2$, and that in general $(n + 1)! = (n + 1)\, n!$. You can also prove that $1 \leq k \leq n$ implies $k \mid n!$.

Define $0! = 1$ as a special case. In general, you can consider 1 as the product of zero terms (in $\mathbb{Z}$ or any ring R), and 0 as the sum of zero terms (in $\mathbb{Z}$ or any additive group). This allows us to include 1 in the set of natural numbers that can be written as the product of (zero or more) primes. In following sections, this idea will be implemented formally in the commutative case.

If n and m are natural numbers, n^m was defined in Chapter 1. One can prove by induction that if $m \geq 1$ then

$$n^m = \prod_{i=1}^{m} n_i$$

where $n_i = n$ for all $i \in \{1, \ldots, m\}$. The right-hand side makes sense even for negative n. We will explore exponentiation further, even for negative m, in future chapters.

5.16 General commutative laws for sums and products (optional)

Now we will consider other forms of the commutative laws. First we define a summation and product using general index sets in the case where addition or multiplication is commutative and associative.

Indexed sums

Suppose U is a set with a binary operation that is written $+$. Often, this operation is commutative. But when it is not, the order of the terms b_i can influence the value of the sum

$$\sum_{i=m}^{n} bi.$$

As we will see in this section, if $+$ is commutative, then the sum can be defined without specifying a particular order for the terms. In other words, we can take sums of terms b_i where the index $i \in I$ is not necessarily an integer, and where I does not necessarily have a set order.

Definition 20. In what follows, we will use *sequential notation* from Section 5.9 even when the index set I is not $\{m, \ldots, n\}$ or $\{n, n+1, \ldots\}$. The index set I can be any convenient set, perhaps one without a fixed order. We will consider functions $b\colon I \to U$ whose domain is such an index set. The image of $i \in I$ will typically be written b_i instead of $b(i)$. The image $b_i \in U$ of $i \in I$ will be called a *term* or more specifically the ith term. The function as a whole will sometimes be written $(b_i)_{i \in I}$ (instead of just b). The domain I is *index set* and the codomain U is the *value set*. An element of I will be called an *index*.

Because we allow other sets to be index sets, we will not assume that I is an ordered set. In spite of this, we can define

$$\sum_{i \in I} b_i$$

as long as we assume that + is commutative and associative. In other words, the sum will be independent of the order in which we add the terms.

In this section we will assume that U is a set with an binary operation + that is associative and commutative. We will also assume that U has an identity element for +, which we will call $0 \in U$. This means that we assume $0 + x = x$ for all $x \in U$. For now you can think of U as being either $\mathbb{N}$ or $\mathbb{Z}$, but it can be any of the commutative rings introduced later in the course.[10]

Before defining the sum over I, we will temporarily use a notation that *does order the terms* and we employ the summation defined earlier in the chapter.

Definition 21 (Summation on index set with ordering bijection)**.** Let U be a set with a binary operation +. Suppose + has an identity $0 \in U$. Suppose $(b_i)_{i\in I}$ has terms in U and that its index set I is finite of size n. Let f be a bijection $\{1, \ldots, n\} \to I$.

If $n \geq 1$ we define

$$\sum_{i\in I}^{(f)} b_i \stackrel{\text{def}}{=} \sum_{j=1}^{n} b_{f(j)}$$

where the right-hand side is as defined earlier in Section 5.10.

If $n = 0$ and $I = \varnothing$, then we consider $\{1, \ldots, 0\}$ to be the empty set, so in this case f must be the identity on the empty set. In this case we define

$$\sum_{i\in\varnothing}^{(f)} b_i \stackrel{\text{def}}{=} 0.$$

Our goal is to show that when + is commutative and associative, we do not need to specify f, and so can drop it from the notation. To do so we will need a few lemmas:

Lemma 63. *Let U, I and $(b_i)_{i\in I}$ be as in the above definition. Suppose that $I = \{x\}$ has size $n = 1$, and let $f\colon \{1\} \to I$ be the (unique) bijection. Then*

$$\sum_{i\in\{x\}}^{(f)} b_i = b_x.$$

Proof. By Definition 21,

$$\sum_{i\in\{x\}}^{(f)} b_i = \sum_{j=1}^{1} b_{f(j)}.$$

[10] A set U with a binary operation that is associative, and that possesses an identity for the operation, is called a *monoid*. In this section we are assuming that U is a commutative monoid. Abelian groups are commutative monoids. Observe that $\mathbb{N}$ is a commutative monoid under addition but not an abelian group. Also, $\mathbb{N}$ is a commutative monoid under multiplication, but this monoid will require different notation. The multiplicative notation for the ideas of this section will be considered in the next section.

By Exercise 30,

$$\sum_{i=1}^{1} b_{f(i)} = b_{f(1)} = b_x.$$

□

Remark 22. In what follows we will sometimes sum $(b_i)_{i\in I}$ on a subset J of the original index set I. It should be understood from context that we are restricting the function $b\colon I \to U$ to J to get the related function $(b_i)_{i\in J}$ which is indexed by J instead of I.

Lemma 64. *Let $U, I, (b_i)_{i\in I}$, and f be as in the above definition. Assume that I has size $n \geq 1$, and let let $x = f(n)$. Let f_1 be the restriction of f to a function $\{1, \ldots, n-1\} \to I - \{x\}$. Then*

$$\sum_{i\in I}^{(f)} b_i = \left(\sum_{i\in I-\{x\}}^{(f_1)} b_i \right) + b_x$$

Proof. First assume $n = 1$, so $I = \{x\}$. By Lemma 63 and Definition 21 (for the case $n = 0$)

$$\sum_{i\in I}^{(f)} b_i = b_x = 0 + b_x = \left(\sum_{i\in\varnothing}^{(f_1)} b_i \right) + b_x.$$

So the result follows.

So now assume that $n > 1$. Then, by Definition 21 and the recursive definition (Definition 12),

$$\sum_{i\in I}^{(f)} b_i \;=\; \sum_{j=1}^{n} b_{f(j)} \;=\; \sum_{j=1}^{n-1} b_{f(j)} + b_{f(n)}.$$

Observe that $b_{f(n)} = b_x$. Using Lemma 56 and Definition 21 we get

$$\sum_{j=1}^{n-1} b_{f(j)} = \sum_{j=1}^{n-1} b_{f_1(j)} = \sum_{i\in I-\{x\}}^{(f_1)} b_i.$$

The result follows. □

Lemma 65. *Let U, I, and $(b_i)_{i\in I}$ be as in the above definition. Suppose I has size $n = 0$ or $n = 1$. If $f\colon \{1, \ldots, n\} \to I$ and $g\colon \{1, \ldots, n\} \to I$ are bijections then*

$$\sum_{i\in I}^{(f)} b_i = \sum_{i\in I}^{(g)} b_i.$$

Proof. If $n = 0$ then f and g must both be the identity on the empty set. Thus $f = g$ and the result follows from the reflexive law of equality.

If $n = 1$ then $I = \{x\}$, and f and g must both be the map $\{1\} \to \{x\}$ sending 1 to x. Thus $f = g$ and the result follows again. □

The above cover what we want to show in the cases when I has size $n = 0$ and $n = 1$. We will use a strong induction argument to show

$$\sum_{i \in I}^{(f)} b_i$$

is independent of f for finite I in general. To do so we require the following from Chapter 3 (with notational changes for the current situation):

Lemma 66. *Suppose that I is a finite set of size n with element $a \in I$. Then there is a bijection $f : \{1, \ldots, n\} \to I$ with the property that $f(n) = a$.*

Now we can prove the main lemma:

Lemma 67. *Let U be a set with a binary operation $+$ that is associative and commutative, and suppose $0 \in U$ is the identity for $+$. Let $(b_i)_{i \in I}$ have terms in U and finite index set I of size n.*

Suppose f and g are bijections $\{1, \ldots, n\} \to I$. Then

$$\sum_{i \in I}^{(f)} b_i = \sum_{i \in I}^{(g)} b_i.$$

Proof. By Lemma 65 we know this holds for $n = 0$ and $n = 1$, so we will assume $n \geq 2$. We will proceed by strong induction on n. More specifically, we will fix $(b_i)_{i \in I}$, and assume that the equation holds for all $(b'_i)_{i \in I'}$ when I' has size n' strictly less than n.

To proceed, let $x = f(n)$ and $y = g(n)$. Let f_1 be the restriction of f to a bijection $\{1, \ldots, n-1\} \to I - \{x\}$. Likewise, let g_1 be the restriction of g to a bijection $\{1, \ldots, n-1\} \to I - \{y\}$. In the case where $x = y$ we can use the strong induction hypothesis and Lemma 64 to conclude

$$\sum_{i \in I}^{(f)} b_i = \left(\sum_{i \in I - \{x\}}^{(f_1)} b_i \right) + b_x = \left(\sum_{i \in I - \{y\}}^{(g_1)} b_i \right) + b_y = \sum_{i \in I}^{(g)} b_i$$

So assume $x \neq y$. Let f_2 be a bijection $\{1, \ldots, n-1\} \to I - \{x\}$ such that $f_2(n-1) = y$. Let g_2 be a bijection $\{1, \ldots, n-1\} \to I - \{y\}$ such that $g_2(n-1) = x$. These exist by Lemma 66. Finally, let f_3 be the restriction of f_2 to a bijection $\{1, \ldots, n-2\} \to I - \{x, y\}$, and let g_3 be the

restriction of g_2 to a bijection $\{1, \ldots, n-2\} \to I - \{x, y\}$. By Lemma 64, the strong induction hypothesis, associativity, and commutativity we get

$$\begin{aligned}
\sum_{i \in I}^{(f)} b_i &= \left(\sum_{i \in I - \{x\}}^{(f_1)} b_i \right) + b_x = \left(\sum_{i \in I - \{x\}}^{(f_2)} b_i \right) + b_x \\
&= \left(\left(\sum_{i \in I - \{x,y\}}^{(f_3)} b_i \right) + b_y \right) + b_x \\
&= \left(\sum_{i \in I - \{x,y\}}^{(f_3)} b_i \right) + (b_y + b_x) \\
&= \left(\sum_{i \in I - \{x,y\}}^{(g_3)} b_i \right) + (b_x + b_y) \\
&= \left(\left(\sum_{i \in I - \{x,y\}}^{(g_3)} b_i \right) + b_x \right) + b_y \\
&= \left(\sum_{i \in I - \{y\}}^{(g_2)} b_i \right) + b_y = \left(\sum_{i \in I - \{y\}}^{(g_1)} b_i \right) + b_y \\
&= \sum_{i \in I}^{(g)} b_i
\end{aligned}$$

□

With this lemma, we can legitimately make the following definition. In other words, the following definition is well-defined: it is independent of choice of f.

Definition 22 (Commutative summation on a finite index set)**.** Let U be a set with a binary operation $+$ that is associative and commutative, and that has an identity element for $+$. Let $(b_i)_{i \in I}$ have terms in U and finite index set I of size n. We define

$$\sum_{i \in I} b_i \stackrel{\text{def}}{=} \sum_{i \in I}^{(f)} b_i$$

where $f \colon \{1, \ldots, n\} \to I$ is any given choice of bijection.

This gives us a new type of sum, but we should confirm that it agrees with our previous definition of summation if $I = \{m, \ldots, n\}$:

Lemma 68. *Let U be as in the above definition. Suppose $I = \{m, \ldots, n\}$ where $m \leq n$ and $m, n \in \mathbb{Z}$, and suppose $(b_i)_{i \in I}$ is a sequence with terms in U indexed by I. Then*

$$\sum_{i \in \{m,\ldots,n\}} b_i = \sum_{i=m}^{n} b_i$$

where the left-hand side represents the sum defined in this section (Definition 22), and the right-hand side corresponds to the sum defined earlier in the chapter (Definition 12).

Proof. Let $N = (n - m) + 1$. Consider the function

$$f \colon \{1, \ldots, N\} \to \{m, \ldots, n\}$$

defined by the equation $f(i) = i + (m - 1)$. The function f has an inverse given by $j \mapsto (j - m) + 1$ so f is a bijection. Thus

$$\begin{aligned}
\sum_{i \in \{m,\ldots,n\}} b_i &= \sum_{i \in \{m,\ldots,n\}}^{(f)} b_i && \text{(Definition 22)} \\
&= \sum_{j=1}^{N} b_{f(j)} = \sum_{j=1}^{(n-m)+1} b_{j+(m-1)} && \text{(Definition 21)} \\
&= \sum_{i=1+(m-1)}^{(n-m)+1+(m-1)} b_{i+(m-1)-(m-1)} && \text{(Theorem 48)} \\
&= \sum_{i=m}^{n} b_i.
\end{aligned}$$

□

Finally we end with some key properties of this new sum.

Theorem 69. *Let U be a set with a binary operation $+$ that is associative and commutative, and that has an identity element $0 \in U$ for $+$. Let $(b_i)_{i \in I}$ have terms in U and finite index set I of size n.*

If $n = 0$ and $I = \varnothing$ then

$$\sum_{i \in \varnothing} b_i = 0.$$

If $n = 1$ and $I = \{x\}$ then

$$\sum_{i \in \{x\}} b_i = b_x.$$

If $n \geq 1$ and $x \in I$ then

$$\sum_{i \in I} b_i = \left(\sum_{i \in I - \{x\}} b_i \right) + b_x.$$

Proof. If $n = 0$ and $I = \varnothing$ then let $f\colon \varnothing \to \varnothing$ be the identity. So, by Definition 22 and Definition 21,

$$\sum_{i\in\varnothing} b_i = \sum_{i\in\varnothing}^{(f)} b_i = 0.$$

If $n = 1$ and $I = \{x\}$ then let $f\colon \{1\} \to \{x\}$ be the map sending 1 to x. So, by Definition 22 and Lemma 63,

$$\sum_{i\in\{x\}} b_i = \sum_{i\in\{x\}}^{(f)} b_i = b_x.$$

If $n \geq 1$ then let $f\colon \{1,\ldots,n\} \to I$ be a bijection such that $f(n) = x$. This exists by Lemma 66. So, by Definition 22 and Lemma 64,

$$\sum_{i\in I} b_i = \sum_{i\in I}^{(f)} b_i = \left(\sum_{i\in I-\{x\}}^{(f_1)} b_i\right) + b_x = \left(\sum_{i\in I-\{x\}} b_i\right) + b_x.$$

where f_1 is the restriction of f to a bijection $f_1\colon \{1,\ldots,n-1\} \to I-\{x\}$. □

Indexed products

In this section we consider the multiplicative versions of the results of the previous section. The proofs are completely parallel to those of the previous section, and so can be safely omitted.

Similarly to the last section, we we assume U is a set with a binary operation, written multiplicatively, with an identity (which we write as 1). We consider $(b_i)_{i\in I}$ with terms in U and indexed by finite sets I. We first define

$$\prod_{i\in I}^{(f)} b_i$$

where $f\colon \{1,\ldots,n\} \to I$ is a bijection. The definition is similar to that for $\sum$ in the last section. Next, the following definition can be shown to be well-defined: the product is independent of choice of f.

Definition 23 (Commutative product on a finite index set)**.** Let U be a set with a binary operation, written multiplicatively, that is associative and commutative, and that has an identity element $1 \in U$. Let $(b_i)_{i\in I}$ have terms in U with finite index set I of size n. We define

$$\prod_{i\in I} b_i \stackrel{\text{def}}{=} \prod_{i\in I}^{(f)} b_i$$

where $f\colon \{1,\ldots,n\} \to I$ is any given choice of bijection.

This new type of product agrees with our previous definition of product in the case where $I = \{m, \ldots, n\} \subseteq \mathbb{Z}$:

Lemma 70. *Let U be as in the above definition. Suppose $I = \{m, \ldots, n\}$ where $m \leq n$ and $m, n \in \mathbb{Z}$, and suppose $(b_i)_{i \in I}$ is a sequence of terms in U indexed by I. Then*

$$\prod_{i \in \{m,\ldots,n\}} b_i = \prod_{i=m}^{n} b_i$$

where the left-hand side represents the product defined in this section (Definition 23), and the right-hand side corresponds to the product defined earlier in the chapter (Definition 13).

We end with some key properties of this new product.

Theorem 71. *Let U be a set with a binary operation, written multiplicatively, that is associative and commutative, and that has an identity element $1 \in U$. Let $(b_i)_{i \in I}$ have terms in U and finite index set I of size n.*

If $n = 0$ and $I = \varnothing$ then

$$\prod_{i \in \varnothing} b_i = 1.$$

If $n = 1$ and $I = \{x\}$ then

$$\prod_{i \in \{x\}} b_i = b_x.$$

If $n \geq 1$ and $x \in I$ then

$$\prod_{i \in I} b_i = \left(\prod_{i \in I - \{x\}} b_i \right) \cdot b_x.$$

General commutative laws

Now we investigate general commutative laws for sums and products. The first version is based on, and generalizes, Theorem 47.[11]

Theorem 72. *Let U be a set with a binary operation $+$ that is associative and commutative, and that has an identity element $0 \in U$. Let $(b_i)_{i \in I}$ have terms in U and a finite index set I of size n. Let I_1 and I_2 be two disjoint subsets such that $I = I_1 \cup I_2$. Consider also the restriction $(b_i)_{i \in I_1}$ and $(b_i)_{i \in I_2}$ of the function $(b_i)_{i \in I}$. Then*

$$\sum_{i \in I} b_i = \sum_{i \in I_1} b_i + \sum_{i \in I_2} b_i.$$

[11] Theorem 47 was described as a general associative law, but the generalization here can also be thought of as a kind of commutative law.

Proof. Let n_1 be the size of I_1 and let n_2 be the size of I_2. Thus I must have size $n_1 + n_2$ (see Chapter 3). Observe that the desired equality holds if $n_1 = 0$ or $n_2 = 0$. For example, if $n_1 = 0$ then $I_1 = \varnothing$, and $\sum_{i\in\varnothing} b_i = 0$. In this case $I_2 = I$. So the equation holds since 0 is an identity.

So assume $n_1 > 0$ and $n_2 > 0$. Next choose bijections $f_1 \colon \{1, \ldots, n_1\} \to I_1$ and $f_2 \colon \{1, \ldots, n_2\} \to I_2$. Consider the function $f \colon \{1, \ldots, n_1+n_2\} \to I_1 \cup I_2$ defined by the rule

$$f(j) = \begin{cases} f_1(j) & \text{if } 1 \le j \le n_1 \\ f_2(j - n_1) & \text{if } n_1 < j \le n_1 + n_2. \end{cases}$$

Now establish that f is injective and surjective, which is straightforward.

Observe that the sequence $(b_{f(i)})$ has domain $\{1, \ldots, n_1 + n_2\}$, and that

$$\begin{aligned}
\sum_{i\in I_1\cup I_2} b_i &= \sum_{i\in I_1\cup I_2}^{(f)} b_i && \text{(Definition 22)} \\
&= \sum_{j=1}^{n_1+n_2} b_{f(j)} && \text{(Definition 21)} \\
&= \sum_{j=1}^{n_1} b_{f(j)} + \sum_{j=1+n_1}^{n_1+n_2} b_{f(j)} && \text{(Theorem 47)} \\
&= \sum_{j=1}^{n_1} b_{f_1(j)} + \sum_{j=1+n_1}^{n_1+n_2} b_{f_2(j-n_1)} && \text{(definition of } f\text{)} \\
&= \sum_{j=1}^{n_1} b_{f_1(j)} + \sum_{j=1}^{n_2} b_{f_2(j)} && \text{(Theorem 48)} \\
&= \sum_{i\in I_1}^{(f_1)} b_i + \sum_{i\in I_2}^{(f_2)} b_i && \text{(Definition 21)} \\
&= \sum_{i\in I_1} b_i + \sum_{i\in I_2} b_i && \text{(Definition 22)}
\end{aligned}$$

Here we needed the general associative law (Theorem 47). □

The multiplicative version is proved in a similar manner:

Theorem 73. *Let U be a set with a binary operation, written multiplicatively, that is associative and commutative, and that has an identity element $1 \in U$. Let $(b_i)_{i\in I}$ have terms in U and a finite index set I of size n. Let I_1 and I_2 be two disjoint subsets such that $I = I_1 \cup I_2$. Consider also the restriction $(b_i)_{i\in I_1}$ and $(b_i)_{i\in I_2}$ of the function $(b_i)_{i\in I}$. Then*

$$\prod_{i\in I} b_i = \prod_{i\in I_1} b_i \cdot \prod_{i\in I_2} b_i.$$

Recall that Theorem 44 can also be thought of as a type of commutative law. This can be extended to the current situation. (The proof is simple, and the idea behind it can be used to extend several results from Sections 5.10 and 5.11).

Theorem 74. *Let U be a set with a binary operation $+$ that is associative and commutative, and that has an identity element $0 \in U$. Let $(b_i)_{i\in I}$ and $(c_i)_{i\in I}$ have terms in U and share a finite index set I. Then*

$$\sum_{i\in I}(b_i + c_i) = \sum_{i\in I} b_i + \sum_{i\in I} c_i.$$

Proof. The expression $(b_i + c_i)_{i\in I}$ refers to the function $(d_i)_{i\in I}$ defined by the rule $i \mapsto b_i + c_i$. So our goal is to prove

$$\sum_{i\in I} d_i = \sum_{i\in I} b_i + \sum_{i\in I} c_i$$

with this (d_i). First recall that $0 + 0 = 0$ so the result follows if $I = \varnothing$. So we can assume I has size $n \geq 1$. Let $f\colon \{1, \ldots, n\} \to I$ be a bijection. So

$$\begin{aligned}
\sum_{i\in I} d_i &= \sum_{i\in I}^{(f)} d_i = \sum_{j=1}^{n} d_{f(j)} && \text{(Definition 22 and 21)}\\
&= \sum_{j=1}^{n} b_{f(j)} + \sum_{j=1}^{n} c_{f(j)} && \text{(Theorem 44)}\\
&= \sum_{i\in I}^{(f)} b_i + \sum_{i\in I}^{(f)} c_i && \text{(Definition 21)}\\
&= \sum_{i\in I} b_i + \sum_{i\in I} c_i && \text{(Definition 22).}
\end{aligned}$$

□

The multiplicative version is proved in a similar manner:

Theorem 75. *Let U be a set with a binary operation, written multiplicatively, that is associative and commutative, and that has an identity element $1 \in U$. Let $(b_i)_{i\in I}$ and $(c_i)_{i\in I}$ have terms in U and share a finite index set I. Then*

$$\prod_{i\in I}(b_i \cdot c_i) = \prod_{i\in I} b_i \cdot \prod_{i\in I} c_i.$$

Before giving the third, and most general, version of the commutative law, we consider a method of changing the index set:

Theorem 76. *Let U be a set with a binary operation $+$ that is associative and commutative, and that has an identity element $0 \in U$. Let $(b_i)_{i\in I}$ have terms in U and finite index set I. Suppose $\sigma\colon I' \to I$ is a bijection, and consider the function $(b_{\sigma i})_{i\in I'}$. Then*

$$\sum_{i\in I'} b_{\sigma i} = \sum_{i\in I} b_i.$$

Proof. If $I = \varnothing$, then the existence of σ forces $I' = \varnothing$ and then both sums are just 0. So we can assume that I and I' have size $n \geq 1$.

Choose a bijection $f\colon \{1,\ldots,n\} \to I'$. Observe that $\sigma \circ f\colon \{1,\ldots,n\} \to I$ is a bijection as well. So

$$\begin{aligned}
\sum_{i\in I'} b_{\sigma i} &= \sum_{i\in I'} b(\sigma i) = \sum_{i\in I'}^{(f)} b(\sigma i) && \text{(Definition 22)}\\
&= \sum_{j=1}^{n} b(\sigma(f(j))) && \text{(Definition 21)}\\
&= \sum_{j=1}^{n} b((\sigma\circ f)(j)) && \text{(Definition of composition)}\\
&= \sum_{i\in I}^{(\sigma\circ f)} b(i) = \sum_{i\in I} b_i && \text{(Definitions 21 and 21)}
\end{aligned}$$

□

The multiplicative version is proved in a similar manner:

Theorem 77. *Let U be a set with a binary operation, written multiplicatively, that is associative and commutative, and that has an identity element $1 \in U$. Let $(b_i)_{i\in I}$ have terms in U and finite index set I. Suppose $\sigma\colon I' \to I$ is a bijection, and consider the function $(b_{\sigma i})_{i\in I'}$. Then*

$$\prod_{i\in I'} b_{\sigma i} = \prod_{i\in I} b_i.$$

We are especially interested in applying the above two theorems in the case where σ is a bijection from I to itself; in other words, we are interested in the case where $I' = I$. It is common to use the term *permutation* for bijection from a finite set to itself, and this term can be used even for infinite sets.

Definition 24. A *permutation* of a set S is a bijection $S \to S$.

Example. If $\sigma\colon I \to I$ is a permutation of $I = \{1, \ldots, n\}$ say, and if $(a_i)_{i\in I}$ is a sequence of terms in $\mathbb{Z}$ indexed by I, then we can use σ to produce another sequence $(a_{\sigma i})_{i\in I}$ with the same terms, but in a different order.[12]

For example, let $(a_i)_{i\in I}$ be the sequence indexed by $I = \{1, 2, 3\}$ defined by $a_1 = 3$, $a_2 = 11$, and $a_3 = 12$. Suppose σ is the permutation of I defined by $1 \mapsto 2$, $2 \mapsto 3$, and $3 \mapsto 1$. Then $(a_{\sigma i})_{i\in I}$ is the sequence $(b_i)_{i\in I}$ where $b_1 = 11$, $b_2 = 12$, and $b_3 = 3$. The two sequences have the same terms, but in a different order.

The general commutative law shows that two such sequences must have the same sum and product.

We can now state most general form of the commutative law:

Theorem 78 (General commutative law)**.** *Let $(b_i)_{i\in I}$ have terms in U and finite index set I. Suppose $\sigma\colon I \to I$ is a permutation of I. If U is a set with an commutative and associative binary operation written as $+$ and with an additive identity, then*

$$\sum_{i\in I} b_i = \sum_{i\in I} b_{\sigma i}.$$

If U is a set with an commutative and associative binary operation written multiplicatively and with a multiplicative identity, then

$$\prod_{i\in I} b_i = \prod_{i\in I} b_{\sigma i}.$$

Proof. This is really just a corollary of the previous theorems. Just consider the case where $I' = I$. □

Example. If σ is defined by the rule $1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 1$, then, under the assumptions of the above theorem, we get

$$b_1 + b_2 + b_3 = b_2 + b_3 + b_1 \qquad b_1\, b_2\, b_3 = b_2\, b_3\, b_1.$$

If σ is defined by the rule $1 \mapsto 3, 2 \mapsto 2, 3 \mapsto 1$. then, under the assumptions of the above theorem, we get

$$b_1 + b_2 + b_3 = b_3 + b_2 + b_1 \qquad b_1\, b_2\, b_3 = b_3\, b_2\, b_1.$$

In these examples, $I = \{1, \ldots, 3\}$. As you see, by choosing σ appropriately, you can rearrange the terms anyway you would like.

[12] We will often use the symbol σ for a general permutation. This should not be confused with the successor function defined in Chapter 1.

Chapter 6

Modular Arithmetic

In this chapter we will consider congruence $\equiv$ modulo m, and explore the associated arithmetic called *modular arithmetic*. This will lead us to the system $\mathbb{Z}_m$ where m is a positive integer. Unlike other number systems, $\mathbb{Z}_m$ is finite: it has m elements. We will define an addition and multiplication operation on $\mathbb{Z}_m$, and show that they have many of the same properties that one finds in other number systems. In fact, we will prove that $\mathbb{Z}_m$ has enough arithmetic properties to be a commutative ring.

We will determine which elements of $\mathbb{Z}_m$ have multiplicative inverses. These will be called the *units* of $\mathbb{Z}_m$. The set of units forms an abelian group. When $m = p$ is a prime, we will show further that every non-zero element of $\mathbb{Z}_p$ has a multiplicative inverse. This will show that $\mathbb{Z}_p$ is a *field*. We will sometimes write $\mathbb{F}_p$ for $\mathbb{Z}_p$ to indicate that we are indicating a field with p elements. We will consider fields in general.

We will end with a discussion of exponentiation in general rings including negative exponents for units, and will show that such exponentiation satisfies the usual properties.

6.1 Congruence modulo m

Let m be a fixed positive integer. We call m the *modulus*.

Definition 1. If $a, b \in \mathbb{Z}$ are such that $\text{Rem}(a, m) = \text{Rem}(b, m)$, then we say that a and b are *congruent* modulo m and write

$$a \equiv b \mod m \qquad \text{or} \qquad a \equiv_m b.$$

Note. Recall that whenever the word *if* is used to define a new concept, it really means, from a logical point of view, *if and only if*. The above gives an example of this. The definition stipulates that

$$\text{Rem}(a, m) = \text{Rem}(b, m) \iff a \equiv b \bmod m \iff a \equiv_m b.$$

Note. The abbreviation "mod m" is short for "modulo m", which in Latin means "using modulus m".

Theorem 1. *Fix m. Then $\equiv_m$ is an equivalence relation on $\mathbb{Z}$.*

Exercise 1. Prove the above.

A very common use of congruences is to assert $a \equiv r \bmod m$ where r is the remainder $\text{Rem}(a, m)$. This is supported in the following theorem. For example, one would commonly say $7 \equiv 2 \bmod 5$. Although it is very common, you do not have to put the remainder on the right-hand side. It is also valid to write $7 \equiv 17 \bmod 5$, or $7 \equiv 7 \bmod 5$, or even $7 \equiv -13 \bmod 5$.

The following gives the common case when the remainder is on the right-hand side.

Theorem 2. *If $a, m \in \mathbb{Z}$ with $m > 0$ then*

$$a \equiv \text{Rem}(a, m) \mod m.$$

Lemma 3. *If $0 \leq c < m$ then $\text{Rem}(c, m) = c$.*

Proof of lemma. Since $c = 0 \cdot m + c$ and $0 \leq c < m$, the quotient-remainder theorem (Chapter 5) forces c to be the remainder $\text{Rem}(c, m)$ by the uniqueness of the remainder. □

Proof of theorem. Let $r = \text{Rem}(a, m)$. Because $0 \leq r < m$ we have that $r = \text{Rem}(r, m)$ by the above lemma. Thus $\text{Rem}(a, m) = \text{Rem}(r, m)$. By definition of congruence, $a \equiv_m r$. □

Since congruence is reflexive, an equality can always be converted to a congruence. The following says that for small integers, a congruence can be converted to an equality.

Theorem 4. *Suppose $a, b, m \in \mathbb{N}$ where $0 \leq a < m$ and $0 \leq b < m$. Then*

$$a \equiv_m b \iff a = b.$$

Proof. Assume $a \equiv_m b$. Thus $\text{Rem}(a, m) = \text{Rem}(b, m)$ by Definition 1. By Lemma 3, $\text{Rem}(a, m) = a$ and $\text{Rem}(b, m) = b$. Thus $a = b$.

The other direction follows from the fact that $\equiv_m$ is reflexive (congruence is an equivalence relation). □

Corollary 5. *Suppose $a, m \in \mathbb{Z}$ with $m > 0$. Then there is exactly one element $b \in \{0, \ldots, m-1\}$ such that*

$$a \equiv b \mod m.$$

Exercise 2. Justify the above corollary. Hint: use Theorems 2 and 4.

The following is another characterization of congruence. It is often chosen as the definition of congruence in number theory books.

Theorem 6. *Let $a, b \in \mathbb{Z}$. Let m be a positive integer. Then*

$$a \equiv_m b \iff m \mid (a-b).$$

Proof. Suppose $a \equiv b$ modulo m. Then a and b have the same remainder (but perhaps different quotients). So we have $a = qm + r$ and $b = q'm + r$ for some $q, q' \in \mathbb{Z}$. Thus

$$a - b = (qm + r) - (q'm + r) = (q - q')m.$$

This implies that $m \mid (a-b)$.

Suppose $m \mid (a-b)$. So $a - b = cm$ for some $c \in \mathbb{Z}$. Thus $a = b + cm$. Apply the Quotient Remainder Theorem (Ch. 5) to b giving us $b = qm + r$ with $0 \leq r < m$. Thus

$$a = b + cm = (qm + r) + cm = (q + c)m + r.$$

Since $0 \leq r < m$, this implies that the quotient and remainder for a divided by m are $q + c$ and r respectively. In particular, a and b have the same remainder r. Thus $a \equiv b \mod m$. □

Exercise 3. Use the above theorem to show that $a \equiv b$ modulo 1 is always true (for all $a, b \in \mathbb{Z}$).

6.2 Modular arithmetic

The first rule of modular arithmetic allows you to add a constant to both sides of a congruence.

Theorem 7. *Let $a, b, c, m \in \mathbb{Z}$ with $m > 0$.*

$$a \equiv b \mod m \implies a + c \equiv b + c \mod m.$$

Proof. By Theorem 6, $m \mid (a-b)$. But

$$(a + c) - (b + c) = a + c + (-b) + (-c) = a - b.$$

So $m \mid \big((a + c) - (b + c)\big)$. The conclusion follows from Theorem 6. □

Using the above twice gives the following

Theorem 8. *Let $a, b, a', b', m \in \mathbb{Z}$ with $m > 0$.*

$$a \equiv_m a' \quad \text{and} \quad b \equiv_m b' \quad \Longrightarrow \quad a + b \equiv_m a' + b'.$$

Proof. By Theorem 7, we can add a to both sides of $b \equiv_m b'$:

$$b + a \equiv b' + a \mod m.$$

By the commutative law we get

$$a + b \equiv a + b' \mod m.$$

By Theorem 7 again, we can add b' to both sides of $a \equiv_m a'$:

$$a + b' \equiv a' + b' \mod m.$$

Now use transitivity. □

Informal Exercise 4. Illustrate Theorems 8 and 10 with several examples.

Not only can you add constants to congruences, you can multiply constants to congruences.

Theorem 9. *Let $a, b, c, m \in \mathbb{Z}$ with $m > 0$.*

$$a \equiv_m b \quad \Longrightarrow \quad ac \equiv_m bc.$$

Exercise 5. Use Theorem 6 to prove the above.

Theorem 10. *Let $a, b, a', b', m \in \mathbb{Z}$ with $m > 0$.*

$$a \equiv_m a' \quad \text{and} \quad b \equiv_m b' \quad \Longrightarrow \quad ab \equiv_m a'b'.$$

Exercise 6. Prove the above. Hint: see Theorem 8 for ideas.

Exercise 7. Let $a, b, m, n \in \mathbb{Z}$ where $m > 0$ and $n > 0$. Although we will give an official definition later for rings in general, for the purpose of this exercise define exponentiation for integers as follows:

$$a^n \stackrel{\text{def}}{=} \prod_{i=1}^{n} a.$$

Prove by induction that if $a \equiv_m b$, then $a^n \equiv_m b^n$.

Informal Exercise 8. Use congruences to show that adding 52 hours to a clock is the same as adding 4 to a clock.

Informal Exercise 9. Ignoring the effect of leap years, consecutive birthdays differ by 365 days. Suppose this is so, where the first of the consecutive birthdays occurs on a Friday. Use congruences to show that the second of the consecutive birthdays must be on a Saturday.

Exercise 10. Suppose that $a, m \in \mathbb{Z}$ with $m > 0$. Show that you can freely add or subtract m in a congruence:

$$a \equiv a + m \equiv a - m \mod m.$$

For instance, if you are given $a = -6$ and $m = 8$, you could add 8 and conclude $-6 \equiv 2$ modulo 8.

Exercise 11. Suppose that $a, d, m \in \mathbb{Z}$ with $m > 0$. Suppose $d \mid m$ where $d > 0$. Show that if $a \equiv_m b$ then $a \equiv_d b$. Hint: use Theorem 6.

6.3 Application to finding remainders in $\mathbb{Z}$

In this section we give quick ways to find remainders when we divide by various small integers. The technique is based on writing a number in base 10, but it generalizes easily to other bases.

Exercise 12. Show that

$$10^n \equiv 1^n \equiv 1 \mod 9$$

and

$$10^n \equiv 1 \mod 3.$$

for all $n \in \mathbb{N}$.

Informal Exercise 13. Let s be the sum of the digits of a number $n \in \mathbb{N}$ written in base 10. Show that

$$n \equiv s \mod 9$$

and

$$n \equiv s \mod 3.$$

Informal Exercise 14. Use the sum of the digits method to find $\text{Rem}(3785, 9)$ and $\text{Rem}(1234567, 3)$. What is the closest number to $54,992$ that is divisible by 9? (without actually dividing by 9)

Informal Exercise 15. Derive your own procedure for finding $\text{Rem}(n, 7)$ where $n \in \mathbb{N}$ has up to three digits. Hint: work with 10^k modulo 7 for $k = 0, 1, 2$. Use this procedure to find $\text{Rem}(258, 7)$ and $\text{Rem}(732, 7)$.

Exercise 16. Prove (without induction) that

$$B^n \equiv 0 \mod B$$

for all positive integers B and n. Hint: write n as $m + 1$.

Informal Exercise 17. Show that if n is a natural number, and m is the last digit of n written in base B, then $n \equiv m \bmod B$. What is a quick way to find the remainder of $n \in \mathbb{N}$ when you divide by 10? what is $\text{Rem}(12329392912013, 10)$? What is the remainder of $[1000000000003575]_{16}$ when you divide by 16?

Informal Exercise 18. Show that for $n \in \mathbb{N}$, $10^n \equiv (-1)^n \bmod 11$. Hint: use Exercise 7.

Informal Exercise 19. Let $n \in \mathbb{N}$, and write n in base 10 as

$$n = \sum_{i=0}^{k} d_i 10^i.$$

Show that

$$n \equiv \sum_{i=0}^{k} (-1)^i d_i \mod 11.$$

Use this to find the remainder of $158,347$ when dividing by 11.

Informal Exercise 20. Observe that $4 \mid 10^2$. Show that to find $\text{Rem}(n, 4)$, you just need to replace $n \in \mathbb{N}$ with the number formed from the last two digits of n. Show that if d_1 and d_0 are the last two digits, then

$$n \equiv 2d_1 + d_0 \bmod 4.$$

Informal Exercise 21. How many digits do you need to consider when calculating $\text{Rem}(n, 8)$? Explain why.

Informal Exercise 22. How many digits do you need to consider when calculating $\text{Rem}(n, 5)$ or $\text{Rem}(n, 2)$? Explain why.

Informal Exercise 23. Find the remainder of 357 when dividing by $2, 3, 4, 5, 7, 9, 10$, and 11 using the techniques in this section.

Informal Exercise 24. Give short cuts for finding the remainder of the number $[100010453000001]_8$ when dividing by $7, 8$ or 9.

6.4 Even and odd integers

By Corollary 5, if $a \in \mathbb{Z}$ then exactly one of the following can occur:

$$a \equiv 0 \mod 2 \qquad \text{or} \qquad a \equiv 1 \mod 2.$$

Definition 2 (Even and odd)**.** If $a \equiv 0 \bmod 2$ then a is called an *even* integer. If $a \equiv 1 \bmod 2$ then a is called an *odd* integer.

Theorem 11. *An even integer plus an even integer is even. An odd integer plus an odd integer is even. An even integer plus an odd integer is odd. An even integer times an even integer is even. An odd integer times an odd integer is odd. An even integer times an odd integer is even.*

Proof. We consider the case of two odd integers. The other cases are left to the reader. If $a, b \in \mathbb{Z}$ are odd then by Theorem 8

$$a + b \equiv 1 + 1 \equiv 2 \equiv 0 \mod 2,$$

so $a + b$ is even. Also, by Theorem 10,

$$ab \equiv 1 \cdot 1 \equiv 1 \mod 2$$

so ab is odd. □

Exercise 25. Prove the other cases in the above theorem.

Informal Exercise 26. Explain why you only need to know the last digit of a positive integer in order to determine if the number is even or odd. Explain why this trick works in base 10, but not in base 5.

Informal Exercise 27. Use the results of Chapter 5 to explain why an even number $n \in \mathbb{N}$ that is also divisible by 3 is divisible by 6. If n is written in base 10, what is a quick procedure to check if it is divisible by 6? Using this procedure, check to see whether 2408, 2349, and 1554 are divisible by 6. Hint: Use the fact that 2 and 3 are relatively prime (why?).

6.5 The finite ring $\mathbb{Z}_m$.

Since $\equiv_m$ is an equivalence relation, we can consider the equivalence classes under this relation. The set of equivalence classes $\mathbb{Z}_m = \{[a]_m \mid a \in \mathbb{Z}\}$ will be shown to be a ring (after we define a suitable $+$ and $\cdot$).

At first one might think that this ring $\mathbb{Z}_m$ is infinite since, for each $a \in \mathbb{Z}$, we can form the equivalence class $[a]$. However, due to the properties of $\equiv_m$, the number of elements in $\mathbb{Z}_m$ is just m.

Definition 3. Fix a positive integer m, and consider the equivalence relation $\equiv_m$ defined above. If $a \in \mathbb{Z}$, then let $[a]$ denote the equivalence class containing a under this relation. In other words, $[a] = \{x \in \mathbb{Z} \mid x \equiv_m a\}$. Define

$$\mathbb{Z}_m = \big\{[a] \mid a \in \mathbb{Z}\big\}.$$

We call $\mathbb{Z}_m$ the set of *integers modulo* m. We often write $\overline{a}$ for $[a]$. We also write $[a]_m$ when we want to be clear about the modulus.

Informal Exercise 28. Describe the set $[5]$ if $m = 1$. Show that $[5] = [-1]$ in this case.

Informal Exercise 29. Describe the set $[5]$ if $m = 2$. Show that $[5]$ consists of the odd integers.

Informal Exercise 30. Describe the set $[5]$ if $m = 3$. Show that $[5] = [2]$.

Informal Exercise 31. Describe the sets $[0], [1]$, and $[2]$ if $m = 3$.

Theorem 12. *Let m be a positive integer and $a, b \in \mathbb{Z}$. Then*

$$a \equiv_m b \iff [a] = [b].$$

Proof. This is a general fact about equivalence classes (from set theory). □

Corollary 13. *Let m be a positive integer and $a, b \in \mathbb{Z}$. Then*

$$[a]_m = [b]_m \iff \text{Rem}(a, m) = \text{Rem}(b, m) \iff m \mid (a - b)$$

Proof. These conditions are all equivalent to $a \equiv_m b$ by earlier results. □

The following shows that when working in $\mathbb{Z}_m$ we can always limit ourselves to $[b]$ with $0 \leq b < m$.

Theorem 14. *Suppose $[a] \in \mathbb{Z}_m$ where m is a positive integer. Then there is exactly one $b \in \{0, \ldots, m-1\}$ such that $[a] = [b]$.*

Proof. Combine Corollary 5 with Theorem 12. □

Corollary 15. *Let m be a positive integer. The rule $x \mapsto [x]$ defines a bijection $f : \{0, \ldots, m-1\} \to \mathbb{Z}_m$.*

Proof. We show that f is injective and surjective. Suppose $f(x) = f(y)$ where $x, y \in \{0, \ldots, m-1\}$. Then $[x] = [y]$. So $x \equiv y$ modulo m by Theorem 12. Therefore, $x = y$ by Theorem 4.

Now we show f is surjective. Let $[a] \in \mathbb{Z}_m$ be an arbitrary element. We must find something in the domain that maps to $[a]$. By Theorem 14 there is a unique $b \in \{0, \ldots, m-1\}$ such that $[a] = [b]$. Thus $f(b) = [a]$. So f is surjective. □

Corollary 16. *Let m be a positive integer. The set $\mathbb{Z}_m$ has m elements.*

Proof. The above corollary gives a bijection $\{0, \ldots, m-1\} \to \mathbb{Z}_m$. However, $\{0, \ldots, m-1\}$ has m elements (Chapter 4). Thus $\mathbb{Z}_m$ has m elements (Chapter 3). □

Now we consider addition and multiplication in $\mathbb{Z}_m$.

Informal Exercise 32. Show that $[3] + [7] = [1]$ in $\mathbb{Z}_9$ using the following definition of addition (and Theorem 12). Hint: first show $[3] + [7] = [10]$.

Definition 4 (Addition)**.** Let m be a positive integer. Suppose $[a], [b] \in \mathbb{Z}_m$. Then $[a]+[b]$ is defined to be $[a+b]$ where addition inside [ ] is as in Chapter 4. This defines a binary operation

$$+ : \mathbb{Z}_m \times \mathbb{Z}_m \to \mathbb{Z}_m.$$

Since this definition involves equivalence classes, and since there are several ways to denote the same class, we need to show that the definition of addition is well-defined. This is done in the following lemma.

Lemma 17. *Let m be a positive integer. If $[a] = [a']$ and $[b] = [b']$ in $\mathbb{Z}_m$ then*

$$[a] + [b] = [a'] + [b'].$$

Proof. By Theorem 12, $a \equiv_m a'$ and $b \equiv_m b'$. Then $a + b \equiv_m a' + b'$ by Theorem 8. So, by Theorem 12, $[a + b] = [a' + b']$. □

Many of the properties of addition for $\mathbb{Z}$ also apply to $\mathbb{Z}_m$. For example, we prove the commutative law.

Theorem 18. *Let $[a], [b] \in \mathbb{Z}_m$ where m is a positive integer. Then*

$$[a] + [b] = [b] + [a].$$

Proof. Observe

$$\begin{aligned} [a] + [b] &= [a + b] && \text{(Def. of addition in } \mathbb{Z}_m) \\ &= [b + a] && \text{(Comm. Law for } + \text{ in } \mathbb{Z}\text{: Ch. 4)} \\ &= [b] + [a] && \text{(Def. of addition in } \mathbb{Z}_m). \end{aligned}$$

□

Exercise 33. Prove the associative law of addition for $\mathbb{Z}_m$.

Now we turn our attention to multiplication.

Informal Exercise 34. Show that $[3] \cdot [7] = [3]$ in $\mathbb{Z}_9$ using the following definition of multiplication (and Theorem 12 or Corollary 13).

Definition 5 (Multiplication)**.** Let m be a positive integer. Then given elements $[a], [b] \in \mathbb{Z}_m$, the product $[a] \cdot [b]$ is defined to be $[ab]$, where multiplication inside [ ] is as in Chapter 4. This defines a binary operation

$$\cdot : \mathbb{Z}_m \times \mathbb{Z}_m \to \mathbb{Z}_m.$$

As with the definition of addition on $\mathbb{Z}_m$, we need to show that this definition is well-defined. This is done in the following lemma.

Lemma 19. *Let m be a positive integer. If $[a] = [a']$ and $[b] = [b']$ in $\mathbb{Z}_m$ then*

$$[a] \cdot [b] = [a'] \cdot [b'].$$

Proof. Combine Theorem 12 with Theorem 10. □

Exercise 35. Let $[a], [b] \in \mathbb{Z}_m$ where m is a positive integer. Then show

$$[a] \cdot [b] = [b] \cdot [a].$$

Now we come to the key theorem of the section.

Theorem 20. *Let m be a positive integer. Then $\mathbb{Z}_m$ is a commutative ring with additive identity $[0]$ and multiplicative identity $[1]$. The additive inverse of $[a]$ is $[-a]$.*

Exercise 36. Prove the above theorem. Hint: some pieces have been done in earlier theorems and exercises. For example, if you want to show $[0]$ is the additive identity, you only need to show $[a]+[0]=[a]$ since $[0]+[a]=[a]+[0]$ from an earlier theorem.

Exercise 37. Show that the additive inverse of $[a] \in \mathbb{Z}_m$ is $[m-a]$. Show that $\overline{2}$ is the additive inverse of $\overline{3}$ in $\mathbb{Z}_5$.

Remark 1. Since the additive inverse of $[a]$ is $[-a]$ you can write

$$-[a] = [-a]$$

where the first $-$ signifies inverse in $\mathbb{Z}_m$, and the second $-$ signifies inverse in $\mathbb{Z}$.

Remark 2. Using the notation $\overline{a}$ for $[a]$ we can write the above definitions and results as

$$\overline{a+b} = \overline{a} + \overline{b} \qquad \overline{ab} = \overline{a}\,\overline{b} \qquad -\overline{a} = \overline{-a} = \overline{m-a}.$$

The additive identity is $\overline{0}$. The multiplicative identity is $\overline{1}$.

Remark 3. In any ring, 0 customarily denotes the additive identity. So we can write

$$0 = [0] = \overline{0}$$

where the left 0 is the identity in $\mathbb{Z}$ and the middle and right 0's are the identity in $\mathbb{Z}_m$. Likewise, 1 can denote the multiplicative identity in any ring:

$$1 = [1] = \overline{1}$$

where the left 1 is the identity in $\mathbb{Z}$ and the middle and right 1 is the identity in $\mathbb{Z}_m$.

Informal Exercise 38. Make addition and multiplication tables for $\mathbb{Z}_m$ for each of $m = 1, 2, 3, 4, 5, 6$. Your answers should be in the form $\overline{a}$ where a is such that $0 \le a < m$, but to save time you do not have to write bars over the answer: if you write '3', everyone will know that you actually mean $\overline{3}$. Hint: use the commutative law to save time.

Exercise 39. Suppose m is a positive integer, and that $m = ab$ where a and b are positive and less than m (in other words, suppose that m is composite). Show that $\mathbb{Z}_m$ is not an integral domain.

Later, we will show that $\mathbb{Z}_p$ is an integral domain if p is a prime number.

6.6 Units in a ring

Every element in a ring has an additive inverse, but only some elements have multiplicative inverses. Any element with a multiplicative inverse is called a *unit*. Recall that we assume all rings have a multiplicative identity.

Definition 6 (Units)**.** Let R be a ring with multiplicative identity 1. If a and b are elements of R are such that $ab = ba = 1$ then we say that a and b are multiplicative inverses. We write $b = a^{-1}$ and $a = b^{-1}$ to indicate that b is the inverse of a and a is the inverse of b. If R is commutative, we only need to check $ab = 1$.

An element $a \in R$ is called a *unit* if it has an inverse. The set of units is written $R^\times$:

$$R^\times \stackrel{\text{def}}{=} \{u \in R \mid u \text{ is a unit}\}.$$

Observe that $R^\times$ is a subset of R.

Warning. The above use of the superscript -1 is different than its use in iteration.

Informal Exercise 40. What are the units of $\mathbb{Z}$? In other words, what is $\mathbb{Z}^\times$?

Informal Exercise 41. Make a multiplication table for $\mathbb{Z}_9$. Use it to find $\mathbb{Z}_9^\times$. List all the inverses of all the units. To save time, you do not have to use bars or brackets in the tables. Hint: remember how to find remainders modulo 9, and use the fact that multiplication is commutative.

Informal Exercise 42. Are $\mathbb{Z}^\times$ and $\mathbb{Z}_9^\times$ closed under addition?

Exercise 43. Let a be an element of a ring R. Show that if a is a unit, then its multiplicative inverse is unique.

Exercise 44. Show that 1 and -1 are units in any ring R. Show that if R is a ring with $0 \neq 1$ then 0 is not a unit. (Most rings have $0 \neq 1$. The *trivial ring* is an exception: it has only one element so all elements are equal).

Exercise 45. Show that if u is a unit in a ring R then so is u^{-1}, and that

$$\left(u^{-1}\right)^{-1} = u.$$

Here is the main theorem of this section. It tells us which elements of $\mathbb{Z}_m$ are units.

Theorem 21. *Let $\overline{a} \in \mathbb{Z}_m$ where m is a positive integer. Then $\overline{a}$ is a unit if and only if a and m are relatively prime.*

Proof. Suppose that $\overline{a}$ is a unit. This means that there is a $b \in \mathbb{Z}$ such that $ab \equiv 1 \bmod m$. In other words, m divides $ab - 1$. This means we can write $ab - 1 = mc$ for some $c \in \mathbb{Z}$. So $ab - cm = 1$. Any common divisor of a and m must divide the linear combination $ab - cm = 1$ (Section 6 of Chapter 5). Thus the only positive common divisor of a and m is 1. This means that a and m are relatively prime.

Now suppose that a and m are relatively prime, and consider the function $f : \mathbb{Z}_m \to \mathbb{Z}_m$ defined by the rule $x \mapsto \overline{a}x$. We first show that f is injective. Suppose $f\left(\overline{b}\right) = f\left(\overline{c}\right)$. Then $\overline{a}\,\overline{b} = \overline{a}\,\overline{c}$. In other words, $ab \equiv_m ac$.

This means that m divides $ab-ac = a(b-c)$. Clearly a also divides $a(b-c)$. Since a and m are relatively prime, we have $am \mid a(b-c)$ (Section 7, Ch. 5). Thus $m \mid (b-c)$. This means that $b \equiv_m c$, so $\overline{b} = \overline{c}$.

We conclude that f is injective. Since f is injective, and maps a finite set to itself, it must also be surjective (Chapter 3). Thus there is an element $\overline{b}$ such that $f\left(\overline{b}\right) = \overline{1}$. By definition of f, we have $\overline{a}\,\overline{b} = \overline{1}$. So $\overline{a}$ is a unit. □

Informal Exercise 46. Use the above theorem to identify $\mathbb{Z}_m^\times$ for $m = 1$ to $m = 12$. Make multiplication tables for $\mathbb{Z}_m^\times$ for $m = 1, 2, 3, 4, 5, 7, 8, 10, 12$. (To save time, you do not have to use bars or brackets in the tables.)

As the tables from the above exercise show, the set $\mathbb{Z}_m$ is closed under multiplication:

Lemma 22. *If $a, b \in R^\times$ are units in a ring R, then ab is a unit. Furthermore, $(ab)^{-1} = b^{-1}a^{-1}$. If R is commutative then $(ab)^{-1} = a^{-1}b^{-1}$*

Proof. (sketch) First show $b^{-1}a^{-1}$ is the inverse of ab. So the inverse of ab exists. In other words, ab is a unit. □

This lemma implies that for $R^\times$ multiplication gives a binary operation

$$R^\times \times R^\times \to R^\times.$$

Multiplication is associative since R is a ring, and $R^\times$ is a subset of R. Since 1 is a unit in any ring, there is an identity for this operation. Furthermore, if $u \in R^\times$ then clearly $u^{-1} \in R^\times$ (see Exercise 45). Thus every element of $R^\times$ has an inverse in $R^\times$. Thus we get the following:

Theorem 23. *If R is a ring, then the units $R^\times$ form a group under multiplication. If R is a commutative ring, then $R^\times$ is an abelian group.*

6.7 The finite field $\mathbb{F}_p$

In this section we will see that every non-zero element of $\mathbb{Z}_p$ is a unit when p is a prime. Commutative rings with this property are very important, and are called *fields*. Because $\mathbb{Z}_p$ is a field we sometimes write $\mathbb{F}_p$ for $\mathbb{Z}_p$. Every field turns out to be an integral domain, so $\mathbb{Z}_p$ is also an integral domain. On the other hand, we saw above that $\mathbb{Z}_m$ is not an integral domain if m is composite.

Theorem 24. *If p is a prime, then every non-zero element of $\mathbb{Z}_p$ is a unit.*

Proof. Let $\overline{a} \in \mathbb{Z}_p$ be non-zero. Observe that a is not a multiple of p (otherwise $a \equiv_p 0$, a contradiction). Since $p \nmid a$ and since p is a prime, a and p are relatively prime (Chapter 5). By Theorem 21, $\overline{a}$ is a unit. □

Definition 7. A *field* is a commutative ring F such that (i) $0 \neq 1$, and (ii) every non-zero element is a unit.

Remark 4. The conditions (i) and (ii) in the above definition can be folded into one condition: $x \in F$ is a unit if and only if $x \neq 0$.

Remark 5. Fields are extremely important in mathematics. The number systems $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all fields. In a field you can make use of all four basic algebraic operations $+, -, \times, \div$, with the only restriction that you cannot divide by zero since zero is not a unit.

Theorem 25. *If p is a prime then $\mathbb{Z}_p$ is a field.*

Proof. Observe that (i) $\overline{0} \neq \overline{1}$ since $p > 1$. In addition, (ii) every non-zero element of $\mathbb{Z}_p$ is a unit by Theorem 24. □

Definition 8. If p is a prime, then $\mathbb{F}_p$ is another name for $\mathbb{Z}_p$. The field $\mathbb{F}_p$ is an example of a finite field.

Remark 6. Every field is an integral domain, but not all integral domains are fields. We leave the verification of these facts to the reader.

6.8 Exponentiation in a ring

In Chapter 1 we used the idea of repeated multiplication to define exponentiation. This idea can be extended to any ring R. For units, we can extend this idea and define exponentiation for negative exponents. Exponentiation in a general ring satisfies familiar rules such as $a^{m+n} = a^m a^n$. We have seen similar rules in the context of iteration. In fact, we will use the rules for iteration to prove the analogous rules for exponentiation.

In this section we will consider exponentiation for general elements in a ring. In the next section we will consider exponentiation of units.

Definition 9 (Exponentiation in rings)**.** Suppose R is a ring and $a \in R$. Let $M_a : R \to R$ be defined by the rule $x \mapsto xa$. If $n \in \mathbb{N}$ then

$$a^n \stackrel{\text{def}}{=} M_a^n(1).$$

Our strategy for studying exponentiation is to find identities between M_a. This approach is very similar to that used in Chapter 4 when we studied the translation function A_a. In what follows, let R be a ring.

Lemma 26. *Let $a \in R$. Then $M_a^0 = id$ and $M_a^1 = M_a$.*

Proof. These are basic properties of iteration (Chapter 1). □

Corollary 27. *Let $a \in R$. Then $a^0 = 1$ and $a^1 = a$.*

Proof. Apply M_a^0 and M_a^1 to 1. The above lemma gives the result. □

Lemma 28. *Let $a, b \in R$. Then $M_a \circ M_b = M_{ba}$.*

Proof. For any $x \in R$,

$$M_a \circ M_b(x) = M_a\big(M_b(x)\big) = M_a(xb) = (xb)a = x(ba) = M_{ba}(x).$$

The conclusion follows. □

Lemma 29. *Let $a, b \in R$. Suppose $ab = ba$ (which holds, for example, if R is a commutative ring). Then M_a and M_b commute.*

Proof. This follows from the above lemma since $M_{ab} = M_{ba}$. □

Lemma 30. *Consider the iteration M_a^n where $a \in R$ and $n \in \mathbb{N}$. Then there is an element $c \in R$ such that $M_a^n = M_c$.*

Proof. Fix $a \in R$. Let S be the set of $n \in \mathbb{N}$ such that the conclusion holds. Observe that $0 \in S_a$ since $M_a^0 = id = M_1$.

Suppose that $n \in S_a$. This means $M_a^n = M_b$ for some $b \in R$. Thus

$$M_a^{n+1} = M_a \circ M_a^n = M_a \circ M_b = M_{ba}. \qquad \text{(Lemma 28)}$$

Thus $n + 1 \in S_a$.

By induction, $S_a = \mathbb{N}$. The result follows. □

The following is very useful for deriving identities.

Lemma 31. *If $n \in \mathbb{N}$ and $a \in R$ then $M_a^n = M_{a^n}$.*

Proof. By Lemma 30, $M_a^n = M_c$ for some $c \in R$. Apply M_a^n and M_c to 1:

$$a^n \stackrel{\text{def}}{=} M_a^n(1) = M_c(1) = 1 \cdot c = c.$$

Thus $c = a^n$. But, $M_a^n = M_c$. So $M_a^n = M_{a^n}$ □

Lemma 32. *Let $a, b \in R$. If $M_a = M_b$ then $a = b$.*

Proof. Suppose $M_a = M_b$. Then $M_a(1) = M_b(1)$. But $M_a(1) = 1 \cdot a = a$ and $M_b(1) = 1 \cdot b = b$. □

We know from Chapter 3 that iteration satisfies the law $f^{m+n} = f^m \circ f^n$ for $m, n \in \mathbb{N}$. We use this fact in the following.

Theorem 33. *Let $a \in R$ where R is a ring, and let $m, n \in \mathbb{N}$. Then*

$$a^{m+n} = a^m a^n.$$

Proof. Observe that

$$\begin{aligned} M_{a^{m+n}} &= M_a^{m+n} && \text{(Lemma 31)} \\ &= M_a^{n+m} && \text{(Comm. Law, Ch. 1)} \\ &= M_a^n \circ M_a^m && (f^{m+n} = f^m \circ f^n, \text{ Ch. 3}) \\ &= M_{a^n} \circ M_{a^m} && \text{(Lemma 31)} \\ &= M_{a^m a^n} && \text{(Lemma 28)} \end{aligned}$$

Thus $a^{m+n} = a^m a^n$ by Lemma 32. □

Theorem 34. *Consider $0 \in R$. If n is a positive integer then*

$$0^n = 0.$$

Proof. Write n as $m + 1$. Then, using the previous theorem,

$$0^n = 0^{m+1} = 0^m 0^1 = 0^m 0 = 0.$$

□

We know from Chapter 3 that iteration of functions satisfies the law $(f^m)^n = f^{mn}$ for all $m, n \in \mathbb{N}$. We use this fact in the following.

Theorem 35. *Let $a \in R$ and $m, n \in \mathbb{N}$. Then*

$$(a^m)^n = a^{mn}.$$

Proof. Start with $M_{a^m} = M_a^m$ (Lemma 31). Then

$$\begin{aligned} (M_{a^m})^n &= \left(M_a^m\right)^n && \text{(Iterate same funct.)} \\ &= M_a^{mn} && \text{(Ch. 3: } (f^m)^n = f^{mn}) \end{aligned}$$

Apply both sides to 1:

$$(a^m)^n \stackrel{\text{def}}{=} (M_{a^m})^n(1) = M_a^{mn}(1) \stackrel{\text{def}}{=} a^{mn}.$$

□

We know from Chapter 4 that if $f : S \to S$ and $g : S \to S$ commute (for some set S) then we have $(f \circ g)^n = f^n \circ g^n$ for all $n \in \mathbb{N}$. We use this fact in the following.

Theorem 36. *Let $a, b \in R$ and $n \in \mathbb{N}$. Suppose $ab = ba$ (which is true, for example, if R is a commutative ring). Then*

$$(ab)^n = a^n b^n.$$

Proof. By Lemma 29, M_a and M_b commute. So

$$\begin{aligned}\big(M_{ab}\big)^n &= \big(M_b \circ M_a\big)^n \qquad \text{(Lemma 28)}\\ &= M_b^n \circ M_a^n \qquad \text{(Ch. 4: } (f\circ g)^n = f^n \circ g^n\text{)}\\ &= M_{b^n} \circ M_{a^n} \qquad \text{(Lemma 31)}\\ &= M_{a^n b^n} \qquad \text{(Lemma 28)}\end{aligned}$$

Apply both sides to 1:

$$(ab)^n \stackrel{\text{def}}{=} \big(M_{ab}\big)^n(1) = M_{a^n b^n}(1) = a^n b^n.$$

□

Exponentiation can also be thought of in terms of finite products:

Theorem 37. *Let $a \in R$ where R is a ring. If n is a positive integer then*

$$a^n = \prod_{i=1}^{n} a_i$$

where $(a_i)_{i=1,\ldots,n}$ is the constant sequence defined by the rule $a_i = a$ for all $i \in \{1, \ldots, n\}$.

Proof. (Sketch) This can be proved by induction. The details are left to the reader. □

Remark 7. If $a \in \mathbb{Z}$ then the definition of a^n given in this section agrees with the definition of Chapter 1 whenever $a \geq 0$. To see this, compare the two definitions and observe that $\mu_a(x) = M_a(x)$ for all $x \geq 0$. By induction, it follows that $\mu_a^n = M_a^n$ for all $n \in \mathbb{N}$. In particular, $\mu_a^n(1) = M_a^n(1)$.

6.9 Exponentiation of units in a ring

If a is a unit in a ring R, then we can define a^u for negative $u \in \mathbb{Z}$. In what follows let R be a ring.

Lemma 38. *If $a \in R$ is a unit then M_a has an inverse function and*

$$(M_a)^{-1} = M_{a^{-1}}.$$

Proof. We know $M_a \circ M_{a^{-1}} = M_1$ by Lemma 28. But $M_1 = id$ by definition of M_a. Likewise $M_{a^{-1}} \circ M_a = id$. The conclusion follows. □

Corollary 39. *If $a \in R$ is a unit then M_a is a bijection, and M_a^u is defined for all $u \in \mathbb{Z}$.*

Now we can define a^u for all $u \in \mathbb{Z}$, even for negative u.

Definition 10. Suppose $a \in R$ is a unit and $u \in \mathbb{Z}$. Then

$$a^u \stackrel{\text{def}}{=} M_a^u(1).$$

Recall from Chapter 4 that if $f : S \to S$ is the identity, then $f^u = id$ for all $u \in \mathbb{Z}$. This is used below.

Theorem 40. *Let* 1 *be the multiplicative identity of a ring* R*. If* $u \in \mathbb{Z}$ *then*

$$1^u = 1.$$

Proof. Observe that M_1 is the identity function. Then M_1^u is the identity function (Chapter 4). Thus $1^u = M_1^u(1) = id(1) = 1$. □

Informal Exercise 47. Find a^{-2} for all non-zero $a \in \mathbb{F}_7$. How do these compare to a^4?

The following generalizes Lemma 30.

Lemma 41. *Consider the iteration* M_a^u *where* $a \in R$ *and* $u \in \mathbb{Z}$*. Then there is an element* $c \in R$ *such that* $M_a^u = M_c$*.*

Proof. If $u \geq 0$ this follows from Lemma 41.

Now assume $u < 0$, so $u = -n$ for some $n \in \mathbb{N}$. Then

$$\begin{array}{rcll} M_a^u & = & M_a^{-n} & \text{(Rewrite)} \\ & = & \left(M_a^{-1}\right)^n & \text{(Ch. 4: } f^{uv} = (f^u)^v\text{)} \\ & = & \left(M_{a^{-1}}\right)^n & \text{(Lemma 38)} \\ & = & M_{a^{-1}}^n & \text{(Rewrite)} \\ & = & M_{(a^{-1})^n} & \text{(Lemma 31)} \end{array}$$

Thus $M_a^u = M_c$ where $c = \left(a^{-1}\right)^n$. □

The following generalizes Lemma 31.

Lemma 42. *Let* $a \in R$ *be a unit. If* $u \in \mathbb{Z}$ *then* $M_a^u = M_{a^u}$*.*

Proof. (Sketch) This is similar to the proof of Lemma 31, but adapted from the natural numbers to the integers. □

The following generalizes Theorem 33.

Theorem 43. *Suppose* $a \in R$ *is a unit and that* $u, v \in \mathbb{Z}$*. Then*

$$a^{u+v} = a^u a^v.$$

Proof. This is similar to the proof of Theorem 33, but adapted from the natural numbers to the integers. □

The following generalizes Theorem 35.

Theorem 44. *Suppose $a \in R$ is a unit and that $u, v \in \mathbb{Z}$. Then*

$$(a^u)^v = a^{uv}.$$

Proof. This is similar to the proof of Theorem 35, but adapted from the natural numbers to the integers. □

The following generalizes Theorem 36.

Theorem 45. *Let $a, b \in R$ be units, and let $u \in \mathbb{Z}$. If $ab = ba$ then*

$$(ab)^u = a^u b^u.$$

Proof. This is similar to the proof of Theorem 36, but adapted from the natural numbers to the integers. □

Theorem 46. *If $a \in R$ is a unit and $u \in \mathbb{Z}$ then a^u is also a unit.*

Proof. By Theorem 43, $a^u \cdot a^{-u} = a^0$ and $a^{-u} \cdot a^u = a^0$. However, by Corollary 27, $a^0 = 1$. The result follows. □

Note. This theory of exponentiation with positive and negative exponents can be developed for groups as well. There is both an additive and a multiplicative version that differ only with the notation used.

Chapter 7

The Rational Numbers $\mathbb{Q}$

In this chapter we study the field of rational numbers $\mathbb{Q}$. We construct the set of rational numbers $\mathbb{Q}$ using equivalence classes $[(x, y)]$ of pairs of elements $x, y, \in \mathbb{Z}$, and we usually write $[x, y]$ for the class $[(x, y)]$. Informally, the coordinates x and y can be thought of as the numerator and denominators of a fraction. We define an addition and a multiplication operation. The first main result of this chapter is that this construction results in a field. We use a canonical embedding $\mathbb{Z} \to \mathbb{Q}$ to view $\mathbb{Z}$ as a subset of the field $\mathbb{Q}$. After this we prove some basic facts about rational numbers.

Although the definitions for addition and multiplication for $\mathbb{Q}$ are inspired by our prior experience with fractions, we do not use such experience with fractions in our proofs. Instead the proofs are based on previous results developed for $\mathbb{Z}$ in earlier chapters. At first we write $[x, y]$ for elements of $\mathbb{Q}$. Later in the chapter we develop the notation x/y for fields in general, not just $\mathbb{Q}$, and prove the main laws that govern the use of fractions. After this we can and will use the notation of fractions in our formal development.

7.1 Basic definitions

We need to define $\mathbb{Q}$ and the operations of addition and multiplication on this set. Before defining $\mathbb{Q}$ we define a related set Q which informally represents quotients of integers. The difference between Q and $\mathbb{Q}$ is that the latter consists of equivalence classes of the former.

Definition 1. Let $Q = \big\{(x, y) \mid x, y \in \mathbb{Z} \text{ and } y \neq 0\big\}$. The first element x of a pair in Q is called the *numerator*, and the second element y is called the *denominator*. Elements of Q are called *numerator-denominator pairs*.

Note. It might be tempting to write (x, y) symbolically as x/y. We do not do so because we will define x/y later when considering division in any field, so using x/y now could create confusion. Also, using the notation x/y might lead one to unintentionally use properties about fractions that have not yet been formally proved.

Informal Exercise 1. Using your prior informal knowledge of fractions as inspiration, how would you add and multiply (x, y) and (w, z)? How would you decide if (x, y) and (w, z) are equivalent?

We now formalize the concepts in the above exercise. First we define rational equivalence.

Definition 2. We say that two elements (x, y) and (z, w) of Q are *rationally equivalent* if $xw = yz$. In this case, we write $(x, y) \sim (z, w)$.

Theorem 1. *The relation $\sim$ is an equivalence relation on the set Q.*

Exercise 2. Prove the above theorem. Hint: the reflexive and symmetry laws involve using the commutative law for $\mathbb{Z}$. The transitivity law requires multiplying an equation by a constant, performing basic manipulations, and using the cancellation law (for a non-zero element of $\mathbb{Z}$).

Definition 3. If $(x, y) \in Q$ then $[(x, y)]$, or just $[x, y]$, denotes the equivalence class containing (x, y) under the above equivalence relation.

Definition 4. Define $\mathbb{Q}$ as follows

$$\mathbb{Q} \stackrel{\text{def}}{=} \big\{[x, y] \mid (x, y) \in Q\big\}.$$

In other words,

$$\mathbb{Q} = \big\{[x, y] \mid x, y \in \mathbb{Z} \text{ with non-zero } y\big\}.$$

Exercise 3. Prove the following two theorems.

Theorem 2. *Let $(x, y) \in Q$. Then $[x, y] = [0, 1]$ if and only if $x = 0$.*

Theorem 3. *Let $(x, y) \in Q$. Then $[x, y] = [1, 1]$ if and only if $x = y$.*

Definition 5. Define two binary operations, *addition* and *multiplication*, for $\mathbb{Q}$ as follows. Let $[x, y]$ and $[z, w]$ be elements of $\mathbb{Q}$. Then

$$[x, y] + [z, w] \stackrel{\text{def}}{=} [xw + yz,\ yw]$$

and

$$[x, y] \cdot [z, w] \stackrel{\text{def}}{=} [xz,\ yw].$$

Since these definitions involve equivalence classes, we must show that they are well-defined. This is the purpose of the following lemmas.

Lemma 4. *Addition on $\mathbb{Q}$ is well-defined.*

Proof. We want to show that if $[x, y] = [x', y']$ and $[z, w] = [z', w']$ then

$$[xw + yz,\ yw] = [x'w' + y'z',\ y'w'].$$

In other words, suppose $(x, y) \sim (x', y')$ and $(z, w) \sim (z', w')$. We must show $(xw + yz, yw) \sim (x'w' + y'z', y'w')$. By definition of $\sim$, we need to show that

$$(xw + yz)(y'w') = (x'w' + y'z')(yw).$$

This can be shown as follows:

$$\begin{aligned}
(xw + yz)(y'w') &= (xw)(y'w') + (yz)(y'w') && \text{(Distr. Law of Ch. 4)}\\
&= (xy')(ww') + (zw')(yy') && \text{(Laws of Ch. 4)}\\
&= (yx')(ww') + (zw')(yy') && \text{(Since } (x, y) \sim (x', y'))\\
&= (yx')(ww') + (wz')(yy') && \text{(Since } (z, w) \sim (z', w'))\\
&= (x'w')(yw) + (y'z')(yw) && \text{(Laws of Ch. 4)}\\
&= (x'w' + y'z')(yw) && \text{(Distr. Law of Ch. 4).}
\end{aligned}$$

□

Lemma 5. *Multiplication on $\mathbb{Q}$ is well-defined.*

Exercise 4. Prove the above lemma.

Exercise 5. Prove the following two theorems.

Theorem 6. *For any $[x, y] \in \mathbb{Q}$,*

$$[x, y] + [0, 1] = [x, y]$$

and

$$[x, y] \cdot [1, 1] = [x, y].$$

Theorem 7. *Suppose $x, y, c \in \mathbb{Z}$ where $y \neq 0$ and $c \neq 0$. Then $(cx, cy) \in Q$ and*

$$[cx, cy] = [x, y].$$

Exercise 6. Prove the following three theorems.

Theorem 8. *Addition in $\mathbb{Q}$ is commutative.*

Theorem 9. *Multiplication in $\mathbb{Q}$ is commutative.*

Theorem 10. *Suppose x, y are non-zero integers. Then*

$$[x, y] \cdot [y, x] = [1, 1].$$

7.2 The field $\mathbb{Q}$

Theorem 11. *Using the above addition and multiplication, the set $\mathbb{Q}$ is a field. The additive identity is $[0,1]$ and the multiplicative identity is $[1,1]$. Suppose $x, y \in \mathbb{Z}$ with $y \neq 0$. Then the additive inverse of $[x,y]$ is $[-x,y]$. If $x \neq 0$ and $y \neq 0$ then the multiplicative inverse of $[x,y]$ is $[y,x]$.*

Exercise 7. Prove the above theorem. Hint: several parts have been proved above.

7.3 The canonical embedding of $\mathbb{Z}$ in $\mathbb{Q}$

Definition 6. The *canonical embedding* $\mathbb{Z} \to \mathbb{Q}$ is the function defined by the rule

$$a \mapsto [a,1].$$

Theorem 12. *The canonical embedding $\mathbb{Z} \to \mathbb{Q}$ is injective.*

Exercise 8. Prove the above theorem.

Exercise 9. Show that $[2,2]$ is in the image of the canonical embedding, but $[1,2]$ is not in the image of the canonical embedding. Conclude that the canonical embedding is not surjective. Hint: suppose $[1,2] = [a,1]$. Derive a contradiction from $(1,2) \sim (a,1)$.

If we identify $a \in \mathbb{Z}$ with its image in $\mathbb{Q}$, then we can think of $\mathbb{Z}$ as a subset of $\mathbb{Q}$. So from now on, if $a \in \mathbb{Z}$, we will think of a and $[a,1]$ as being the same element of $\mathbb{Q}$. By Theorem 7 we have that $[ca,c]$ and a are considered as the same element of $\mathbb{Q}$ (if c is non-zero), so a pair does not have to have the second number 1 to be thought of as an integer. For example, $[3,3]$ is an integer.

By this convention, $0 \in \mathbb{Z}$ is identified with $[0,1]$, and $1 \in \mathbb{Z}$ is identified with its image $[1,1]$. By Theorem 11, $0 = [0,1]$ is the additive identity and $1 = [1,1]$ is the multiplicative identity as expected.

Since we now think of $\mathbb{Z}$ as a subset of $\mathbb{Q}$ we have to be careful with $+$ and $\cdot$ in $\mathbb{Z}$. We defined these operations for $\mathbb{Z}$ in one way in Chapter 4, and then defined them for $\mathbb{Q}$ in the current chapter. Do we get the same answer for integers $a, b \in \mathbb{Z}$ as for the corresponding elements $[a,1]$ and $[b,1]$ in $\mathbb{Q}$? The answer is yes since

$$[a,1] + [b,1] = [a+b,\ 1] \qquad \textit{and} \qquad [a,1] \cdot [b,1] = [a \cdot b, 1 \cdot 1] = [ab,1].$$

Likewise for additive inverse: by Theorem 11

$$-[a,1] = [-a,1].$$

This equality shows that if a is identified with $[a,1]$, then the definitions of additive inverse, either as an integer or as a fraction, gives the same result. We summarize the above observations as follows.

Theorem 13. *Consider $\mathbb{Z}$ as a subset of $\mathbb{Q}$. Then the addition, multiplication, and additive inverse operators on $\mathbb{Q}$ extend the corresponding operators on $\mathbb{Z}$.*

Remark 1. Since subtraction (in any ring) is defined in terms of addition and additive inverse, the above theorem tells us that the subtraction of $\mathbb{Q}$ extends that of $\mathbb{Z}$.

Note. In this chapter we have constructed $\mathbb{Q}$ from $\mathbb{Z}$ using equivalence classes. Everything we have done works if $\mathbb{Z}$ is replaced with an arbitrary integral domain. In other words, if R is an integral domain, the above techniques can be used to construct a field F and a canonical embedding of R into F. We can think of F as a field containing R. The field F is called the *field of fractions of R*. Thus $\mathbb{Q}$ is the field of fractions of $\mathbb{Z}$.

If R is already a field, then the canonical embedding can be shown to be a bijection.

Optional Exercise 10. Verify that the construction and the main results concerning the construction can really be extended from $\mathbb{Z}$ to a general integral domain R. Why does R have to be an integral domain? In other words, what goes wrong if R is a more general ring?

Optional Exercise 11. Show that if R is a field, then the canonical embedding $R \to F$ is a bijection where F is the field of fractions of F. In other words, if R is already a field then its field of fractions is, in some sense, just itself.

7.4 Division and fractional notation in fields

Before studying the field $\mathbb{Q}$ in more detail, it is helpful to have the concept of division and to set up fractional notation. These concepts are valid in any field F, not just $\mathbb{Q}$.

Definition 7. Suppose $x, y \in F$ where F is a field and where $y \neq 0$. Then x/y is defined to be $x \cdot y^{-1}$. We also write this as $\frac{x}{y}$.

Remark 2. The rule $(x, y) \mapsto x/y$ defines a function $F \times F^{\times} \to F$. This is almost, but not quite, a binary operation. If fails to be a binary operation due to the fact that its domain is not all of $F \times F$. We call this "almost binary" operation the *division operation*.

Observe that a field has all four traditional arithmetic operations: addition, subtraction, multiplication, and division.

Most of the familiar identities and laws concerning fractions and division are valid for general fields, and can be easily proved using the identity $(xy)^{-1} = x^{-1}y^{-1}$, an identity that is valid in any field. In fact, we proved this identity for units in any commutative ring (see Chapter 6). Here are some examples of fraction identities.

Theorem 14. *Suppose that $x \in F$ and $y, z \in F^\times$ where F is a field. Then*

$$\frac{zx}{zy} = \frac{x}{y}.$$

Proof. Observe that

$$\begin{aligned}
(zx)/(zy) &= (zx)(zy)^{-1} && \text{(Def. 7)} \\
&= (zx)\left(z^{-1}y^{-1}\right) && \text{(Inverse Law for fields)} \\
&= \left(xy^{-1}\right)\left(zz^{-1}\right) && \text{(Comm/Assoc. Laws for fields)} \\
&= \left(xy^{-1}\right)\cdot 1 = x/y && \text{(Def. of inverse, Def. 7).}
\end{aligned}$$

□

When you have a common denominator, the formula for addition is very simple. This is just a special case of the distributive law.

Theorem 15. *Suppose that $x, y, z \in F$ where F is a field and $y \neq 0$. Then*

$$\frac{x}{y} + \frac{z}{y} = \frac{x+z}{y}.$$

Proof. Observe that

$$\begin{aligned}
x/y + z/y &= xy^{-1} + zy^{-1} && \text{(Def. 7)} \\
&= (x+z)y^{-1} && \text{(Distr. Law for rings)} \\
&= (x+z)/y && \text{(Def. 7).}
\end{aligned}$$

□

Theorem 16. *Suppose $x, z \in F$ and $y, w \in F^\times$ where F is a field. Then*

$$\frac{x}{y} + \frac{z}{w} = \frac{xw+yz}{yw}.$$

Proof. Observe that

$$\begin{aligned}
(xw+yz)/(yw) &= (xw+yz)(yw)^{-1} && \text{(Def. 7)} \\
&= (xw)(yw)^{-1} + (yz)(yw)^{-1} && \text{(Distr. Law for rings)} \\
&= (xw)/(yw) + (yz)/(yw) && \text{(Def. 7)} \\
&= x/y + z/w && \text{(Thm. 14).}
\end{aligned}$$

□

Exercise 12. Let $x, z \in F$ and $y, w \in F^\times$ where F is a field. Prove the following

$$\frac{x}{y}\cdot\frac{z}{w} = \frac{xz}{yw}, \qquad \frac{0}{y} = 0, \qquad \frac{y}{y} = 1,$$

$$\frac{x}{1} = x, \qquad x\frac{z}{y} = \frac{xz}{y}, \qquad y\frac{x}{y} = x.$$

Exercise 13. Let $x, y \in F$ where F is a field and y is not zero. Then show that x/y and $(-x)/y$ are additive inverses. Conclude that

$$-\frac{x}{y} = \frac{-x}{y} \qquad \text{and} \qquad -\frac{-x}{y} = \frac{x}{y}.$$

Exercise 14. Let $x, y \in F^\times$ where F is a field. Show that x/y and y/x are multiplicative inverses. Conclude that

$$\frac{1}{x/y} = \frac{y}{x}.$$

Theorem 17. *Let $x, z \in F$ and $y, w \in F^\times$ where F is a field. Then*

$$\frac{x}{y} = \frac{z}{w} \iff xw = yz,$$

and

$$\frac{x}{y} = \frac{z}{y} \iff x = z.$$

Proof. Multiply both sides of each equation by the appropriate constant. □

7.5 Representing elements of $\mathbb{Q}$

The notation and properties from the previous section apply to any field. We will now return to the study of $\mathbb{Q}$, but will use the fractional notation whenever possible. Of course, $[x, y]$ can be written as the fraction x/y:

Theorem 18. *Let $x, y \in \mathbb{Z}$ where $y \neq 0$. Think of $\mathbb{Z}$ as a subset of $\mathbb{Q}$ via the canonical embedding. Then*

$$[x, y] \;=\; \frac{x}{y}.$$

Proof. By definition of multiplication in $\mathbb{Q}$, we have $[x, y] = [x, 1] \cdot [1, y]$. However, $[1, y] = [y, 1]^{-1}$. Thus

$$[x, y] \;=\; [x, 1] \cdot [y, 1]^{-1}.$$

We identify x with $[x, 1]$ and y with $[y, 1]$. Therefore,

$$[x, y] \;=\; x \cdot y^{-1} = x/y.$$

□

Corollary 19. *Think of $\mathbb{Z}$ as a subset of $\mathbb{Q}$ via the canonical embedding. Then*

$$\mathbb{Q} = \left\{ \frac{a}{b} \;\middle|\; a, b \in \mathbb{Z} \text{ and } b \neq 0 \right\}.$$

Remark 3. Another consequence of the above theorem is that the canonical embedding is given by the law $a \mapsto a/1$.

In $\mathbb{Q}$ we can be picky and insist that the denominator be positive:

Theorem 20. *If $r \in \mathbb{Q}$ then there are integers a, b such that*

$$r = \frac{a}{b}$$

and such that b is positive. In particular,

$$\mathbb{Q} = \left\{ \frac{a}{b} \;\middle|\; a, b \in \mathbb{Z} \text{ and } b > 0 \right\}.$$

Exercise 15. Prove the above theorem. Hint: use Theorem 14 if necessary.

Now we show that we can be even pickier and insist that a and b have no factor in common greater than one:

Lemma 21. *If $r \in \mathbb{Q}$ then there are relatively prime integers $a, b \in \mathbb{Z}$ such that $b > 0$ and such that $r = a/b$.*

Proof. By Theorem 20 there are $a', b' \in \mathbb{Z}$ such that $b' > 0$ and $r = a'/b'$. This theorem does not guarantee that a', b' are relatively prime, so let g be the GCD of a' and b'. Thus a' and b' are multiples of g, so we can find $a, b \in \mathbb{Z}$ such that $a' = ga$ and $b' = gb$. Since g and b' are postive, b must also be positive. By Theorem 14, $r = a/b$.

Are a and b relatively prime? Let d be the GCD of a and b. We must show $d = 1$. Since $d \mid a$ we have $dg \mid ag$. In other words, $dg \mid a'$. Likewise, $dg \mid b'$. Thus dg is a common divisor of a' and b', but g is the greatest such common divisor. So $dg \leq g$. This implies that $d \leq 1$, which implies that $d = 1$, since d and g are both positive. Thus a and b are relatively prime. □

Theorem 22. *If $r \in \mathbb{Q}$ then there is a unique pair a, b of relatively prime integers such that $b > 0$ and*

$$r = \frac{a}{b}.$$

Proof. The existence is established by the previous lemma, so we focus on uniqueness. If $r = 0$, then $a = 0$ and $b = 1$ is the unique pair that works (if $b > 0$ and $a = 0$ then the GCD of b and a is just b). So assume $a \neq 0$, and suppose c, d is another such pair. Since $a/b = c/d$ we get $ad = bc$. Thus $b \mid ad$. Of course, $a \mid ad$. By a theorem of Chapter 5, $ab \mid ad$ since a and b are relatively prime. Thus $b \mid d$. A similar argument shows that $d \mid b$. By a result of Chapter 5, $|b| = |d|$. Since b and d are positive, $b = d$. This in turn implies that $a = c$. □

If we do not insist on the relatively prime condition, we can always find a common denominator for any two elements of $\mathbb{Q}$:

Theorem 23. *If $u, v \in \mathbb{Q}$ then we can find integers a, b, d with $d > 0$ such that*

$$u = \frac{a}{d} \quad \text{and} \quad v = \frac{b}{d}.$$

Exercise 16. Prove the above theorem.

Division is related to divisibility:

Theorem 24. *Let $a, b \in \mathbb{Z}$ with $b \neq 0$. Then $a/b \in \mathbb{Z}$ if and only if $b \mid a$.*

Proof. If $a/b \in \mathbb{Z}$ then $ab^{-1} = c$ for some $c \in \mathbb{Z}$. Multiply both sides by b. Thus $a = bc$. In other words, $b \mid a$.

If $b \mid a$ then $a = bc$ for some $c \in \mathbb{Z}$. Multiply both sides by b^{-1}. □

Remark 4. If $a, b \in \mathbb{Z}$ are such that $b \mid a$ and $b \neq 0$, then a/b was defined in Chapter 5 as the unique integer c such that $bc = a$. By multiplying $bc = a$ by b^{-1} we see that $c = ab^{-1}$. Thus the current (Chapter 7) definition of division is equivalent to the definition of Chapter 5.

7.6 Positive and negative rational numbers

The set $\mathbb{Q}$ is not just a field, but is an *ordered field*. We will define the notion of ordered field in the following chapter, but a key part of the definition is the idea of a positive subset. In this section we define the subset of positive rational numbers.

Definition 8 (Positive and Negative)**.** A number $r \in \mathbb{Q}$ is said to be a *positive* rational number if it can be written as a/b where a and b are positive integers. A number $r \in \mathbb{Q}$ is said to be a *negative* rational number if $-r$ is positive.

Remark 5. We already know, from Chapter 4, what positive and negative integers are. The above extends the definitions to rational numbers. Lemma 28 below shows that the new definitions truly extend the old definitions.

Theorem 25. *The set of positive rational numbers is closed under addition and multiplication: if $u, v \in \mathbb{Q}$ are positive, then so are $u + v$ and uv.*

Exercise 17. Prove the above. Hint: write $u = a/b$ and $v = c/d$ where a, b, c, d are positive integers. Use properties of positive integers.

Theorem 26. *Let $a/b \in \mathbb{Q}$ where $a, b \in \mathbb{Z}$ with $b \neq 0$. Then a/b is a positive rational number if and only if either (i) both a and b are positive integers or (ii) both a and b are negative integers.*

Proof. First suppose that a/b is a positive rational number. We will show that either both a and b are positive integers, or that a and b are both negative integers. By Definition 8, $a/b = c/d$ for some positive integers c and d. Thus $ad = bc$. Now we consider cases. First suppose that b is a positive integer. Then bc is a positive integer (Chapter 4). Thus ad is a positive integer. Since d is a positive integer, we must also have that a is a positive integer.

In the second case, suppose that b is a negative integer. Then bc is a negative integer (Chapter 4). Thus ad is a negative integer. Since d is a positive integer, we must also have that a is a negative integer.

Now we prove the converse. If a and b are positive integers, the result follows from Definition 8. If a and b are negative integers, then

$$a/b = (-a)/(-b)$$

by Theorem 14. Now use Definition 8 with $-a$ and $-b$. □

Theorem 27. *Let $a/b \in \mathbb{Q}$ where $a, b \in \mathbb{Z}$ with $b \neq 0$. Then a/b is a negative rational number if and only if either (i) a is a positive integer and b is a negative integer, or (ii) a is a negative integer and b is a positive integer.*

Proof. First suppose that a/b is a negative rational number. By Definition 8, the rational number $-(a/b)$ is positive, but $-(a/b) = (-a)/b$ by Exercise 13. So, by the previous theorem, $-a$ and b are either both positive or both negative. The result follows from results of Chapter 4.

Conversely, suppose (i) or (ii) holds. This implies that $-a$ and b are either both positive or both negative. Thus, by the previous theorem, $(-a)/b$ is positive. But $-(a/b) = (-a)/b$ by Exercise 13. So $-(a/b)$ is a positive rational number. Thus a/b is a negative rational number by Definition 8. □

We now show that the definitions of positive and negative numbers really do extend the definitions of positive and negative integer.

Lemma 28. *Let $a \in \mathbb{Z}$. Then $a/1$ is a positive rational number if and only if a is a positive integer. Likewise, $a/1$ is a negative rational number if and only if a is a negative integer.*

Proof. If $a/1$ is a positive rational number then, since 1 is a positive integer, it follows that a is a positive integer (Theorem 26). Conversely, if a is a positive integer, then $a/1$ is a positive rational number since 1 is a positive integer (Theorem 26).

If $a/1$ is a negative rational number then, since 1 is a positive integer, it follows that a is a negative integer (Theorem 27). Conversely, if a is a negative integer, then $a/1$ is a negative rational number since 1 is a positive integer (Theorem 27). □

Theorem 29 (Trichotomy version 1)**.** *If $r \in \mathbb{Q}$ then exactly one of the following occurs: (i) $r = 0$, (ii) r is positive, (iii) r is negative.*

Exercise 18. Prove the above theorem. Hint: you can use Theorem 20 to simplify your proof.

7.7 The incompleteness of $\mathbb{Q}$

In geometry we learn the Pythagorean Theorem which allows us to compute one side of a right triangle assuming we know the lengths of the other two sides. As an application, one easily shows that the diagonal of the unit square has length $\sqrt{2}$. In other words, the length d of the diagonal has the property that $d^2 = 2$. There is a problem: there is no such d in the field $\mathbb{Q}$. This elementary observation shows that $\mathbb{Q}$ does not have all the numbers required to do even basic geometry. So in some sense (to be made precise in Chapter 9) the rational numbers $\mathbb{Q}$ are "incomplete". This compels us to construct a richer number system $\mathbb{R}$ called the *real numbers* to fill in all the gaps in $\mathbb{Q}$. Any real number that is not in $\mathbb{Q}$ is called an *irrational real number*. The number $\pi \in \mathbb{R}$ is another important number missing from $\mathbb{Q}$.

We now formally show that there is indeed no $r \in \mathbb{Q}$ such that $r^2 = 2$. There are several proofs of this fact, and you may have seen a proof in another course. The proof given here is designed to build on our familiarity with modular arithmetic.

Theorem 30. *There is no $r \in \mathbb{Q}$ with the property that $r^2 = 2$.*

For this we need a lemma:

Lemma 31. *If $c \in \mathbb{Z}$ is odd then $c^2 \equiv 1$ modulo 4. If $c \in \mathbb{Z}$ is even then $c^2 \equiv 0$ modulo 4.*

Proof of lemma. Suppose $c \in \mathbb{Z}$ is odd. By the quotient-remainder theorem and the definition of odd integer, $c = 2q + 1$. Thus

$$c^2 = (2q + 1)^2 = 4q^2 + 4q + 1,$$

but

$$4q^2 + 4q + 1 \equiv 0 + 0 + 1 \equiv 1 \bmod 4.$$

Now suppose $c \in \mathbb{Z}$ is even. Then $c = 2d$ for some $d \in \mathbb{Z}$. So $c^2 = 4d^2$. Since $4 \mid c^2$ the result follows. □

Proof of theorem. Suppose such an r exists. By Theorem 22, $r = a/b$ where a and b are relatively prime integers. This implies that a and b cannot both be even. From the assumption $r^2 = 2$ we get the equation $a^2 = 2b^2$. This, in turn, implies that

$$a^2 \equiv 2b^2 \mod 4.$$

Case 1: a and b are both odd. In this case, $a^2 \equiv b^2 \equiv 1 \bmod 4$ by the previous lemma. Substituting into $a^2 \equiv 2b^2$ gives $1 \equiv 2$ modulo 4. This is a contradiction.

Case 2: a is odd, b is even. In this case, $a^2 \equiv 1 \bmod 4$ as before, but $b^2 \equiv 0 \bmod 4$ (by the previous lemma). Substituting into $a^2 \equiv 2b^2$ gives $1 \equiv 0$ modulo 4. This is a contradiction.

Case 3: a is even, b is odd. This case is similar to the previous case. Here $a^2 \equiv 0 \bmod 4$ but $b^2 \equiv 1 \bmod 4$. Substituting into $a^2 \equiv 2b^2$ gives us $0 \equiv 2$ modulo 4. This is a contradiction.

So in any case, we get a contradiction. So no such $r \in \mathbb{Q}$ exists. □

Exercise 19. Adapt the proof of Theorem 30 to show that if $n \equiv 2 \bmod 4$ then there is no $r \in \mathbb{Q}$ such that $r^2 = n$. This shows, for example, that $\sqrt{10}$ is irrational.

Remark 6. One can generalize the above theorem to show that if $n \in \mathbb{Z}$ is not equal to some m^2 (with $m \in \mathbb{N}$) then there is no $r \in \mathbb{Q}$ with $r^2 = n$. Informally we say that if n is not a perfect square then $\sqrt{n}$ is irrational. This result can, in turn, be generalized to other powers beyond 2. We will not prove such results here.

Chapter 8

Sequences and Limits

As we saw in Chapter 7, the field $\mathbb{Q}$ is in some sense "incomplete". There are numbers missing from $\mathbb{Q}$ that are essential for mathematics. Our goal is to construct a number system $\mathbb{R}$ that does not suffer from this problem. We will not construct $\mathbb{R}$ until Chapter 10; we need to do some preliminary work first. We need to understand the concept of limit, and the concepts of infimum and supremum, before we can formalize the idea of "complete" and before we have the tools necessary to carry out the construction of $\mathbb{R}$.

In this chapter we study infinite sequences and their limits. We will do so in the context of ordered fields. The most important ordered fields in mathematics are $\mathbb{Q}$ and $\mathbb{R}$, and although there are other ordered fields studied in mathematics these are in fact the only two ordered fields we will see in this course. Our approach to sequences and limits in ordered fields will thus be a unified approach that will work for both $\mathbb{Q}$ and $\mathbb{R}$. In Chapter 10 we will need sequences of rational numbers in order to construct real numbers. After we have constructed $\mathbb{R}$ we will use sequences of real numbers. Not only are such sequences important in this class, but sequences of real numbers are crucial in all of advanced mathematics, most notably in the field of real analysis.

So in this chapter we aim to prove results for a general ordered field, or sometimes a general Archimedean ordered field. Since $\mathbb{Q}$ is an Archimedean ordered field, the results will automatically be true for $\mathbb{Q}$. After we construct $\mathbb{R}$ and prove that it is too an Archimedean ordered field, the results will automatically become theorems for $\mathbb{R}$ as well.

8.1 Ordered fields

Informally an ordered field is a field with an order relation $<$ that satisfies the usual rules we learn in elementary algebra and arithmetic. Our formal definition does not mention $<$, but focuses on the subset of *positive* elements. The actual order relation $<$ will be formally defined in terms of these positive elements later.

Definition 1. An *ordered field* F is a field with a designated subset P such that (i) P is closed under addition and multiplication, and (ii) for any element $u \in F$ exactly one of the following occurs: $u = 0$, $u \in P$, $-u \in P$.

Remark 1. When we say that P is *closed under addition and multiplication*, we mean that if $x, y \in P$ then $x+y$ and $x \cdot y$ are in P. The second condition, that exactly one of $u = 0$, $u \in P$, $-u \in P$ holds, is called the *first form of trichotomy*.

Definition 2. Let F be an ordered field with designated subset P. The elements in P are called the *positive elements*. We sometimes write the designated subset P as $F_{>0}$.

Definition 3. Let F be an ordered field with designated subset P. If $u \in F$ is such that $-u \in P$ then u is said to be *negative*. (So the first form of trichotomy says that every element satisfies exactly one of the following properties: it is zero, positive, or negative.)

The first example of an ordered field is $\mathbb{Q}$:

Theorem 1. *The field $\mathbb{Q}$ is an ordered field.*

Exercise 1 (Easy)**.** Prove the above theorem using theorems from Chapter 7.

Exercise 2. Show that the field $\mathbb{F}_5$ cannot qualify as an ordered field. Hint: try all possible subsets for P, and show that none work.

Remark 2. This extends: no $\mathbb{F}_p$ can be an ordered field. In fact, ordered fields must be infinite. Later we will study the complex numbers $\mathbb{C}$ which is an example of an infinite field that is not an ordered field.

For most of the rest of the chapter we will consider theorems about a general ordered field F. The only ordered field we have constructed so far is $\mathbb{Q}$, so you can initially think of F as being something like $\mathbb{Q}$. You can freely use all the standard laws of algebra that are true in a general field. In Chapter 10 we will construct the real numbers $\mathbb{R}$, another ordered field, and everything we prove about ordered fields will apply to $\mathbb{R}$ as well.

As the name suggests, an ordered field is ordered by a certain order relation. This is defined as follows.

Definition 4. Let F be an ordered field, and let $x, y \in F$. If $y-x$ is positive then we write $x < y$. We also write $y > x$ in this case.

Exercise 3. Show that $y - x$ is negative if and only if $y < x$. Here x and y are in an ordered field.

Exercise 4. Prove the two following theorems.

Theorem 2. *Suppose $u \in F$ where F is an ordered field. Then u is positive if and only if $u > 0$. Similarly, u is negative if and only if $u < 0$.*

Theorem 3 (Transitivity)**.** *Suppose $x, y, z \in F$ where F is an ordered field. If $x < y$ and $y < z$ then $x < z$.*

Theorem 4. *Suppose $x, y, z \in F$ where F is an ordered field. If $x < y$ then $x + z < y + z$.*

Exercise 5. Prove the above theorem. Hint: simplify $(y + z) - (x + z)$.

Theorem 5. *Let $x, y, x', y' \in F$. If $x < y$ and $x' < y'$ then $x + x' < y + y'$.*

Proof. By Theorem 4 used twice, we have

$$x + x' < y + x' < y + y'.$$

(We also use the commutative law for $+$ and the transitive law for $<$). □

Theorem 6 (Trichotomy version 2)**.** *Suppose $x, y \in F$ where F is an ordered field. Then exactly one of the following occurs: (i) $x = y$, (ii) $y < x$, or (iii) $x < y$.*

Exercise 6. Prove the above. Hint: use the first version of trichotomy for $y - x$ to divide into three disjoint cases. Show for each that that being in that case is equivalent to satisfying one of (i), (ii), or (iii).

Recall the definition of an ordered set from Chapter 2. We now know that every ordered field is an ordered set. Products behave as expected:

Theorem 7. *Suppose $x, y \in F$ where F is an ordered field. If x and y are positive, then xy is positive. If x is positive, but y is negative, then xy is negative. If x and y are negative, then xy is positive.*

Proof. The first statement follows from the definition of ordered field: the positive elements are closed under multiplication.

In the second statement, $-y$ is positive by definition of negative. Thus the product $x(-y)$ is positive by closure. But $x(-y) = -(xy)$ since F is a field (this is true in any ring). Thus $-(xy)$ is positive, so xy is negative.

In the third statement, $-x$ and $-y$ are positive. So $(-x)(-y)$ is positive by closure. But, since F is a field,

$$(-x)(-y) = -\big(x(-y)\big) = -\big(-(xy)\big) = xy.$$

Thus xy is positive. □

Theorem 8. *Suppose $x, y, z \in F$ where F is an ordered field. If $x < y$, and if z is positive, then $xz < yz$. If $x < y$, and if z is negative, then $xz > yz$.*

Exercise 7. Prove the above theorem. Hint: multiply $y - x$ and z.

The follow statement is already known for $F = \mathbb{Q}$. The point of proving it here is to show that it is true of any other possible ordered field F.

Theorem 9. *The element $1 \in F$ is positive, and -1 is negative.*

Proof. Since $0 \neq 1$ in any field, we have that 1 is either positive or negative (by the first version of trichotomy). Suppose 1 is negative. Then $1 \cdot 1$ is positive by Theorem 7. But $1 \cdot 1 = 1$, so 1 is positive, a contradiction.

Since 1 is positive, -1 is negative by definition of negative. □

Theorem 10. *Suppose x is a positive element of an ordered field F. Then the inverse x^{-1} is also positive. Suppose x is a negative element of F. Then the inverse x^{-1} is also negative.*

Proof. Suppose x is positive. Observe that x^{-1} cannot be 0: otherwise

$$1 = xx^{-1} = x \cdot 0 = 0$$

which is not allowed in a field. Observe that x^{-1} cannot be negative: otherwise $xx^{-1} = 1$ must be negative (Theorem 7) contradicting Theorem 9. Thus, by trichotomy, x^{-1} is positive.

The proof of the second claim is similar. □

Theorem 11. *Suppose x, y are positive elements of an ordered field F. If $x < y$ then $y^{-1} < x^{-1}$.*

Proof. Multiply both sides of $x < y$ by $x^{-1}y^{-1}$. □

Now we consider the special case where $F = \mathbb{Q}$. Recall that $\mathbb{Z}$ is regarded as a subset of $\mathbb{Q}$. We have an order for $\mathbb{Z}$ from Chapter 4, and an order for $F = \mathbb{Q}$ defined in the current section. We now show that the new order extends the old order.

Lemma 12. *The order relation $<$ on $\mathbb{Q}$ extends the order relation $<$ on $\mathbb{Z}$. In other words, if $a, b \in \mathbb{Z}$, then $a < b$ (as defined in Chapter 4) if and only if $a < b$ (as defined in this section).*

Proof. Suppose that $a < b$ in the sense of Chapter 4. By the results of Chapter 4, $b - a$ must be a positive integer. By a result of Chapter 7, this means $b - a$ is a positive rational number. Thus $a < b$ in the sense of this section.

This proves one direction. The converse is similar. □

Theorem 13. *Suppose a and b are integers, and d is a positive integer. Consider $a/d, b/d \in \mathbb{Q}$. Then $a/d > b/d$ if and only if $a > b$.*

Proof. If $a/d > b/d$, then multiply both sides by d to show $a > b$. Conversely, suppose that $a > b$. Multiply both sides by d^{-1}. Now $d^{-1} > 0$ by Theorem 10. Thus $a/d > b/d$. □

Less than or equal in ordered fields

Now we define and investigate $\leq$ in an ordered field F.

Definition 5. If $x, y \in F$ then $x \leq y$ means $(x < y) \vee (x = y)$. We also write $y \geq x$ in this case.

Theorem 14. *Let $x, y \in F$. Then the negation of $x < y$ is $y \leq x$. The negation of $y \leq x$ is $x < y$.*

Proof. By version 2 of trichotomy (Theorem 6),

$$\neg\big(x < y\big) \iff y \leq x.$$

The contrapositive of the above gives

$$\neg\big(y \leq x\big) \iff x < y.$$

□

Theorem 15 (Mixed transitivity)**.** *Let $x, y, z \in F$. If $x < y$ and $y \leq z$ then $x < z$. Likewise, if $x \leq y$ and $y < z$ then $x < z$.*

Proof. Suppose that $x < y$ and $y \leq z$. By definition of $y \leq z$, we have either $y < z$ or $y = z$. In the first case, use transitivity of $<$ (Theorem 3). In the second case, use substitution. In either case $x < z$. This proves the first statement. The proof of the second is similar. □

Theorem 16 (Transitivity)**.** *Let $x, y, z \in F$. If $x \leq y$ and $y \leq z$ then $x \leq z$.*

Proof. Suppose that $x \leq y$ and $y \leq z$. By definition of $x \leq y$, we have either $x < y$ or $x = y$. In the first case, use mixed transitivity (Theorem 15). In the second case, use substitution. In either case $x \leq z$ as desired. □

Theorem 17. *Let $x, y, z \in F$. If $x \leq y$ then $x + z \leq y + z$.*

Proof. By definition of $x \leq y$, we have either $x < y$ or $x = y$. In the first case, $x + z < y + z$ by an earlier result (Theorem 4). In the second case, $x + z = y + z$. In either case $x + z \leq y + z$ as desired. □

Theorem 18. *Let $x, y, z \in F$ where $x \leq y$. If $z \geq 0$ then $xz \leq yz$. If $z \leq 0$ then $yz \leq xz$.*

Proof. Assume $z \geq 0$ and $x \leq y$. In the special cases where $z = 0$ or $x = y$ then $xz = yz$, so $xz \leq yz$ holds. So we can assume that $z > 0$ and $x < y$. Now use Theorem 8. This proves the first statement. The proof of the second statement is similar. □

Theorem 19. *Let $x, y, x', y' \in F$. If $x \leq y$ and $x' \leq y'$ then $x + x' \leq y + y'$.*

Proof. By Theorem 17 twice, we have

$$x + x' \leq y + x' \leq y + y'.$$

(We also use the commutative law for addition, and the transitive law for $\leq$.) □

8.2 Absolute values in ordered fields

Absolute values can be defined and developed in any ordered field. Throughout this section, let F be an ordered field.

Definition 6. The *absolute value* $|x|$ of $x \in F$ is defined as follows.

$$|x| \stackrel{\text{def}}{=} \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0. \end{cases}$$

Theorem 20. *If $x \in F$ then $|x| \geq 0$.*

Proof. If $x \geq 0$ then $|x| = x$, so $|x| \geq 0$. If $x < 0$ then $|x| = -x$. Adding $-x$ to both sides of $x < 0$ gives $0 < -x$. Thus $|x| = -x > 0$ in this case. □

Remark 3. The above theorem shows that the absolute value defines a function $F \to F_{\geq 0}$ where $F_{\geq 0}$ is the set $\{x \in F \,|\, x \geq 0\}$.

Informal Exercise 8. Is the function $F \to F_{\geq 0}$ defined by $x \mapsto |x|$ injective? Is it surjective?

The following are easy consequences of the definition.

Theorem 21. *Let $x \in F$. Then*

$$|x| = 0 \iff x = 0.$$

Theorem 22. *Let $x \in F$. Then*

$$|x| > 0 \iff x \neq 0.$$

Theorem 23. *Let $x \in F$. Then $|x| = |-x|$.*

Proof. We use trichotomy to divide the proof into three cases.
If $x > 0$ then $|x| = x$ and $|-x| = -(-x) = x$ (since $-x < 0$).
If $x = 0$ then $|x| = x = 0$ and $|-x| = -x = 0$ (since $-x = 0$).
If $x < 0$ then $|x| = -x$ and $|-x| = -x$ (since $-x > 0$). □

Exercise 9. Prove the following corollary.

Corollary 24. *Let $x, y \in F$. Then $|x - y| = |y - x|$.*

Absolute value is compatible with multiplication.

Theorem 25 (Compatability with multiplication)**.** *Let $x, y \in F$. Then*

$$|xy| = |x| \cdot |y|.$$

Proof. We divide the proof into cases using trichotomy.

If both x and y are positive then so is xy by Theorem 7. So, by the definition of absolute value, $|xy| = xy$ and $|x||y| = xy$.

If both x and y are negative, then xy is positive (Theorem 7). So, by the definition of absolute value, $|xy| = xy$ and

$$|x||y| = (-x)(-y) = -\big(x(-y)\big) = -\big(-(xy)\big) = xy.$$

If either $x = 0$ or $y = 0$ then $|xy| = 0$ and $|x||y| = 0$.

If x is positive and y is negative, then xy is negative (Theorem 7). So, by the definition of absolute value, $|xy| = -xy$ and $|x||y| = x(-y) = -xy$.

The case where x is negative and y is positive is similar. □

Absolute value is compatible with inverses and division as well.

Theorem 26. *Let $x \in F$ be nonzero. Then*

$$\left|x^{-1}\right| = |x|^{-1}.$$

Let $x, y \in F$. Then

$$\left|\frac{x}{y}\right| = \frac{|x|}{|y|}.$$

Exercise 10. Prove the above theorem. Hint: for the first equality, start with proving that $\left|xx^{-1}\right| = 1$, and solve for $\left|x^{-1}\right|$ (with the help of the previous Theorem).

Lemma 27. *If $x \in F$ then $x \leq |x|$ and $-x \leq |x|$.*

Proof. We use trichotomy to divide into cases.

If $x = 0$ then $|x| = 0$ and $-x = 0$. So obviously $x \leq |x|$ and $-x \leq |x|$.

If $x > 0$ then $x = |x|$, so $x \leq |x|$. Also $-x < 0$ and $0 < |x|$, so $-x \leq |x|$.

If $x < 0$ then $-x = |x|$, so $-x \leq |x|$. Since $x < 0$ and $0 \leq |x|$ (Theorem 20), we have $x \leq |x|$. □

Theorem 28. *Suppose $x, y \in F$ where $y \geq 0$. Then*

(i) $|x| < y$ if and only if $-y < x < y$,

(ii) $|x| > y$ if and only if $x > y$ or $x < -y$, and

(iii) $|x| = y$ if and only if $x = y$ or $x = -y$.

Proof. (ia) Suppose that $|x| < y$. Now $x \leq |x|$ (by Lemma 27). So $x < y$ by mixed transitivity (Theorem 15). Also $-x \leq |-x|$ (Lemma 27) and $|-x| = |x|$ (Theorem 23), so $-x \leq |x|$. Thus $-x < y$ by transitivity (Theorem 15). Adding $x - y$ to both sides gives $-y < x$ (Theorem 4). We have both $x < y$ and $-y < x$, so $-y < x < y$.

(ib) Suppose $-y < x < y$. If $x \geq 0$ then $|x| = x$, so $|x| < y$. Otherwise, $x < 0$. Adding $y - x$ to both sides of $-y < x$ gives $-x < y$ (see Theorem 4). Since $|x| = -x$, we have $|x| < y$.

We leave the proofs of (ii) and (iii) to the reader. □

Exercise 11. Prove (ii) and (iii) of the above theorem.

Corollary 29. *Suppose $x, y \in F$ where $y \geq 0$. Then*
(i) $|x| \leq y$ if and only if $-y \leq x \leq y$,
(ii) $|x| \geq y$ if and only if $x \geq y$ or $x \leq -y$, and

The following is sometimes called the "triangle inequality" since the analogous vector version says that the third side of a triangle can be no larger than the sum of the lengths of the other two sides.

Theorem 30 (Triangle inequality)**.** *If $x, y \in F$ then*

$$|x + y| \leq |x| + |y|.$$

Proof. We have $x \leq |x|$ and $y \leq |y|$ (Lemma 27). Thus $x + y \leq |x| + |y|$ by Theorem 19.

We have $-x \leq |x|$ and $-y \leq |y|$ (Lemma 27). Adding $x - |x|$ to both sides of $-x \leq |x|$ gives $-|x| \leq x$ (see Theorem 17). Likewise, $-|y| \leq y$. So $-(|x| + |y|) \leq x + y$ (Theorem 19). Thus

$$-(|x| + |y|) \leq x + y \leq |x| + |y|.$$

By Corollary 29 (i),

$$|x + y| \leq |x| + |y|.$$

□

8.3 Intervals and density in ordered fields

We begin with the definition of finite intervals:

Definition 7 (Intervals)**.** Let F be an ordered field, and let $x, y \in F$ such that $x < y$. Then we define four intervals with endpoints x, y. The associated *open interval* is defined as follows:

$$(x, y) \stackrel{\text{def}}{=} \{z \in F \mid x < z < y\}.$$

The associated *closed interval* is defined as follows:

$$[x, y] \stackrel{\text{def}}{=} \{z \in F \mid x \leq z \leq y\}.$$

The associated *half-closed intervals* are defined as follows

$$[x, y) \stackrel{\text{def}}{=} \{z \in F \mid x \leq z < y\} \qquad (x, y] \stackrel{\text{def}}{=} \{z \in F \mid x < z \leq y\}.$$

It is common also to define *infinite intervals*. For example, (x, ∞) is defined as $\{z \in F \mid z > x\}$.

Exercise 12. Define four other types of infinite intervals (including the interval $(-\infty, \infty)$). Show that all five types of infinite intervals are nonempty.

It is obvious that all the closed and half-closed intervals are nonempty (if $x < y$ as in the above definition). However, is it obvious that (x, y) is nonempty? We will end this section with a proof that (x, y) is indeed nonempty. In other words, for all $x, y \in F$ with $x < y$ there is a $z \in F$ with $x < z < y$. So you can always find new elements between two given elements. Because of this we say that F is *dense*.

The proof is simple: show that the average $(x + y)/2$ is between x and y. One question: is $2 \in F$? Of course, in the case of $\mathbb{Q}$ the answer is clearly yes: $\mathbb{Q}$ contains $\mathbb{Z}$. In general, it can be shown that $\mathbb{Z}$ can be embedded into any ordered field. Instead of showing this now, we make an *ad hoc* definition of the number two. This makes 2 a member of any ordered field. Later we will need 3 as well, so we define 3 along with 2.

Definition 8. Let 1 be the multiplicative identity of an ordered field F. Define 2 to be $1 + 1$. Define 3 to be $2 + 1$.

Remark 4. Observe that 2 is positive in any ordered field F since $1 \in F$ is positive and the set of positive elements is closed under addition. As a consequence, the multiplicative inverse $2^{-1} = 1/2$ exists and is positive in any ordered field. Similarly 3 and $1/3$ are positive in any ordered field.

Remark 5. If we needed to, we could define 4 to be $3 + 1$, and so on. This trick works not just for ordered fields, but for any ring whatsoever. However, in some rings, as in the ring $\mathbb{Z}_4$, we would have $4 = 0$. In ordered fields, positive elements are closed under addition, so we never get 0 by this process. So there is an injection from $\mathbb{N}$ into any ordered field.

Theorem 31 (Density)**.** *Let F be an ordered field. Let $x, y \in F$ be such that $x < y$. Then we can find an element $z \in F$ with $x < z < y$. In other words, then interval (x, y) is nonempty, and the field F is dense.*

Proof. Let $z = (x + y)/2$.

Since $x < y$ we have $x+x < x+y$. Now $x+x = 1 \cdot x+1 \cdot x = (1+1)x = 2x$. So $2x < x + y$. Since 2 is positive, 2^{-1} is positive. Thus $x < (x + y)/2$.

Since $x < y$ we have $x+y < y+y$. Now $y+y = 1 \cdot y+1 \cdot y = (1+1)y = 2y$. So $x + y < 2y$. Since 2 is positive, 2^{-1} is positive. Thus $(x + y)/2 < y$. □

8.4 The Archimedean property

There are two ordered fields that we will study in this course: $\mathbb{Q}$, and later $\mathbb{R}$. Both of these have an important property: they are *Archimedean*. Before explaining what this means, we need the following preliminary definition:

Definition 9 (Subfield)**.** Suppose that K and F are fields, and that K is a subset of F. Suppose also that the ring operations $+$ and $\times$ of K are compatible with the ring operations of F in the sense that the ring operations of F restricted to K gives the ring operations on K. Then we say that K is a *subfield* of F.

Suppose further that K and F are ordered fields, and that the order relation $<$ on K is compatible with the order relation on F in the sense that for $x, y \in K$ we have $x <_K y$ if and only if $x <_F y$. Then we say that K is an *ordered subfield* of F.

We are most interested in ordered fields that contain $\mathbb{Q}$ as an ordered subfield. Such an ordered field F must then contain $\mathbb{N}$ as a subset, and every positive integer is in the set P of positive elements of F.

Definition 10 (Archimedean ordered field)**.** Let F be an ordered field. We say that F is an *Archimedean ordered field* if (i) it contains $\mathbb{Q}$ as an ordered subfield, and (ii) for all $x > 0$ and y in F, there is an $n \in \mathbb{N}$ such that $nx \geq y$.

Remark 6. In the appendix we will discuss the fact that (i) is in some sense true of any ordered field, so we can actually remove this from the definition. The condition (ii) is the main condition. Informally it says that if you have an x which is possibly very small, the integer multiples nx become arbitrarily large. This means that F cannot have so called *infinitesimal* elements. A version of this property was used in geometry by the famous ancient Greek mathematician Archimedes, hence the name.

Note. There is a version of calculus and analysis, called *nonstandard analysis*, where infinitesimals are allowed. This uses a version of the real numbers that is not Archimedean. Advocates of nonstandard analysis claim that some of the simplicity of the original form of calculus of Newton and Leibniz can be preserved if we keep infinitesimals. In addition, infinitesimals dx, dy and so on are often used informally in applied mathematics and in the sciences. However, the majority of current mathematicians use the version of $\mathbb{R}$ with no infinitesimals. We follow this usage: our version of $\mathbb{R}$ will be Archimedean with no infinitesimals.

In an Archimedean field, the integers are unbounded:

Theorem 32. *Let F be an ordered field with ordered subfield $\mathbb{Q}$. Then F is Archimedean if and only if the following holds: for all $y \in F$ there is an $n \in \mathbb{N}$ such that $n \geq y$.*

Exercise 13. Prove the above theorem.

Corollary 33. *Let $y \in F$ where F is an Archimedean ordered field. Then there are integers m, n such that*

$$m \leq y \leq n.$$

Proof. The existence of a suitable n follows from the previous theorem. To find m, apply the above theorem to $-y$. By Theorem 32 there is an integer n' such that $-y \leq n'$. Thus $-n' \leq y$. So if $m = -n'$, then $m \leq y$. □

Exercise 14. Let F be an Archimedean field. Suppose $u > 0$ in F. Show that there is a positive integer n such that $1/n \leq u$. Hint: use Theorem 32.

Theorem 34. *Let F be an ordered field with $\mathbb{Q}$ as an ordered subfield. Then F is Archimedean if and only if the following holds: if $u > 0$ is a positive element of F then there is a positive $n \in \mathbb{N}$ such that $1/n \leq u$.*

Proof. One direction is an exercise (see above). For the other direction, we assume that for all $u > 0$ there is a positive $n \in \mathbb{N}$ such that $1/n \leq u$.

Let $x > 0$ and y be in F. We must show there is an $n \in \mathbb{N}$ with $nx \geq y$. If $y \leq 0$ then $n = 0$ works, so assume that $y > 0$. Let $u = x/y$, and let $n \in \mathbb{N}$ be such that $1/n \leq u$. Then $nx \geq y$ as desired. □

Remark 7. In the above we can also find an n such that $1/n < u$ (strict inequality). This follows from the above and the fact that $1/(k+1) < 1/k$.

Theorem 35. *The field $\mathbb{Q}$ is an Archimedean ordered field.*

Proof. By Theorem 34 we just need to show that if $r \in \mathbb{Q}$ is positive then we can find a positive integer n with $1/n \leq r$. So assume $r \in \mathbb{Q}$ is positive. Hence we can write $r = a/b$ with a, b both positive integers. Since $1 \leq a$ we have $1/b \leq a/b = r$. □

We can strengthen Corollary 33:

Theorem 36. *Let F be an Archimedean ordered field. Suppose $x \in F$. Then there is a unique integer $n \in \mathbb{Z}$ such that $n \leq x < n + 1$.*

Remark 8. In other words, $x = n + y$ where $0 \leq y < 1$. The number y is sometimes called the *fractional part* of x (although it is not always a fraction in the sense of being in $\mathbb{Q}$). We call n the *floor* of x.[1]

[1] In contrast, the smallest integer greater than or equal to x is called the *ceiling of* x. If $n < x < n + 1$ then $n + 1$ is the ceiling of x, but if $n = x$ then n is both the floor and ceiling of x. The floor of x is written $\lfloor x \rfloor$ and the ceiling is written $\lceil x \rceil$.

Proof. By Corollary 33 there are integers a and b such that $a \leq x \leq b$. In particular the set of integers less than or equal to x has upper bound $b \in \mathbb{Z}$. By a result of Chapter 4, this implies that there is a largest integer n such that $n \leq x$. In particular, $n + 1 > x$.

To prove uniqueness, suppose $n_1, n_2 \in \mathbb{N}$ satisfy both $n_1 \leq x < n_1 + 1$ and $n_2 \leq x < n_2 + 1$. So $n_1 < n_2 + 1$ and $n_2 < n_1 + 1$. This implies that $n_1 - 1 < n_2 < n_1 + 1$. Thus $n_2 = n_1$ since n_1 is the only integer between $n_1 - 1$ and $n_1 + 1$. □

The following will be important to us in the case where $F = \mathbb{R}$.

Theorem 37. *Let F be an Archimedean ordered field. Then $\mathbb{Q}$ is dense in F. In other words, if $x < y$ with $x, y \in F$, then there exists an $r \in \mathbb{Q}$ such that $x < r < y$.*

Proof. Let n be a positive integer such that $1/n < (y - x)$. Such an n exists by Theorem 34. In particular $1 < ny - nx$.

By Theorem 36 there is an integer m such that $m \leq nx < m + 1$. Since $m \leq nx$ and $1 < ny - nx$ we have $m + 1 < nx + (ny - nx) = ny$. So $nx < m + 1 < ny$. Hence $x < (m + 1)/n < y$. So $r = (m + 1)/n$ is a rational number between x and y. □

Remark 9. The converse is true as well: if $\mathbb{Q}$ is a dense ordered subfield of an ordered field F then F is an Archimedean ordered field. This follows from Theorem 34.

8.5 Infinite sequences and limits

In this section we consider infinite sequences in an ordered field F. Recall the definition of infinite sequences from Section 9 of Chapter 5. An infinite sequence with values in F is a function whose domain is a set (called the *index set*) of the form $\{i \in \mathbb{Z} \mid i \geq n_0\}$, and whose codomain is F. We use notation such as $(a_i)_{i \geq n_0}$ to denote such a sequence. We simply write (a_i) when the domain is not important to the discussion.

An important concept associated to sequences is that of a *limit*. What do we mean by the limit of a sequence? Informally, a sequence (a_i) has limit b if the terms of the sequence eventually get and stay arbitrarily close to b. This informal description is a bit ambiguous and is unsuitable to use in a proof, so we give a more precise definition.

Definition 11 (Limit)**.** Suppose F is an ordered field, that (a_i) is a sequence in F, and that $b \in F$. We say that b is the *limit* of (a_i) if the following holds: for all positive $\varepsilon \in F$ there is an $N \in \mathbb{N}$ such that if $i \geq N$ then $|a_i - b| < \varepsilon$. We can write this with three quantifiers as follows:

$$(\forall \varepsilon \in F_{>0})(\exists N \in \mathbb{N})(\forall i \in \mathbb{N})\Big(i \geq N \implies |a_i - b| < \varepsilon\Big).$$

Here $F_{>0}$ denotes the positive elements of F.

Not all sequences have limits. If a limit exists the we say the sequence *converges*. This is captured in the following definition:

Definition 12 (Convergence)**.** A sequence that has a limit is said to *converge*. We can write this with *four* quantifiers as follows:

$$(\exists b \in F)(\forall \varepsilon \in F_{>0})(\exists N \in \mathbb{N})(\forall i \in \mathbb{N})\Big(i \geq N \implies |a_i - b| < \varepsilon\Big).$$

A sequence that does not have a limit is said to *diverge*.

Exercise 15. Use the rules of quantifiers in logic to negate the definition of limit. Complete the following sentence: *the sequence* (a_i) *does not have limit* b *means that there is a positive* $\varepsilon \in F$ *such that for all* $N \in \mathbb{N}$ *...*

In a similar manner, use the rules of basic logic to negate the definition of converges. In other words, complete the following sentence: *the sequence* (a_i) *diverges means that for all* $b \in F$ *there exists a positive ...*

Exercise 16. Suppose F is an Archimedean ordered field, and consider the sequence $(i)_{i\in\mathbb{N}}$. In other words, consider the sequence given by the identity function. Show that this sequence diverges. Hint: work with $\varepsilon = 1$, and either use the previous exercise or give a proof by contradiction.

Notice that in the above definition we used the term *the* limit. It sounds like we are treating limits as if they are unique. This is justified by the following theorem.

Theorem 38. *A convergent sequence in an ordered field has a unique limit.*

Proof. Suppose otherwise that (a_i) is a sequence in F with two distinct limits b and c. By trichotomy we get that $b < c$ or $c < b$. Assume that $b < c$. The case where $c < b$ is similar. Let $\varepsilon = (c - b)/2$. By definition of positive and Theorem 2, $c - b > 0$. Since $2 > 0$ we have $2^{-1} > 0$. Thus the product $(c - b)2^{-1}$ is positive. In other words $\varepsilon > 0$.

By definition of limit, there is a $N_1 \in \mathbb{N}$ such that $|a_i - b| < \varepsilon$ for all $i \geq N_1$. Likewise there is a $N_2 \in \mathbb{N}$ such that $|a_i - c| < \varepsilon$ for all $i \geq N_2$. Let i be the maximum of N_1 and N_2. Then since $i \geq N_1$,

$$|a_i - b| < \varepsilon.$$

Hence

$$-\varepsilon < a_i - b < \varepsilon.$$

Using properties of ordered fields we get

$$a_i < b + \varepsilon = b + (c - b)/2 = (b + c)/2.$$

Next, since $i \geq N_2$,

$$|a_i - c| < \varepsilon,$$

a similar argument gives us

$$(b + c)/2 < a_i.$$

Putting these together we conclude

$$a_i < (b + c)/2 < a_i,$$

so $a_i < a_i$ by transitivity. This contradicts trichotomy. □

Remark 10. We often write

$$\lim_{i \to \infty} a_i = b$$

when we wish to assert (1) that the sequence (a_i) converges, and (2) that the unique limit of the sequence (a_i) is b.

Informal Exercise 17. Draw a picture of a number line representing F. Draw b and c in the above proof, and indicate the sets defined by $|x - b| < \varepsilon$ and $|x - c| < \varepsilon$ where ε is as in the above proof. Observe that the sets do not intersect so there can be no a_i simultaneously in both, which is why we got a contradiction. This explains why we chose $\varepsilon = (c - b)/2$. Note, we could have chosen $\varepsilon = (c - b)/4$, for instance, and obtained a contradiction. However, $\varepsilon = 2(c - b)$ does not work. Why not?

Remark 11. In the above theorem we get to choose ε to be whatever we want since we are *assuming* a limit exists. If instead you are trying to *prove* a limit exists, you cannot choose ε, but must allow ε to be an arbitrary positive element of F.

Remark 12. When proving that a particular sequence (a_i) converges to a value b, we need to first find out which N will work *before* we start writing our proof. Our proof will start with assuming an arbitrary $\varepsilon > 0$, and the next step of the proof will be to say which N we choose. The method is to start with the inequality that we want to be true, namely $|a_i - b| < \varepsilon$, and working backwards, we solve for i. We are looking for an inequality of the form $i > x$, where x is usually an expression involving ε. This tells us which values of i will give us the desired inequality. Recalling that the definition requires N to be a natural number, we then choose N to be a natural number satisfying $N \geq x$, which exists in an Archimedean ordered field. Note that since this involves *working backwards* from the inequality we are trying to prove, this work is merely behind-the-scenes scratch work, and it therefore *does not* appear in our proof. In the proof, we simply state which N we choose, without explaining how we figured it out.

Exercise 18. Let F be an Archimedean ordered field. Show that

$$\lim_{j\to\infty} \frac{1}{j} = 0.$$

(In this exercise as well as the following two, assume that the domain of the sequence is the positive integers, since this sequence is not defined when the denominator is 0).

Exercise 19. Let F be an Archimedean ordered field. Show that

$$\lim_{i\to\infty} \frac{2}{i} = 0.$$

Exercise 20. Let F be an Archimedean ordered field. Show that

$$\lim_{k\to\infty} \frac{k+1}{3k} = \frac{1}{3}.$$

Exercise 21. Show that convergent sequences are bounded. In other words, show that if $(a_i)_{i\geq n}$ is a convergent sequence in an ordered field, then there is a bound $M \in F$ such that $|a_i| \leq M$ for all $i \geq n$. Show that if F is an Archimedean ordered field, then we can choose $M \in \mathbb{N}$.

Hint: Using the definition of convergence and a choice of ε (such as $\varepsilon = 1$) first find upper and lower bounds for (a_i) for $i \geq N$ for some N. These bounds work for most terms, but not necessarily when $i < N$. Then find upper and lower bounds for the case when $i < N$. There are only a finite number of terms in this case. Now take the largest of the absolute values of the upper and lower bounds you found above. Prove that this number works. Drawing a picture might help to visualize what is happening.

8.6 Equivalence relation for sequences

Definition 13 (Equivalent sequences)**.** Suppose that $(a_i)_{i\geq n_1}$ and $(b_i)_{i\geq n_2}$ are two sequences in an ordered field F. We write $(a_i) \sim (b_i)$ if the following occurs: for all positive $\varepsilon \in F$ there is a $N \in \mathbb{N}$ such that, for all $i \in \mathbb{N}$,

$$i \geq N \quad \Rightarrow \quad |a_i - b_i| < \varepsilon.$$

Remark 13. Informally the above definition says that the terms of the two sequences get and stay arbitrarily close to each other. Note the N in the above definition should be greater than or equal to both n_1 and n_2.

Lemma 39. *The relation $\sim$ is reflexive on the set of all sequences in F.*

Proof. We need to show $(a_i) \sim (a_i)$ for any given sequence $(a_i)_{i\geq n_0}$ in F. Let ε be an arbitrary positive element of F. We must find an $N \in \mathbb{N}$ such that

$$i \geq N \quad \Rightarrow \quad |a_i - a_i| < \varepsilon.$$

Let us propose $N = n_0$. If $i \geq N$ then $|a_i - a_i| < \varepsilon$ since $|a_i - a_i| = 0$ and ε is positive. So N has the desired property. □

Lemma 40. *The relation $\sim$ is transitive on the set of all sequences in F.*

Proof. Suppose $(a_i) \sim (b_i)$ and $(b_i) \sim (c_i)$. We need to show $(a_i) \sim (c_i)$. Let ε be an arbitrary positive element of F. We must find an $N \in \mathbb{N}$ such that

$$i \geq N \quad \Rightarrow \quad |a_i - c_i| < \varepsilon.$$

To find N we need to use the fact that $(a_i) \sim (b_i)$ and $(b_i) \sim (c_i)$. We work with $\varepsilon' = \varepsilon/2$. Since $(a_i) \sim (b_i)$ there is a $N_1 \in \mathbb{N}$ such that

$$i \geq N_1 \quad \Rightarrow \quad |a_i - b_i| < \varepsilon'.$$

Likewise, there is a $N_2 \in \mathbb{N}$ such that

$$i \geq N_2 \quad \Rightarrow \quad |b_i - c_i| < \varepsilon'.$$

Let us propose for N the larger of N_1 or N_2. Note that if $i \geq N$ then, since $N \geq N_1$, by transitivity we get $i \geq N_1$. As above we get $|a_i - b_i| < \varepsilon'$. Similarly if $i \geq N$ then $i \geq N_2$, so $|b_i - c_i| < \varepsilon'$. Thus if $i \geq N$ then

$$|a_i - c_i| = |(a_i - b_i) + (b_i - c_i)| \leq |a_i - b_i| + |b_i - c_i| < \varepsilon' + \varepsilon'.$$

Here we use the triangle inequality and the fact that $i \geq N_1$ and $i \geq N_2$. Since $\varepsilon = 2\varepsilon'$,

$$|a_i - c_i| < \varepsilon.$$

Thus N has the desired property. □

Remark 14. In the proof above we chose $N = \max\{N_1, N_2\}$. This is a common trick in proofs of this type. We are often looking for an N such that if $i \geq N$ then both Condition 1 and Condition 2 hold. If we know that Condition 1 holds when $i \geq N_1$ and Condition 2 holds when $i \geq N_2$, then we choose $N = \max\{N_1, N_2\}$. Note that the maximum of N_1 and N_2 is greater than or equal to both of them. Then when $i \geq N$, since $N \geq N_1$, by transitivity $i \geq N_1$ and so Condition 1 holds. Similarly when $i \geq N$ we also have $i \geq N_2$ and so Condition 2 holds as well.

Lemma 41. *The relation $\sim$ is symmetric on the set of all sequences in F.*

Exercise 22. Prove the above.

From the above lemmas we get the following:

Theorem 42. *Let F be an ordered field. Then the relation $\sim$ is an equivalence relation on the set of sequences of (a_i) in F.*

Exercise 23. Show that if (a_i) and (b_i) have the same limit, then $(a_i) \sim (b_i)$. Hint: use $\varepsilon' = \varepsilon/2$ to find N_1 and N_2. Choose N to be the maximum of N_1 and N_2.

Theorem 43. *Let (a_i) and (b_i) be sequences in an ordered field F and suppose $(a_i) \sim (b_i)$. If (a_i) has a limit, then (b_i) converges and has the same limit as (a_i).*

Exercise 24. Prove the above theorem. Hint: Let a be the limit of (a_i). Given an $\varepsilon > 0$, you are looking for an $N \in \mathbb{N}$ such that $|b_i - a| < \varepsilon$ when $i \geq N$. What happens if you add and subtract a_i inside of the absolute value? How does that help you to use your two hypotheses?

Remark 15. Suppose (a_i) and (b_i) have the property that $a_i = b_i$ for sufficiently large i. In other words, suppose that there is a k such that $a_i = b_i$ for all $i \geq k$. That is, suppose the two sequences agree after some point. Then $(a_i) \sim (b_i)$. This is easily proved from the definition. So by Theorem 43, if one converges then both do with the same limit.

In particular, if we take a sequence (a_i) and change a finite number of terms, then the resulting sequence is equivalent to (a_i). Likewise, if we change the domain of $(a_i)_{i \geq n_0}$ by replacing n_0 with a larger integer, then the resulting sequence is equivalent.

Because of this, the limit of $(a_i)_{i \geq n_0}$ does not depend on the initial element n_0 of the domain of the sequence. So from now on we will ignore the starting point of sequences when considering limits: it does not matter where the sequence starts, or the behavior of any finite number of terms, but only what happens in the long term.

Note. Several of the definitions and proofs given in this chapter are similar in nature. Be very careful to learn the subtle differences between them. At the same time, since the proof techniques are similar it pays to learn them well.

8.7 Limit laws

The concept of limits, and their basic laws, are extremely important in calculus. In this section we develop many of the basic limit laws.

Theorem 44. *Suppose $(a_i)_{i \geq n_0}$ is a constant sequence with value a. In other words, suppose $a_i = a$ for all $i \geq n_0$. Then (a_i) converges to a. In other words,*

$$\lim_{i \to \infty} a = a.$$

Exercise 25 (Easy)**.** Prove the above theorem. Hint, for all $\varepsilon > 0$, the choice $N = n_0$ will work.

Theorem 45. *Suppose (a_i) and (b_i) are two converging sequences in an ordered field F. Suppose there is a $N \in \mathbb{N}$ such that $a_i \leq b_i$ for all $i \geq N$. Then*

$$\lim_{i \to \infty} a_i \leq \lim_{i \to \infty} b_i$$

Proof. Let a be the limit of (a_i) and let b be the limit of (b_i). Suppose the conclusion fails, and that $a > b$. Let $\varepsilon = (a-b)/2$. Note: $\varepsilon > 0$. By the convergence hypothesis for (a_i), there is a N_1 such that $i \geq N_1$ implies $|a_i - a| < \varepsilon$. Likewise, there is a N_2 such that $i \geq N_2$ implies that $|b_i - b| < \varepsilon$. The existence of N_1 and N_2 follow from the definition of limit. Let N be as in the hypothesis of the theorem, and let i be the maximum of N, N_1, and N_2.

Since $i \geq N_1$ we have $|a_i - a| < \varepsilon$. Thus $-\varepsilon < a_i - a < \varepsilon$. Since $a_i - a > -\varepsilon$ we have $a_i > a - \varepsilon$. Using the rules of arithmetic for fractions developed in the previous chapter and the fact that $\varepsilon = (a-b)/2$, we get $a - \varepsilon = (a+b)/2$. So $a_i > (a+b)/2$.

Since $i \geq N_2$ we have $|b - b_i| < \varepsilon$. Thus $-\varepsilon < b_i - b < \varepsilon$. Since $b_i - b < \varepsilon$ we have $b_i < b + \varepsilon$. Using the rules of arithmetic for fractions developed in the previous chapter and the fact that $\varepsilon = (a-b)/2$, we get $b + \varepsilon = (a+b)/2$. So $b_i < (a+b)/2$.

Combining the above inequalities, we get

$$b_i < (a+b)/2 < a_i,$$

and so we have $a_i > b_i$. But since $i \geq N$ this contradicts the hypothesis of the theorem that $a_i \leq b_i$. □

Remark 16. This result cannot be generalized to $<$. Without more information, given $a_i < b_i$ for all i we cannot conclude strict inequality in the limit. Consider for instance $a_i = 1 - 1/i$ and $b_i = 1 + 1/i$.

Corollary 46. *Suppose (a_i) is a converging sequence in an ordered field F with limit a. Suppose that b is an upper bound of (a_i). Then $a \leq b$. In other words,*

$$\lim_{i \to \infty} a_i \leq b.$$

Suppose instead that b is a lower bound of (a_i). Then

$$\lim_{i \to \infty} a_i \geq b.$$

Proof. Apply Theorem 45 to (a_i) and the constant sequence with terms b. □

Remark 17. In the above, b does not have to be a bound for all a_i. It is enough that it is a valid bound for a_i with $i \geq N$ for some fixed N. (The proof above generalizes to this case).

Remark 18. If $(a_i)_{i \geq n}$ and $(b_i)_{j \geq m}$ are two sequences, then we can define $(a_i + b_i)_{i \geq p}$ to be the sequences whose value is the sum. Here the starting index p can be defined as the maximum of n and m. As in Remark 15 the starting index is irrelevant to our investigation of limits, so we will usually not indicate it. Similar comments apply to sequences of products $(a_i b_i)$.

For the sequence $(1/b_i)$ we have to choose a starting index p so that $b_i \neq 0$ for $i \geq p$. The problem is that there might not be such a p. If there is such a p, we choose the smallest such p as our starting index, and then $(1/b_i)$ is well-defined. Similar comments apply to sequences of quotients (a_i/b_i).

Theorem 47 (Sum law)**.** *Let (a_i) and (b_i) be sequences with values an ordered field F. If (a_i) and (b_i) converge then the sequence (a_i+b_i) converges and*

$$\lim_{i\to\infty}(a_i + b_i) = \lim_{i\to\infty} a_i + \lim_{i\to\infty} b_i.$$

Proof. Let a be the limit of (a_i) and let b be the limit of (b_i). Let $\varepsilon > 0$ be in F. We must show that there is a $N \in \mathbb{N}$ such that $|(a_i + b_i) - (a + b)| < \varepsilon$ for all $i \geq N$.

Let $\varepsilon' = \varepsilon/2$. Since (a_i) converges to a, there is a N_1 such that $|a_i - a| < \varepsilon'$ for all $i \geq N_1$. Likewise, there is a N_2 such that $|b_i - b| < \varepsilon'$ for all $i \geq N_2$. Let N be the maximum of N_1 and N_2. Suppose $i \geq N$. Then

$$\begin{aligned}
|(a_i + b_i) - (a + b)| &= |(a_i - a) + (b_i - b)| && (F \text{ is a field}) \\
&\leq |a_i - a| + |b_i - b| && (\text{triangle inequality}) \\
&< \varepsilon' + \varepsilon' && (i \geq N_1 \text{ and } i \geq N_2) \\
&= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && (\text{choice of } \varepsilon') \\
&= \left(\frac{1}{2} + \frac{1}{2}\right)\varepsilon && (F \text{ is a field}) \\
&= \varepsilon && (1 + 1 \stackrel{\text{def}}{=} 2 \text{ in } F).
\end{aligned}$$

□

Theorem 48. *Let (a_i) be a sequence with values in an ordered field F, and let $c \in F$. If (a_i) converges then (ca_i) converges and*

$$\lim_{i\to\infty} c\, a_i = c \lim_{i\to\infty} a_i.$$

Corollary 49. *Let (a_i) be a sequence with values in an ordered field F. If (a_i) converges, then so does $(-a_i)$, and*

$$\lim_{i\to\infty} -a_i = -\lim_{i\to\infty} a_i$$

Exercise 26. Prove the above theorem and corollary. Hint: in the case that $c \neq 0$ choose $\varepsilon' = \varepsilon/|c|$. What if $c = 0$?

Theorem 50 (Product law)**.** *Let (a_i) and (b_i) be sequences with values in an ordered field F. If (a_i) and (b_i) converge, then (a_ib_i) converges and*

$$\lim_{i\to\infty} a_ib_i = \lim_{i\to\infty} a_i \cdot \lim_{i\to\infty} b_i.$$

Proof. Let a be the limit of (a_i) and let b be the limit of (b_i). Let $\varepsilon > 0$ be in F. We must show that there is an $N \in \mathbb{N}$ such that $|a_i b_i - ab| < \varepsilon$ for all $i \geq N$.

By Exercise 21 there is a bound M_1 such that $|a_i| \leq M_1$ for all i in the domain of (a_i). Let M be the maximum of $M_1, |b|$, and 1. Thus $|b| \leq M$ and $0 < M$ and $|a_i| \leq M$ for all a_i. Let $\varepsilon' = \varepsilon/(2M)$. Observe that ε' is positive. Since (a_i) converges to a, there is a N_1 such that $|a_i - a| < \varepsilon'$ for all $i \geq N_1$. Likewise, there is a N_2 such that $|b_i - b| < \varepsilon'$ for all $i \geq N_2$. Let N be the maximum of N_1 and N_2. Suppose $i \geq N$. Then

$$\begin{aligned}
|a_i b_i - ab| &= |a_i b_i - a_i b + a_i b - ab| && (F \text{ is a field}) \\
&= |a_i(b_i - b) + b(a_i - a)| && (F \text{ is a field}) \\
&\leq |a_i(b_i - b)| + |b(a_i - a)| && (\text{triangle inequality}) \\
&= |a_i||b_i - b| + |b||a_i - a| && (|xy| = |x||y|) \\
&\leq M|b_i - b| + M|a_i - a| && (\text{bound on } |a_i| \text{ and } |b|) \\
&< M\varepsilon' + M\varepsilon' && (i \geq N_1 \text{ and } i \geq N_2) \\
&= M\frac{\varepsilon}{2M} + M\frac{\varepsilon}{2M} && (\text{choice of } \varepsilon') \\
&= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && (MM^{-1} = 1) \\
&= \left(\frac{1}{2} + \frac{1}{2}\right)\varepsilon && (F \text{ is a field}) \\
&= \varepsilon && (1 + 1 \stackrel{\text{def}}{=} 2 \text{ in } F).
\end{aligned}$$

□

Corollary 51 (Power law)**.** *Let (a_i) be a sequence in an ordered field F and let $n \in \mathbb{N}$. If (a_i) converges, then so does (a_i^n) and*

$$\lim_{i\to\infty} a_i^n = \left(\lim_{i\to\infty} a_i\right)^n.$$

Exercise 27. Prove the above corollary by a simple induction argument.

Lemma 52. *Let (b_i) be a sequence in an ordered field F. Suppose that (b_i) converges to a non-zero value b. Then there is an $N \in \mathbb{N}$ such that*

$$|b_i| \geq |b|/2$$

for all $i \geq N$.

Remark 19. In particular, in the above situation $b_i \neq 0$ if $i \geq N$. So $(1/b_i)$ is well-defined, and (a_i/b_i) is well-defined where (a_i) is any sequence in F. See Remark 18.

Proof. Let $\varepsilon = |b|/2$. Observe that $\varepsilon > 0$. Since (b_i) has limit b, there is an N such that the following holds: if $i \geq N$ then $|b_i - b| < \varepsilon$. In this case $|b_i - b| < |b|/2$ and so, with the triangle inequality,

$$|b| = |b_i + (b - b_i)| \leq |b_i| + |b - b_i| < |b_i| + |b|/2.$$

Thus if $i \geq N$

$$|b_i| > |b| - |b|/2 = (2-1)|b|/2 = |b|/2$$

(note we defined $2 = 1 + 1$ in any ordered field, so $2 - 1 = 1$). In particular, if $i \geq N$ then $|b_i| \geq |b|/2$. □

Theorem 53 (Inverse law)**.** *Let (b_i) be a sequence with values in an ordered field F. Suppose (b_i) converges and*

$$\lim_{i\to\infty} b_i \neq 0.$$

Then (b_i^{-1}) is a well-defined convergent sequence and

$$\lim_{i\to\infty} b_i^{-1} = \left(\lim_{i\to\infty} b_i\right)^{-1}.$$

Proof. The previous lemma shows (b_i^{-1}) is well-defined. Let $b \neq 0$ be the limit of (b_i). Let $\varepsilon > 0$ be in F. We must show that there is an $N \in \mathbb{N}$ such that $|1/b_i - 1/b| < \varepsilon$ for all $i \geq N$.

By the previous lemma, there exists an $N_1 \in \mathbb{N}$ such that if $i \geq N_1$ then $|b_i| \geq |b|/2$. In particular, if $i \geq N_1$ then

$$\frac{1}{|b_i|} \leq \frac{2}{|b|}.$$

Let $\varepsilon' = \varepsilon|b|^2/2$. Observe that $\varepsilon' > 0$. Since (b_i) converges to b, there is an $N_2 \in \mathbb{N}$ such that if $i \geq N_2$ then $|b_i - b| < \varepsilon'$.

Let N be the maximum of N_1 and N_2. Suppose $i \geq N$. Then

$$\begin{aligned}
\left|\frac{1}{b_i} - \frac{1}{b}\right| &= \left|\frac{b - b_i}{b_i b}\right| && (F \text{ is a field}) \\
&= \frac{|b - b_i|}{|b_i||b|} && (|xy^{-1}| = |x||y|^{-1} \text{ and } |xy| = |x||y|) \\
&< \frac{\varepsilon'}{|b_i||b|} && (i \geq N_2) \\
&= \varepsilon' \frac{1}{|b|}\frac{1}{|b_i|} && (F \text{ is a field}) \\
&\leq \varepsilon' \frac{1}{|b|}\frac{2}{|b|} && (i \geq N_1) \\
&= \frac{\varepsilon|b|^2}{2}\frac{1}{|b|}\frac{2}{|b|} && (\text{choice of } \varepsilon') \\
&= \varepsilon. && (F \text{ is a field})
\end{aligned}$$

□

Exercise 28. Prove the following two corollaries.

Corollary 54 (Quotient law)**.** *Let (a_i) and (b_i) be sequences with values in an ordered field F. Suppose (a_i) and (b_i) converge. Suppose also that the limit of (b_i) is not zero. Then (a_i/b_i) converges and*

$$\lim_{i\to\infty} \frac{a_i}{b_i} = \frac{\lim_{i\to\infty} a_i}{\lim_{i\to\infty} b_i}.$$

Corollary 55 (Power law 2)**.** *Let (a_i) be a sequence in an ordered field F and let $n \in \mathbb{Z}$. If (a_i) converges to a nonzero limit, then so does (a_i^n) and*

$$\lim_{i\to\infty} a_i^n = \left(\lim_{i\to\infty} a_i\right)^n.$$

8.8 Suprema and infima

Recall that if a nonempty subset S of $\mathbb{Z}$ has an upper bound, then S actually has a maximum. Likewise, if such S has a lower bound then it has a minimum. This is *not* true for ordered fields such as $\mathbb{R}$ and $\mathbb{Q}$. Consider the interval $(0, 1)$ in $\mathbb{Q}$ (or in $\mathbb{R}$). It is bounded above and below, but has no maximum or minimum. However, the interval $(0, 1)$ has a least upper bound 1 and a greatest lower bound 0. Least upper bounds (suprema) and greatest lower bounds (infima) are the "next best thing" to maxima and minima. They differ from maxima and minima in the sense that they are sometimes not in the set S itself, but are always "arbitrarily close" to S.

When we construct the real numbers we will see that *any* nonempty subset with an upper bound has a least upper bound, and any nonempty subset with a lower bound has a greatest lower bound. According to the following example, $\mathbb{Q}$ lacks this property: it has bounded sets without least upper bounds.

Example (Informal). Consider the set $S = \{x \in \mathbb{Q} \mid x^2 \leq 2\}$. This set has many upper bounds, such as 3 or even 3/2, but it turns out that no $b \in \mathbb{Q}$ can be a *least* upper bound. In fact, a positive rational number r is an upper bound if and only if $r^2 \geq 2$. But if $r^2 \geq 2$ then, since $\mathbb{Q}$ has no square root of 2, we know $r^2 > 2$. It turns out that given such an r one can show that there is smaller positive rational number r' such that $(r')^2 > 2$. So no given r is a least upper bound.

On the other hand, $S = \{x \in \mathbb{R} \mid x^2 \leq 2\}$ has a least upper bound in $\mathbb{R}$. It turns out to be $b = \sqrt{2}$.

Definition 14 (Supremum)**.** Let S be a nonempty subset of an ordered field F. A *supremum* M of S is a least upper bound in the following sense.

1. $M \geq x$ for all $x \in S$.
2. If $B \in F$ is such that $B \geq x$ for all $x \in S$ then $M \leq B$.

Definition 15 (Infimum)**.** Let S be a nonempty subset of an ordered field F. An *infimum* m of S is a greatest lower bound in the following sense.

1. $m \leq x$ for all $x \in S$.
2. If $b \in F$ is such that $b \leq x$ for all $x \in S$ then $b \leq m$.

Remark 20. The plural of "supremum" is "suprema" and the plural of "infimum" is "infima". This reflects the Latin origin of these words.

Suprema are similar to maxima, but they do not have to be in the set S. Likewise infima are like minima. For example, if they exist then they are unique.

Exercise 29 (Easy)**.** Show that if S is a nonempty unbounded subset of an ordered field in the sense that S has no upper bounds, then S has no supremum. Note: sometimes in this case the supremum is defined to be ∞. Likewise a nonempty set with no lower bound is sometimes said to have infimum $-\infty$. *In this course we will say that the supremum or infimum do not exist in these cases.*

Exercise 30. Prove the following two theorems.

Theorem 56. *Let S be a nonempty subset of an ordered field F. If S has a supremum then it has a unique supremum. Similarly, if S has an infimum then it has a unique infimum.*

Theorem 57. *Let S be a nonempty subset of an ordered field F. If S has a maximum, then the supremum of S is the maximum. If S has a minimum, then the infimum of S is the minimum.*

Exercise 31. Prove the following corollary.

Corollary 58. *Let S be a nonempty subset of an ordered field F. If a bound is in the set S itself, then it must be a supremum or infimum. More precisely, if $M \in S$ is an upper bound, it must be the supremum and the maximum of S. If $m \in S$ is a lower bound, it must be the infimum and the minimum of S.*

Suprema and infima do not have to be in the set they bound, but they are "arbitrarily close" to the set. This is made precise in the following theorem:

Theorem 59. *Let S be a nonempty subset of an ordered field F. Then an upper bound M is the supremum of S if and only if the following is true: for all $\varepsilon > 0$ there is an element of S in the interval $(M - \varepsilon, M]$. Similarly a lower bound m is the infimum of S if and only if the following is true: for all $\varepsilon > 0$ there is an element of S in the interval $[m, m + \varepsilon)$.*

Proof. We prove the claim for upper bounds. The claim for lower bounds is similar. So let M be an upper bound. First we assume that for all $\varepsilon > 0$ there is an element of S in the interval $(M - \varepsilon, M]$. We will show that M is a supremum using the definition. This requires us to show that if B is an upper bound then $M \leq B$. Suppose otherwise that such an upper bound of S has the property that $B < M$. Since B is an upper bound, there are no elements of S in $(B, M]$. Let $\varepsilon = M - B$. So the interval $(B, M]$ is just the interval $(M - \varepsilon, M]$. But this interval has no elements of S, a contradiction.

Now suppose M is a supremum. We will show that if $\varepsilon > 0$ then the interval $(M - \varepsilon, M]$ intersects S. By the definition of supremum, $M - \varepsilon$ cannot be an upper bound since $M - \varepsilon < M$. Since $M - \varepsilon$ is not an upper bound, there is an element $x \in S$ such that $x > M - \varepsilon$ for some $x \in S$. Since M is an upper bound, $x \leq M$. Combining the inequalities gives us $x \in (M - \varepsilon, M]$ as desired. □

The above shows that suprema and infima are on the "boundary" of S in some sense. The following illustrates this as well with sequences. We state and prove the version for suprema, but of course there is a version for infima as well. We will need this theorem in the next chapter when we prove the intermediate value theorem.

Theorem 60. *Let S be a nonempty subset of an Archimedean ordered field F. If M is the supremum of S then M is the limit of a sequence (a_i) of elements $a_i \in S$ and is the limit of a sequence (b_i) of elements $b_i \notin S$. If $a < M$ is given we can assume that each a_i is in the interval $[a, M]$. Similarly, if $b > M$ is given we can assume that each b_i is in the interval $[M, b]$.*

Proof. We will prove the case involving the supremum, and leave the other case as an exercise. So we will assume that $a < M$ where M is the supremum of S.

For each $i \in \mathbb{N}^+$, choose $\varepsilon_i = \min\{1/i, M - a\} > 0$. By Theorem 59, there is an $a_i \in (M - \varepsilon_i, M]$ such that $a_i \in S$.

First we show that each $a_i \in [a, M]$. Since $a_i \in (M - \varepsilon_i, M]$, we have

$$M - \varepsilon_i < a_i \leq M.$$

Also since $\varepsilon_i \leq M - a$ we get

$$M - \varepsilon_i \geq M - (M - a) = a,$$

which yields: $a \leq M - \varepsilon_i < a_i \leq M$.

Next we show that $|a_i - M| < 1/i$. Since $a_i \leq M$, $a_i - M$ is negative or 0, and so

$$|a_i - M| = -(a_i - M) = M - a_i.$$

Then using $M - \varepsilon_i < a_i$, we conclude $M - a_i < \varepsilon_i$. Finally since $\varepsilon_i \leq 1/i$,

$$|a_i - M| = M - a_i < \varepsilon_i \leq 1/i.$$

Finally we prove that (a_i) converges to M. Assume $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $N > 1/\varepsilon$. This is possible because F is an Archimedean ordered field.

Assume $i \in \mathbb{N}$ and $i \geq N$. Then

$$|a_i - M| < 1/i \leq 1/N < 1/(1/\varepsilon) = \varepsilon.$$

Thus M is the limit of the sequence (a_i) where each of the elements a_i are in $S \cap [a, M]$. □

Exercise 32. Using the above theorem and proof for the case involving the supremum for ideas, state and prove the corresponding theorem for the case involving the infimum.

Suprema and infima do not always exist for bounded sets in $\mathbb{Q}$. So we cannot prove they exist in general (at least not without the extra completeness property, see Chapter 9). However, we can prove a partial result that will be useful when we study Cauchy sequences in Chapter 9. We prove the version for suprema only, but obviously an analogous result holds for infima as well. First a definition.

Definition 16. Let S be a nonempty subset of an ordered field F, and let $\varepsilon > 0$ be in F. An *ε-almost-supremum* A of S is an upper bound of S such that there is an $x \in S$ in the interval $(A - \varepsilon,\ A]$.

Theorem 61. *Let S be a nonempty subset of an Archimedean ordered field F, and let $\varepsilon > 0$ be in F. If S is bounded from above, then S has an ε-almost-supremum.*

Proof. Let $x_0 \in S$. Such a element exists since S was assumed to be nonempty. Let B be an upper bound of S. Since F is an Archimedean ordered field, there is an integer $n \in \mathbb{N}$ such that $n\,\varepsilon \geq (B - x_0)$ (Definition 10). This implies that $x_0 + \varepsilon n \geq B$, so $x_0 + \varepsilon n$ is an upper bound of S. By the well-ordering principle, there is a smallest $n_0 \in \mathbb{N}$ such that $x_0 + \varepsilon n_0$ is an upper bound of S. Let $A = x_0 + \varepsilon n_0$ for such an n_0. So A is an upper bound.

Observe that $A - \varepsilon = x_0 + \varepsilon(n_0 - 1)$ is not an upper bound of S. (There are two cases. If $n_0 = 0$ this is true since $x_0 > A - \varepsilon$. If $n_0 > 0$, this is true based on the choice of n_0). Thus there is an element $x \in S$ with $A - \varepsilon < x$. Since A is an upper bound, $x \leq A$. So $x \in (A - \varepsilon, A]$. This means that A is an ε-almost-supremum. □

8.9 Embedding of $\mathbb{Q}$ in ordered fields (optional)

Let F be an ordered field. We can embed $\mathbb{N}$ into F using the idea mentioned in Remark 5. Informally, $n \in \mathbb{N}$ is mapped to $1 + 1 + \ldots + 1$ where the sum

has n terms, and where $1 \in F$ is the multiplicative identity. Since F has additive inverses, we can extend this embedding to give an embedding of $\mathbb{Z}$ into F. Finally since F has multiplicative inverses for nonzero elements, we can extend this to an embedding of $\mathbb{Q} \to F$. These embeddings can all be shown to be injective, and so we can think of $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ as subsets of F. We can show that $\mathbb{Q}$ is actually an ordered subfield of F. This gives a sketch of the argument that every ordered field contains $\mathbb{Q}$ as an ordered subfield.

Instead of making this a formal theorem, we just make the assumption that F containes $\mathbb{Q}$ whenever it is convenient. We can get away with this short-cut since the only ordered fields we will see, $\mathbb{Q}$ and $\mathbb{R}$, obviously contain $\mathbb{Q}$. However, it is nice to know that it is automatically true for ordered fields in general.

Once we know that every ordered field contains $\mathbb{Q}$ as an ordered subfield, we can simplify the definition of *Archimedean* by replacing the definition with the second requirement.

Chapter 9

Completeness and Continuity

The goal of this chapter is to give a formal definition of the notion of a *complete* ordered field, and explore issues related to this concept. We will construct a complete ordered field $\mathbb{R}$ in the next chapter.

Recall that in Chapter 7 we showed that $\mathbb{Q}$ does not have a square root of 2. Informally this illustrates the incompleteness of $\mathbb{Q}$. One of the goals of this chapter is to show that any complete ordered field possesses a square root of 2, and in fact a square root of any nonnegative element. To do so we will prove a basic version of the intermediate value theorem. Since $\mathbb{Q}$ lacks a square root for 2 we conclude that it is, as expected, incomplete in the formal sense.

The intermediate value theorem requires the notion of *continuous function*. This is one of the key concepts of analysis, and we will just scratch the surface of this concept in this course. We want just enough results about continuity to prove the intermediate value theorem, and apply it to functions such as $f(x) = x^2$ in order to obtain square roots.

9.1 Completeness

We begin with a formal definition of the concept of *complete*. There are several equivalent ways to define this concept. For example, the following uses the existence of suprema, but obviously one could form an alternative definition that uses the existence of infima instead.

Definition 1 (Completeness)**.** Let F be an ordered field. We say that F is *complete* if every nonempty subset $S \subseteq F$ which is bounded from above has a supremum (least upper bound).

The existence of suprema gives us the existence of infima as well:

Theorem 1. *Suppose F is a complete ordered field. Suppose $S \subseteq F$ is a nonempty subset which is bounded from below. Then S has an infimum (greatest lower bound).*

Proof. Consider the set $S' = \{-x \mid x \in S\}$ of additive inverses. Observe that S' is nonempty and bounded from above. Since F is complete, the set S' has a supremum M. Let $m = -M$. Then observe that m is the infimum of S. □

Exercise 1. Complete the above proof by giving detailed justifications for the two observations in the proof.

Theorem 2. *Suppose F is a complete ordered field containing $\mathbb{Q}$ as an ordered subfield. Then F is an Archimedean ordered field.*

Proof. From a results in Chapter 8 it is enough to show that for all $y \in F$ there is an $n \in \mathbb{N}$ such that $n \geq y$.

Suppose instead that this condition fails. Then there is a $y \in F$ such that $n < y$ for all $n \in \mathbb{N}$. This implies that $\mathbb{N}$ is bounded. Since F is complete, this means that $\mathbb{N}$ has a supremum M. Since $M - 1 < M$, the definition of supremum tells us that $M - 1$ is not an upper bound for $\mathbb{N}$, so there must be an integer $n \in \mathbb{N}$ such that $M - 1 < n$. From this we have that $M < n + 1$. Thus M is also not an upper bound of $\mathbb{N}$, a contradiction. □

Note. We noted in Chapter 8 that is is possible to think of $\mathbb{Q}$ as a subfield of any ordered field. If we do this, then the theorem can be stated more succinctly as *every complete ordered field is an Archimedean ordered field.*

9.2 Continuous functions

Now we explore the concept of continuity in order to set the stage for the intermediate value theorem. Due to the emphasis we place on on sequences in this course, we use the sequential definition of continuity. The next section (optional) gives another very common definition of continuity, and shows that the two definitions are equivalent.

Definition 2 (Continuity)**.** Let S be a subset of an ordered field F. Then a function $f\colon S \to F$ is said to be *continuous* on S if for all converging sequences (a_i) with terms and limit in S, the sequence $(f(a_i))$ also converges, and

$$\lim_{i\to\infty} f(a_i) \;=\; f\left(\lim_{i\to\infty} a_i\right).$$

The most basic examples of continuous functions are the identity and constant functions:

Theorem 3. *Let S be a subset of an ordered field F. Any constant function $x \mapsto c$ with $c \in F$ is a continuous function $S \to F$. The identity function $x \mapsto x$ is a continuous function $F \to F$.*

Exercise 2. Prove the above. Hint: showing this for the identity function is truly trivial. For a constant function, observe that $f(a_i)$ is a constant sequence. What do you know about limits of constant sequences?

Definition 3. Let f and g be functions $S \to F$ where F is a field. The function $f + g$ is defined to be the function $S \to F$ which sends $x \in S$ to $f(x) + g(x)$. The function $f \cdot g$ is defined to be the function $S \to F$ which sends $x \in S$ to $f(x) \cdot g(x)$.

Theorem 4 (Closure under addition)**.** *Let S be a subset of an ordered field F. If f, g are continuous on S then $f + g$ is also continuous on S. In other words, the set of continuous functions is closed under addition.*

Proof. By the definition of continuity (Definition 2), we take an arbitrary converging sequence (a_i) with limit a. We assume that each a_i is in S and that a is in S, and we must show that the sequence $((f + g)(a_i))$ converges with limit $(f + g)(a)$.

By Definition 3, $(f+g)(a_i) = f(a_i)+g(a_i)$, and $(f+g)(a) = f(a)+g(a)$. Since f and g are continuous on S the sequences $(f(a_i))$ and $(g(a_i))$ both converge. By the limit laws of Chapter 8, the sequence $\big(f(a_i)+g(a_i)\big)$ must also converge since it is the sum of convergent sequences. In fact,

$$\begin{aligned}
\lim_{i\to\infty}(f+g)(a_i) &= \lim_{i\to\infty}\big(f(a_i)+g(a_i)\big) && \text{(Sum of functions (Def 3))}\\
&= \lim_{i\to\infty} f(a_i) + \lim_{i\to\infty} g(a_i) && \text{(Limit law for + (Ch. 8))}\\
&= f\left(\lim_{i\to\infty} a_i\right) + g\left(\lim_{i\to\infty} a_i\right) && (f \text{ and } g \text{ continuous})\\
&= f(a) + g(a) && (a \text{ is the limit})\\
&= (f+g)(a). && \text{(Sum of functions (Def 3))}
\end{aligned}$$

□

Exercise 3. Prove the following.

Theorem 5 (Closure under multiplication)**.** *Let S be a subset of an ordered field F. If f, g are continuous on S then the function $f \cdot g$ is also continuous on S. In other words the set of continuous functions is closed under multiplication.*

Definition 4. Let S be a subset of an ordered field F. Then let $\mathcal{C}(S)$ be the set of continuous functions $S \to F$.

By Definition 4, Theorem 4, and Theorem 5, the set $C(S)$ has binary operations $+$ and $\cdot$. Is $C(S)$ a ring? Let us begin with the associative law:

Lemma 6. *The operation $+$ is associative on $\mathcal{C}(S)$.*

Proof. We must show that if $f, g, h \in \mathcal{C}(S)$ then $(f+g)+h = f+(g+h)$. To show functions are equal, it is enough to show that they have equal value for an arbitrary x in the domain. So let $x \in S$ be in the domain. Then

$$\begin{aligned}
\big((f+g)+h\big)(x) &= (f+g)(x)+h(x) && \text{(Sum of functions (Def 3))}\\
&= \big(f(x)+g(x)\big)+h(x) && \text{(Sum of functions (Def 3))}\\
&= f(x)+\big(g(x)+h(x)\big) && \text{(Addition in field } F \text{ is assoc.)}\\
&= f(x)+(g+h)(x) && \text{(Sum of functions (Def 3))}\\
&= (f+(g+h))(x). && \text{(Sum of functions (Def 3))}
\end{aligned}$$

Since $x \in S$ is arbitrary, $(f+g)+h = f+(g+h)$. □

Exercise 4. Show that $+$ for $\mathcal{C}(S)$ is also commutative. Show multiplication for $\mathcal{C}(S)$ is associative and commutative. Show that the distributive law holds for $\mathcal{C}(S)$.

Theorem 7. *Let S be a subset of an ordered field F. Then the set of continuous functions $\mathcal{C}(S)$ is a commutative ring.*

Exercise 5. Complete the proof of the above theorem. What are the 0 and 1 elements in the ring $\mathcal{C}(S)$?

Informal Exercise 6. Assume the existence of $\mathbb{R}$ (informally at this point). Informally, a continuous function on $[0,1]$ is a function $[0,1] \to \mathbb{R}$ whose graph is a connected curve.

Show that $\mathcal{C}([0,1])$ is not an integral domain. Do so by sketching the graph of two continuous functions whose product is zero.

Example. Let S be a subset of an ordered field F. If $f(x) = x$, then we know that $f \in \mathcal{C}(S)$ since it is the identity function (Theorem 3). Thus $f \cdot f \in \mathcal{C}(S)$ by closure under multiplication. However $f \cdot f$ is just the function $x \mapsto x^2$. Thus $g(x) = x^2$ is a continuous function.

By induction, we can similarly show $g(x) = x^k$ is continuous for all $k \in \mathbb{N}$.

Example. The previous example shows that $g(x) = x^k$ is continuous for integers $k \in \mathbb{N}$. We can call such functions *monomial functions*. Since constant functions are continuous, and continuous functions are closed under multiplication, the function $g(x) = cx^k$ is continuous as well.

Observe that any polynomial function is the sum of functions of the form $g(x) = cx^k$. Since continuous functions are closed under sum, we conclude that any polynomial function is continuous.

9.3 The δ-ε definition of continuity (optional)

Definition 2 is sometimes called the definition of *sequential continuity*. There is another definition of continuity that is often used in analysis which is called the δ-ε definition of continuity. In this section we will show that both definitions are equivalent. This fact will not be used in this course, but it is an important fact in analysis.

Definition 5 (Second Definition of Continuity)**.** Let S be a subset of an ordered field F. Let $f\colon S \to F$ be a function and let $a \in S$. We say that f is *continuous at* a in the δ-ε sense if the following holds. For all $\varepsilon > 0$ in F there is a $\delta > 0$ such that for all $x \in S$ if $|x - a| < \delta$ then $|f(x) - f(a)| < \varepsilon$.

If f is continuous in this sense for all $a \in S$ then we say that $f\colon S \to F$ is continuous on S (in the δ-ε sense).

Notice that we defined continuity at a point. We have not yet done this for the sequential form of continuity. We do so now:

Definition 6. Let S be a subset of an ordered field F. Let $f\colon S \to F$ be a function and let $a \in S$. We say that f is *continuous at* a in the sequential sense if the following holds. For all sequences (a_i) converging to a with terms in S, the sequence $\big(f(a_i)\big)$ converges to $f(a)$.

By definition f is continuous in the sequential sense if and only if it is continuous at a in the sequential sense for all $a \in S$.

We will now prove the equivalence of the two notions of continuity using two lemmas, one for each direction of the equivalence.

Lemma 8. *Let S be a subset of an ordered field F. Let $f\colon S \to F$ be a function and let $a \in S$. If f is continuous at a in the δ-ε sense, then it is continuous at a in sequential sense.*

Proof. Assume (a_i) is a sequence converging to a with terms in S. Our goal according to Definition 6 is to show that the sequence $\big(f(a_i)\big)$ converges to $f(a)$. To achieve this goal, we use the definition of limit in Chapter 8. So assume $\varepsilon > 0$ is given. Our goal reduces to the following goal: find an $N \in \mathbb{N}$ such that if $i \geq N$ then $|f(a_i) - f(a)| < \varepsilon$.

By assumption, f is continuous at a in the δ-ε sense, so by Definition 5 there is a $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ if $|x - a| < \delta$ and $x \in S$. Now since the sequence (a_i) converges to a, there is an $N \in \mathbb{N}$ such that if $i \geq N$ then $|a_i - a| < \delta$. (We use the definition of limit from Chapter 8, using δ for our epsilon). By the above property of δ for f, $|a_i - a| < \delta$ implies that $|f(a_i) - f(a)| < \varepsilon$.

In summary, for this choice of N, if $i \geq N$, then $|a_i - a| < \delta$. This implies in turn that $|f(a_i) - f(a)| < \varepsilon$. So this choice of N achieves our goal. $\square$

Lemma 9. *Let S be a subset of an Archimedean ordered field F. Assume that $f\colon S \to F$ is a function and that $a \in S$. If f is continuous at a in the sequential sense, then it is continuous at a in the δ-ε sense.*

Proof. We prove the contrapositive. So we suppose that f is not continuous at a in the δ-ε sense. Our goal is to show that f is not continuous at a in the sequential sense.

We negate the definition of continuity at a in the δ-ε sense. This means that there exists an $\epsilon_0 > 0$ such that for all $\delta > 0$ there is an $x \in S$ such that $|x - a| < \delta$ but $|f(x) - f(a)| \geq \varepsilon$.

We use this to define a sequence (a_i). For any positive $k \in \mathbb{N}$, let $\delta = 1/k$. By the above property there is an element $a_k \in S$ such that $|a_k - a| < 1/k$ but $|f(a_k) - f(a)| \geq \varepsilon_0$.

The first claim is that (a_i) has limit a. To see this, let $\varepsilon > 0$ be given. (This ε is independent of the ε_0 above.) Since F is Archimedean, there is an $N \in \mathbb{N}$ such that $1/N < \varepsilon$ (see Chapter 8). If $i \geq N$ then

$$|a_i - a| < \frac{1}{i} \leq \frac{1}{N} < \varepsilon.$$

This (a_i) converges to a as desired.

The second claim is that $\big(f(a_i)\big)$ does not converge to $f(a)$. This follows from the fact that $|f(a_i) - f(a)| \geq \varepsilon_0$. In other words, for this particular epsilon value, there is no N' such that $i \geq N'$ implies $|f(a_i) - f(a)| < \varepsilon_0$.

If we combine the two claims, we see that f cannot be continuous at a in the sense of Definition 6. □

We now combine the above lemmas to give the following:

Theorem 10. *Let S be a subset of an Archimedean ordered field F. Assume that $f\colon S \to F$. Then f is continuous at $a \in S$ according to the sequential definition if and only if it is continuous at $a \in S$ according to the δ-ε definition. Therefore, f is continuous on S according to the sequential definition if and only if it is continuous on S according to the δ-ε definition.*

9.4 Intermediate value theorem

One of the key foundational results in real analysis is the intermediate value theorem. It can be proved from the completeness property.

First we remind the reader of a result proved in Chapter 8, which we restate here for the convenience of the reader.

Theorem 11. *Suppose S is a nonempty subset of an Archimedean ordered field F, and suppose that S has a supremum M. Then M is the limit of a sequence (a_i) of elements $a_i \in S$ and is the limit of a sequence (b_i) of elements $b_i \notin S$. If $a < M$ is given we can assume that each a_i is in the*

interval $[a, M]$. *Similarly, if* $b > M$ *is given we can assume that each* b_i *is in the interval* $[M, b]$.

This gives us the main tool that we need to prove the following. (In the following, C is between A and B. This is intended to include the case where C is A or B itself.)

Theorem 12 (Intermediate Value Theorem)**.** *Let* F *be a complete ordered field that contains* $\mathbb{Q}$ *as an ordered subfield. Let* $[a, b]$ *be a closed interval in* F *where* $a < b$ *are elements of* F. *Suppose* $f\colon [a, b] \to F$ *is continuous. If* $C \in F$ *is any value between* $A = f(a)$ *and* $B = f(b)$ *then there is an element* $c \in [a, b]$ *such that* $f(c) = C$.

Proof. Without loss of generality we can assume $A < C < B$ (the case where $B < C < A$ is similar, and the case where C is A or B is trivial). Define the following set

$$S \stackrel{\text{def}}{=} \{u \in [a, b] \mid f(u) \leq C\}.$$

Observe that S is nonempty since $a \in S$, and that b is an upper bound for S. Since F is complete, the set S has a supremum, call it c. We claim that c satisfies the conclusion of the theorem.

We begin by showing $f(c) \leq C$. By Theorem 11 there is a sequence (a_i) of elements in S that converges to c. Since $a_i \in S$ we have $f(a_i) \leq C$. Since f is continuous, the sequence $(f(a_i))$ converges to $f(c)$. Thus

$$f(c) = \lim_{i \to \infty} f(a_i) \leq C.$$

Next we show that c is in the interval $[a, b)$. (For the sake of the theorem, we only need $c \in [a, b]$, but we need $c < b$ below in the proof.) Observe that $a \leq c$ since $a \in S$ and c is an upper bound for S. Since b is an upper bound of S, we have $c \leq b$ because c is the least upper bound of S. Finally, we have $b \neq c$ since $f(b) = B > C$ and $f(c) \leq C$ (established above). So $a \leq c < b$.

Finally we show $f(c) \geq C$, which with the above will yield the result. By Theorem 11 there is a sequence (b_i) of elements not in S converging to c, and since $c < b$ we can also assume that each $b_i \in [c, b]$. In particular, each b_i is in the interval $[a, b]$ since $[c, b]$ is a subset of $[a, b]$. Since $b_i \notin S$ we have $f(b_i) \geq C$. (Actually $f(b_i) > C$, but we just need $\geq$). Since f is continuous, the sequence $(f(b_i))$ converges to $f(c)$. Thus

$$f(c) = \lim_{i \to \infty} f(b_i) \geq C.$$

□

As an application of this theorem, we show that square roots exists in complete ordered fields.

Corollary 13. *Suppose $C \in F$ where F is a complete ordered field that contains $\mathbb{Q}$ as an ordered subfield. If $C \geq 0$, then there is an element $c \in F$ such that $c^2 = C$.*

Proof. If $C < 1$, let $b = 1$, but if $C \geq 1$, let $b = C$. Consider the function $f\colon F \to F$ defined by $x \mapsto x^2$. This function is continuous since it is the product of the identity function with itself, and continuous functions are closed under products. So the restriction $f\,|_{[0,b]}$ to $[0, b]$ is also continuous.

Claim: $f(0) \leq C \leq f(b)$. To see $f(0) \leq C$, combine the facts that $C \geq 0$ and $f(0) = 0$. To show $C \leq f(b) = b^2$ we divide into two cases. First consider the case where $C < 1$, and $b = 1$. So $f(b) = 1^2 = 1$. Thus $C \leq f(b)$ as desired. Next assume that $C \geq 1$. Then

$$C = C \cdot 1 \leq C \cdot C = b \cdot b = f(b).$$

The hypotheses of the Intermediate Value Theorem are satisfied. By the Intermediate Value Theorem, there is a $c \in [0, b]$ such that $f(c) = C$. In other words, $c^2 = C$ as desired. □

Exercise 7. Suppose $C \geq 0$ in a complete ordered field F. Show that there is $c \in F$ such that $c^3 = C$. What if $C < 0$?

Corollary 14. *The field $\mathbb{Q}$ is not complete.*

Proof. In Chapter 7 we proved that there is no $r \in \mathbb{Q}$ such that $r^2 = 2$. However, if $\mathbb{Q}$ were complete, then there would be such an r by the previous corollary. □

9.5 Cauchy sequences

If a sequence converges, then the terms of the sequence get and stay arbitrarily close to each other. This is shown in the following theorem. Such sequences are called *Cauchy sequences*. Cauchy sequences will play a key role in our construction of $\mathbb{R}$. In addition, they play an important role in analysis quite generally. Informally, a Cauchy sequence is a sequence that seems like it "ought to converge". It might not actually converge in incomplete ordered fields, though. For example, not every Cauchy sequence in $\mathbb{Q}$ converges in $\mathbb{Q}$. However, in a complete ordered field, every Cauchy sequences will converge.

Theorem 15. *Suppose (a_i) is a convergent sequence in an ordered field F. Then for all positive ε in F there is an $N \in \mathbb{N}$ such that for all $i, j \in \mathbb{N}$*

$$i, j \geq N \quad \Rightarrow \quad |a_i - a_j| < \varepsilon.$$

Proof. Since we assume that (a_i) converges, it has a limit. Let b be the limit of the sequence (a_i).

Let $\varepsilon > 0$ be an arbitrary positive element of F. We must find a $N \in \mathbb{N}$ that satisfies the statement of the theorem.

Let $\varepsilon' = \varepsilon/2$. By the definition of limit there is a $N \in \mathbb{N}$ with $|a_i - b| < \varepsilon'$ for all $i \geq N$. So, for $i, j \geq N$ we have

$$|a_i - a_j| = |(a_i - b) + (b - a_j)| \leq |a_i - b| + |b - a_j| < \varepsilon' + \varepsilon'.$$

Here we have used the triangle inequality. Since $2\varepsilon' = \varepsilon$ we conclude that $|a_i - a_j| < \varepsilon$. Thus N has the desired property. □

The above theorem says that all convergent sequences satisfy the following definition:

Definition 7 (Cauchy sequence)**.** Suppose (a_i) is an infinite sequence in an ordered field F. We say that (a_i) is *Cauchy* if the following occurs: for all positive ε in F there is a $N \in \mathbb{N}$ such that for all $i, j \in \mathbb{N}$

$$i, j \geq N \quad \Longrightarrow \quad |a_i - a_j| < \varepsilon.$$

Remark 1. We can reinterpret Theorem 15 through its contrapositive: *if a sequence is not Cauchy, it cannot converge.*

Is the converse true? In other words, do all Cauchy sequences converge? The answer is no for $F = \mathbb{Q}$. The problem with $\mathbb{Q}$ is that it has 'holes'. For example, we saw that there is no $r \in \mathbb{Q}$ with $r^2 = 2$. Define a sequence by the rule $a_i = n_i/10^i$ where n_i is the largest integer such that $a_i^2 < 2$. This sequence will not be convergent in $\mathbb{Q}$, but can be shown to be Cauchy. Even though this Cauchy sequence does not converge in $\mathbb{Q}$, it will turn out that it is convergent in $\mathbb{R}$, and has limit $\sqrt{2}$.

Informal Exercise 8. Find the first five terms of (a_i) defined in the above remark. Assume the index set is the set of $i \geq 0$. In other words, start with $i = 0$. Hint: punch $\sqrt{2}$ into your calculator.

Our approach in the next chapter will be to assume that all Cauchy sequences in $\mathbb{Q}$ should determine a real number. Non-Cauchy sequences cannot possibly converge, so should not determine real numbers. There is a problem: different sequences can determine the same real number. For example, the sequence defined by the rule $b_i = n_i/2^i$ where n_i is the largest integer such that $b_i^2 < 2$ determines the same real number as the sequence (a_i) discussed above (in fact, they both determine $\sqrt{2}$: the sequence (a_i) is related to the decimal expansion of $\sqrt{2}$ and (b_i) is related to the base 2 expansion of $\sqrt{2}$). How do we tell if two sequences determine the same number? We can use the equivalence relation defined in Chapter 8: two Cauchy sequence determine the same real number if and only if $(a_i) \sim (b_i)$. We will make this approach more precise in Chapter 10. When we do, we will need the following.

Theorem 16. *Let F be an ordered field. If $(a_i) \sim (b_i)$ and if (a_i) is Cauchy, then (b_i) is Cauchy.*

Exercise 9. Prove the above theorem. The proof is similar to the proof of the claim that if $(a_i) \sim (b_i)$ and if (a_i) converges, then (b_i) converges. You might wish to choose $\varepsilon' = \varepsilon/3$. The key step of the proof is

$$|b_i - b_j| = |(b_i - a_i) + (a_i - a_j) + (a_j - b_j)| < \varepsilon' + \varepsilon' + \varepsilon'.$$

We conclude with a lemma that shows that every Cauchy sequence is bounded.

Lemma 17. *If $(a_i)_{i \geq n_0}$ is a Cauchy sequence in an ordered field F, then there is a bound $B \in F$ such that $|a_i| \leq B$ for all $i \geq n_0$.*

Proof. First we show that $a_i \leq B_1$ for some positive upper bound B_1.

Since (a_i) is Cauchy, there is a $N \in \mathbb{N}$ such that $|a_i - a_j| < 1$ for all $i, j \geq N$ (choose $\varepsilon = 1$). Let A be the maximum of $0, a_{n_0}, \ldots, a_N$, and let $B_1 = A + 1$. Since $A \geq 0$ we have B_1 positive. We will show that B_1 is in fact an upper bound for (a_i).

First consider the case where $i \leq N$. In this case

$$a_i \leq A < A + 1.$$

Since $B_1 = A + 1$, we have $a_i \leq B_1$ as desired.

Next consider the case where $i > N$. Since $i, N \geq N$, we have the inequality $|a_i - a_N| < 1$. Thus $-1 < a_i - a_N < 1$. So

$$a_i < a_N + 1 \leq A + 1 = B_1,$$

and we get $a_i \leq B_1$ in this case as well.

The proof of the existence of a negative lower bound is similar. (Subtract one from a mininum). Write the lower bound as $-B_2$ where B_2 is positive. So we get

$$-B_2 \leq a_i \leq B_1$$

for all $i \geq n_0$. Let B be the maximum of B_1 and B_2. Then

$$-B \leq a_i \leq B.$$

So $|a_i| \leq B$ as desired. □

9.6 Cauchy criterion for completeness

Our goal is to prove the following theorem. We will need this later to show that our construction of $\mathbb{R}$ gives a complete ordered field. The converse is true as well, and will be proved in a later section.

Theorem 18 (Cauchy criterion)**.** *Let F be an Archimedean ordered field. If every Cauchy sequence in F converges in F then F is complete.*

The proof of this uses the notion of an *ε-almost-supremum* We first build a Cauchy sequence out of such "almost-Sups", and the limit can be shown to be the actual supremum.

Recall the definition (from Chapter 8): let S be a nonempty subset of an ordered field F, and let $\varepsilon > 0$ be in F. An *ε-almost-supremum* A of S is an upper bound of S such that there is an $x \in S$ in the interval $(A - \epsilon, A]$. In Chapter 8 we proved the following (which we restate for convenience):

Theorem 19. *Let S be a nonempty subset of an Archimedean ordered field F, and let $\varepsilon > 0$ be in F. If S is bounded from above, then S has an ε-almost-supremum.*

Using this theorem we can prove the Cauchy criterion:

Proof of Theorem 18. Let S be a nonempty subset of F with an upper bound. Our goal is to show that S has a supremum. This will show F is complete as desired.

For each positive integer $n \in \mathbb{N}$, let A_n be an $1/n$-almost-supremum of S. This exists by Theorem 19 (proved in Chapter 8).

Claim: (A_i) is a Cauchy sequence. Let $\varepsilon > 0$ be given. To prove the claim we need to find an N such that if $i, j \geq N$ then $|A_i - A_j| < \varepsilon$. Since F is an Archimedean ordered field there is an N such that $1/N \leq \varepsilon$. Now suppose $i, j \geq N$. Without loss of generality, suppose $A_i \geq A_j$. Since A_i is an $1/i$-almost-supremum, $A_i - 1/i$ is not an upper bound of S. Since A_j is an upper bound of S, we have $A_i - 1/i < A_j$. Hence

$$|A_i - A_j| = A_i - A_j < 1/i \leq 1/N \leq \varepsilon.$$

Thus (A_i) is Cauchy. By the assumption of the theorem, (A_i) has a limit, call it A.

Claim: A is an upper bound of S. To see this we must show $x \leq A$ for all $x \in S$. Suppose otherwise, that $x > A$ for some $x \in S$. Let $\varepsilon = x - A$. Since A is the limit of (A_i), there is an N such that

$$i \geq N \implies |A_i - A| < \varepsilon.$$

In particular, $|A_N - A| < \varepsilon$. Observe that $A < x \leq A_N$ since A_N is an upper bound of S. In particular, $A < A_N$ so

$$A_N - A = |A_N - A| < \varepsilon = x - A.$$

This implies that $A_N < x$, a contradiction since A_N is an upper bound of S.

Claim: for all $\varepsilon > 0$ there is an $x \in S$ such that $|A - x| < \varepsilon$. To show this, fix $\varepsilon > 0$. Since F is an Archimedean ordered field, there is a positive

integer N such that $1/N \leq \varepsilon/2$ (Chapter 8). Since A is the limit of (A_i) there is a positive integer N' such that $|A_i - A| < \varepsilon/2$ for all $i \geq N'$. Let n be the maximum of N and N'. Since A_n is an $1/n$-almost supremum, there is an $x \in S$ with $|A_n - x| < 1/n$. So

$$|A - x| \leq |A - A_n| + |A_n - x| < \frac{\varepsilon}{2} + \frac{1}{n} \leq \frac{\varepsilon}{2} + \frac{1}{N} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

We know that A is an upper bound, but is it the least upper bound? We show that A is indeed the supremum of S showing that if $A' < A$ then A' is not an upper bound. To do so, let $\varepsilon = A - A'$. By the above claim, there is an $x \in S$ such that $|A - x| < \varepsilon$. (Note $A - x = |A - x|$ since A is an upper bound of S). So

$$A - x = |A - x| < \varepsilon = A - A'.$$

This implies $A' < x$, so A' is not an upper bound. □

9.7 Bounded monotonic sequences converge

Monotonic sequences are commonly used in mathematics and are often easier to deal with than arbitrary sequences. One of the most useful basic facts about $\mathbb{R}$ is that every bounded monotonic sequences converges. We will show that this follows from the completeness property.

Definition 8 (Monotonic)**.** Let $(a_i)_{i \geq n_0}$ be a sequence in an ordered field F. The sequence is said to be *increasing* if $a_{i+1} \geq a_i$ for all $i \geq n_0$. The sequence is *decreasing* if $a_{i+1} \leq a_i$ for all $i \geq n_0$. In either case (a_i) is said to be *monotonic*. (Observe that a constant sequence is considered both increasing and decreasing).

The sequence $(a_i)_{i \geq n_0}$ is said to be *strictly increasing* if $a_{i+1} > a_i$ for all $i \geq n_0$. The sequence $(a_i)_{i \geq n_0}$ is *strictly decreasing* if $a_{i+1} < a_i$ for all $i \geq n_0$. In either case (a_i) is said to be *strictly monotonic*. (Observe that a constant sequence is monotonic, but not strictly monotonic)

The following is a simple consequences of the definition. It is stated for increasing sequences, but the statement holds, with the obvious modifications, for decreasing sequence. There are obvious versions for strictly monotonic sequences as well.

Lemma 20. *Suppose that $(a_k)_{k \geq n_0}$ is a monotonically increasing sequence in an ordered field F. If $j \geq i \geq n_0$ then $a_j \geq a_i$.*

Proof. Fix $i \geq n_0$, and consider the set $S_i = \{u \in \mathbb{Z} \mid u \geq n_0 \text{ and } a_u \geq a_i\}$ we will show by induction that all $j \geq i$ are in S_i.

For the base case, observe that $a_i \geq a_i$ (reflexive). Thus $i \in S_i$.

Now suppose $k \in S_i$. This implies $a_{k+1} \geq a_k \geq a_i$ (the first inequality by Definition 8, the second since $k \in S_i$). So $k + 1 \in S_i$.

By induction, S_i contains all $j \geq i$. In particular $a_j \geq a_i$ if $j \geq i$. □

An increasing sequence is automatically bounded from below: if a_{n_0} is the first term of such a sequence then the above lemma shows a_{n_0} is a lower bound. So to say that such a sequence is *bounded* really means that it is also bounded from above. Obviously the same idea, but reversed, applies to decreasing sequences.

Theorem 21 (Convergence of bounded monotonic sequences)**.** *Let (a_i) be a bounded monotonic sequence in a complete field F. Then (a_i) converges.*

Proof. We assume that $(a_i)_{i \geq n_0}$ is an increasing sequence. The decreasing case is similar. Since F is complete, and the set $\{a_i \mid i \geq n_0\}$ is bounded, this set has a supremum B. We will show that B is in fact the limit.

Let $\varepsilon > 0$ be given. By a result in Chapter 8, since B is the supremum, the interval $(B - \varepsilon, B]$ must contain an element of $\{a_i \mid i \geq n_0\}$. In other words, there is an $N \in \mathbb{N}$ such that $B - \varepsilon < a_N \leq B$. We will show that N has the desired property (as in the definition of limit). So suppose $i \geq N$. Then $a_N \leq a_i$ since the sequence is increasing. But B is an an upper bound for the sequence. So $a_i \leq B$. Thus

$$B - \varepsilon < a_N \leq a_i \leq B.$$

Observe $|B - a_i| = (B - a_i)$ since $a_i \leq B$. Observe also that $B - a_i < \varepsilon$ since $B - \varepsilon < a_i$. Therefore, $|B - a_i| < \varepsilon$ as desired. □

Exercise 10 (Optional)**.** Use ε-almost suprema (and infima) from Chapter 8 to show that if F is an Archimedean ordered field, then every bounded monotonic sequence is Cauchy. (Even if F is not complete). Hint: it is enough to do the increasing case. The proof is similar to that of the previous theorem. For each $n \in \mathbb{N}$, let B_n be an $1/n$-almost supremum. Now given an arbitrary $\varepsilon > 0$, choose $1/n < \varepsilon$. Show that you can choose an $N \in \mathbb{N}$ such that $a_N \in (B_n - 1/n, B_n]$. Show that $i, j \geq N$ implies $|a_i - a_j| < 1/n < \varepsilon$.

9.8 Accumulation points (optional)

Consider the sequence defined by the equation $a_k = (-1)^k$. It is obviously not a Cauchy sequences, so it cannot converge. (Recall that all convergent sequences must be Cauchy). However, it does have an infinite number of terms equal to 1 and an infinite number of terms equal to -1. So 1 an -1 are in some sense limits in a more general sense. A similar phenomenon occurs for the sequence defined by $b_k = 1/k + (-1)^k$; this sequence seems to have values that "accumulate" near both 1 and -1. This leads to an important concept which we now define:

Definition 9 (Accumulation point)**.** Let (a_i) be a sequence in an ordered field F. We say that the element $b \in F$ is an *accumulation point* of (a_i) if the following occurs: for all $\varepsilon > 0$ and $n \in \mathbb{N}$ there is an integer $i \geq n$ such that $|a_i - b| < \varepsilon$.

Remark 2. We can rephrase the above condition to an equivalent condition: *for all* $\varepsilon > 0$ *there are an infinite number of* $i \in \mathbb{N}$ *such that* $|a_i - b| < \varepsilon$. Note that there are two ways for a value b to be an accumulation point of the sequence: either the value b itself occurs an infinite number of times in the sequence, or there are terms of the sequence that get *arbitrarily close* to b.

Exercise 11. Let F be an Archimedean ordered field. Show that 1 is an accumulation point of the sequence defined by $a_k = (-1)^k$ and of the sequence defined by $b_k = 1/k + (-1)^k$.

Exercise 12. If a sequence has a limit, show that its limit is the unique accumulation point.

Exercise 13. From the previous exercise we see that a convergent sequence has exactly one accumulation point. Is the converse true? In other words, if a sequence has exactly one accumulation point, does it follow that the sequence converges?

The following will be useful later in showing the converse of Theorem 18.

Theorem 22. *Let* (a_i) *be a Cauchy sequence in an ordered field* F*. If* a *is an accumulation point of* (a_i) *then* a *is the limit of* (a_i)*.*

Proof. By the definition of limit, for any given $\varepsilon > 0$ we must show that there is an $N \in \mathbb{N}$ such that $|a_i - a| < \varepsilon$ for all $i \geq N$.

So suppose $\varepsilon > 0$ is given. Let $\varepsilon' = \varepsilon/2$. Since (a_i) is a Cauchy sequence, there is an N such that $|a_i - a_j| < \varepsilon'$ for all $i, j \geq N$. By the definition of accumulation point, there is an $j_0 \geq N$ such that $|a_{j_0} - a| < \varepsilon'$. Now assume that $i \geq N$. Then

$$|a_i - a| \leq |a_i - a_{j_0}| + |a_{j_0} - a| < \varepsilon' + \varepsilon' = \varepsilon.$$

□

Exercise 14. Show that if $(a_i) \sim (b_i)$ are equivalent sequences, then they have the same accumulation points. Is the converse true?

9.9 Lim infs and lim sups (optional)

We will see that in any complete ordered field, bounded sequence (bounded in both direction) must have accumulation points. In this case there is a greatest accumulation point call the *superior limit* and a least accumulation point called the *inferior limit*.

The superior limit, often called lim sup, is formed by seeing how the bounds to the sequence change as a larger and larger number of terms are removed from the *beginning* of the sequence. These bounds will tend to go down, or stay the same, as more terms are removed. What happens in the long term, as measured by the infimum of these bounds, is the *superior limit*. The *inferior limit*, often called lim inf, is formed in a similar way except with lower bounds. The following definition makes this idea precise.

Definition 10 (Lim sup and lim inf)**.** Suppose $(a_i)_{i \geq n_0}$ is a bounded sequence in a complete ordered field F. In other words, suppose there is a bound B such that $|a_i| \leq B$ for all $i \geq n_0$. For each $k \geq n_0$ consider the following set

$$S_k = \{a_i \mid i \geq k\}.$$

Observe that S_k is a nonempty set with upper bound B and lower bound $-B$. Let M_k be the supremum of S_k and let m_k be the infimum of S_k. These exist since F is complete. Observe that $-B \leq m_k \leq M_k \leq B$, so the sets $\{m_k \mid i \geq k\}$ and $\{M_k \mid i \geq k\}$ are themselves bounded by $-B$ and B. The superior limit is defined as follows:

$$\limsup_{i \to \infty} a_i \stackrel{\text{def}}{=} \inf\{M_k \mid k \geq n_0\}.$$

The inferior limit is defined as follows:

$$\liminf_{i \to \infty} \stackrel{\text{def}}{=} \sup\{m_k \mid k \geq n_0\}.$$

These both exist since F is complete.

Remark 3. Often lim sups and lim infs are defined even for unbounded sequences. For example, if a sequence (a_i) has no upper bound then the lim sup of (a_i) is sometimes said to be ∞. Similarly, if (a_i) has no lower bound then the lim inf of (a_i) would be $-\infty$. In this book, however, we will stick to bounded sequences where lim sup and lim inf are elements of F (assuming F is complete).

Lemma 23. *Let (a_i) be a bounded sequence in a complete ordered field F. Suppose $M \in F$ is such that*

$$M > \limsup_{i \to \infty} a_i.$$

Then all but a finite number of terms of (a_i) are strictly smaller than M. In other words, there is an $N \in \mathbb{N}$ such that if $i \geq N$ then $a_i < M$.

Proof. Let $X = \limsup_{i \to \infty} a_i$. By definition X is the greatest lower bound of the M_k in the above definition. Since $M > X$, observe that M is not a lower bound of $\{M_k\}$. Thus there is an $N \in \mathbb{N}$ such that $M_N < M$. Since M_N is the least upper bound of $S_N = \{a_i \mid i \geq N\}$, we have $a_i \leq M_N$ for all $i \geq N$. Since $M_N < M$ we have $a_i < M$ for all $i \geq N$. □

We can give an alternate characterization of lim sups that is sometimes easier to work with than Definition 10.

Theorem 24. *Let (a_i) be a bounded sequence in a complete ordered field F. Then $\limsup_{i\to\infty} a_i$ is the minimal element $X \in F$ with the following property: For all $M > X$ there is an $N \in \mathbb{N}$ such that if $i \geq N$ then $a_i < M$. In other words, $\limsup_{i\to\infty} a_i$ is the smallest element of F such that any larger element is a strict upper bound for all but a finite number of terms of (a_i).*

Proof. The above lemma shows that $X = \limsup_{i\to\infty} a_i$ has the desired property. Now we need to show that no smaller element Y has the property. Suppose $Y < X$ has the property. Let Z be chosen so that $Y < Z < X$. By the property assumed for Y there is an $N \in \mathbb{N}$ such that if $i \geq N$ then $a_i < Z$. This implies that Z is an upper bound of the set $S_N = \{a_i \mid i \geq N\}$. So $M_N \leq Z$ where M_N is the supremum of S_N. Since X is defined to be the infimum of the set of such M_k we have $X \leq M_N$. Now we get the following contradiction:

$$X \leq M_N \leq Z < X.$$

□

For lim infs we have the following:

Theorem 25. *Let (a_i) be a bounded sequence in a complete ordered field F. Then $\liminf_{i\to\infty} a_i$ is the maximal element $x \in F$ with the following property: For all $m < x$ there is an $N \in \mathbb{N}$ such that if $i \geq N$ then $a_i > m$. In other words, $\liminf_{i\to\infty} a_i$ is the largest element of F such that any smaller element is a strict lower bound for all but a finite number of terms of (a_i).*

Exercise 15. How would you change the proof of Lemma 23 and Theorem 24 in order to prove Theorem 25?

We give a third characterization of lim sups and lim infs in terms of accumulation points. This is perhaps the simplest way of describing them.

Theorem 26. *Suppose (a_i) is a bounded sequence in a complete ordered field F. Then $\limsup_{i\to\infty} a_i$ is an accumulation point of (a_i). In fact, it is the greatest accumulation point of (a_i).*

Proof. Let $X = \limsup_{i\to\infty} a_i$. We will first show that X is an accumulation point. We assume we are given $\varepsilon > 0$ and $n \in \mathbb{N}$. According to the definition of accumulation point, our goal is to show that there is an $i \geq n$ such that $|a_i - X| < \varepsilon$.

We have $X + \varepsilon > X$. So by Theorem 24 there is an $N \in \mathbb{N}$ such that $i \geq N$ implies $a_i < X + \varepsilon$. Let k be the maximum of n and N. As in the definition of lim sup, let $S_k = \{a_i \mid i \geq k\}$ and let M_k be the supremum of S_k.

By definition of lim sup, we have $X \leq M_k$, so $X - \varepsilon < M_k$. Observe that $X - \epsilon$ is not an upper bound of S_k (since M_k is the least upper bound).

So there is an $a_i \in S_k$ with $X - \varepsilon < a_i$. Since $i \geq k \geq N$, we have $a_i < X + \varepsilon$. So

$$X - \varepsilon < a_i < X + \varepsilon.$$

This implies $|a_i - X| < \varepsilon$. Observe also that $i \geq k \geq n$. This concludes the proof that $X = \limsup_{i\to\infty} a_i$ is an accumulation point.

Now suppose that $Y > X$. We must show that Y is not an accumulation point. Let $Z \in F$ be such that $X < Z < Y$. By Theorem 24, there is an $N \in \mathbb{N}$ such that if $i \geq N$ then $a_i < Z$. For $i \geq N$ we use $a_i < Z < Y$ to conclude

$$|a_i - Y| = Y - a_i = (Y - Z) + (Z - a_i) \geq (Y - Z).$$

So if $\varepsilon = Y - Z$ then $|a_i - Y| \geq \varepsilon$ for all $i \geq N$. This shows that Y is not an accumulation point of (a_i). □

The proof of the following is similar to the proof of the above theorem, so we omit it.

Theorem 27. *Suppose (a_i) is a bounded sequence in a complete ordered field F. Then $\liminf_{i\to\infty} a_i$ is an accumulation point of (a_i). In fact, it is the least accumulation point of (a_i).*

Corollary 28. *Every bounded sequence in a complete ordered field F. has an accumulation point.*

Corollary 29. *Suppose (a_i) is a bounded sequence in a complete ordered field F. Then*

$$\liminf_{i\to\infty} a_i \leq \limsup_{i\to\infty} a_i.$$

Furthermore (a_i) converges if and only if equality hold. In this case,

$$\lim_{i\to\infty} a_i = \liminf_{i\to\infty} a_i = \limsup_{i\to\infty} a_i.$$

Proof. Since $\lim_{i\to\infty} a_i$ is the least accumulation point and $\limsup_{i\to\infty} a_i$ is the greatest accumulation point, we have

$$\liminf_{i\to\infty} a_i \leq \limsup_{i\to\infty} a_i.$$

By Exercise 12, if the sequence (a_i) has a limit, that limit is the unique accumulation point, so

$$\lim_{i\to\infty} a_i = \liminf_{i\to\infty} a_i = \limsup_{i\to\infty} a_i.$$

Finally, suppose equality holds, and let X be its value:

$$\liminf_{i\to\infty} a_i = \limsup_{i\to\infty} a_i = X.$$

We need to show that X is the limit of (a_i). Let $\varepsilon > 0$ be given. We will find an $N \in \mathbb{N}$ such that $|a_i - X| < \varepsilon$ for all $i \geq N$. First use Theorem 24 to obtain an $N_1 \in \mathbb{N}$ such that if $i \geq N_1$ then $a_i < X + \varepsilon$. Use Theorem 25 to get an $N_2 \in \mathbb{N}$ such that if $i \geq N_2$ then $a_i > X - \varepsilon$. Thus if $i \geq N$ where N is the maximum of N_1 and N_2, then

$$X - \varepsilon < a_i < X + \varepsilon.$$

In particular, $|a_i - X| < \varepsilon$ as desired.

□

Remark 4. The above shows that for *bounded* sequences, convergence is equivalent to the existence of exactly one accumulation point.

9.10 Cauchy sequences converge (optional)

Earlier we showed that if F is an Archimedean ordered field such that every Cauchy sequences converges, then F is complete. Now we show the converse.

Theorem 30. *If F is a complete ordered field, then every Cauchy sequence converges.*

Proof. Let (a_i) be a Cauchy sequence in a complete ordered field F. By Lemma 17 the sequence (a_i) is bounded. By Theorem 26 the lim sup yields an accumulation point for (a_i). So, by Theorem 22, (a_i) converges. □

This gives another characterization of *completeness* for Archimedean ordered fields.

Corollary 31. *Let F be an Archimedean ordered field. Then F is complete if and only if every Cauchy sequence converges.*

Chapter 10

Constructing the Real Numbers

In this chapter we introduce the field of real numbers $\mathbb{R}$. There are several ways to introduce the real numbers. Three popular approaches are to introduce $\mathbb{R}$ with (i) new axioms, with (ii) Dedekind cuts of $\mathbb{Q}$, or with (iii) Cauchy sequences in $\mathbb{Q}$. We will use the third approach and construct real numbers as equivalence classes of Cauchy sequences of rational numbers. This approach is chosen since it avoids the need for additional axioms by building on the previously developed number systems, and it gives students practice with sequences in general and Cauchy sequences in particular.

The main theorem of this chapter is that $\mathbb{R}$, as constructed from Cauchy sequences, is a complete ordered field.

10.1 The real numbers

Our idea for constructing $\mathbb{R}$ is based on two intuitive principles: (1) every Cauchy sequence in $\mathbb{Q}$ should determine a real number, and (2) equivalent sequences should determine the same real number. The second principle can be reexpressed as the requirement that every Cauchy sequence in an equivalence class $\big[(a_i)\big]$ should determine the same real number. This idea leads us to the idea that real numbers correspond to equivalence classes $\big[(a_i)\big]$ of Cauchy sequences.

Finally we take one more conceptual step: real numbers don't merely *correspond* to equivalence classes of Cauchy sequence, but can be *defined* as equivalence classes of Cauchy sequences. In other words, if we wish to

construct the real numbers then they have to be defined somehow, why not define them via this intuitive correspondence?[1]

Definition 1 (Real number)**.** If (a_i) is a Cauchy sequence in $\mathbb{Q}$, then let $\big[(a_i)\big]$ be the equivalence class containing (a_i) under the equivalence relation $\sim$ on the set of sequences in $\mathbb{Q}$. We call $\big[(a_i)\big]$ a *real number*.

Definition 2. The set of real numbers $\mathbb{R}$ is defined as follows:

$$\mathbb{R} \stackrel{\text{def}}{=} \left\{ \big[(a_i)\big] \;\middle|\; (a_i) \text{ is a Cauchy sequence in } \mathbb{Q} \right\}.$$

In order to make $\mathbb{R}$ into a field we need to define an addition and multiplication operation on $\mathbb{R}$.

Definition 3 (Addition and multiplication)**.** Let $\big[(a_i)\big]$ and $\big[(b_i)\big]$ be real numbers. Then

$$\big[(a_i)_{i\geq n_0}\big] + \big[(b_i)_{i\geq m_0}\big] \stackrel{\text{def}}{=} \big[(a_i + b_i)_{i\geq l_0}\big]$$

and

$$\big[(a_i)_{i\geq n_0}\big] \cdot \big[(b_i)_{i\geq m_0}\big] \stackrel{\text{def}}{=} \big[(a_i b_i)_{i\geq l_0}\big].$$

Here l_0 is the maximum of n_0 and m_0. Our definitions give two binary operations $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

In order to check that these definitions are well-defined we need to verify three facts that are not totally obvious: (i) $(a_i + b_i)$ and $(a_i b_i)$ are Cauchy sequences, (ii) if $(a_i') \sim (a_i)$ then we can replace (a_i) with (a_i') in the definition and the resulting sum and product will give the same real number, and (iii) if $(b_i') \sim (b_i)$ then we can replace (b_i) with (b_i') in the definition and the resulting sum and product will give the same real number.

The remainder of this section will be devoted to verifying the facts needed to confirm that the definition is well-defined.

Lemma 1. *Suppose that $(a_i)_{i\geq n_0}$ and $(b_i)_{i\geq m_0}$ are Cauchy sequences in an ordered field F. Then $(a_i + b_i)_{i\geq l_0}$ and $(a_i b_i)_{i\geq l_0}$ are also Cauchy. Here l_0 is the maximum of n_0 and m_0.*

Proof. First we prove the result for products, and leave the easier sum case to the reader.

Let $\varepsilon > 0$. We must find a suitable N. Recall that Cauchy sequences are bounded (Chapter 9), so there is a bound A such that $|a_i| \leq A$ for all terms a_i of the first sequence. Likewise, there is a bound B such that $|b_i| \leq B$

[1] As mentioned above, there is another intuitive correspondence with Dedekind cuts of rational numbers. So we could also define real numbers as Dedekind cuts. The Cauchy sequence approach and the Dedekind cut approach lead to isomorphic ordered fields, so from the mathematical point of view it does not matter which approach is followed.

for all terms b_i of the second sequence. Clearly we can assume that A and B are chosen to be positive.

Let $\varepsilon_1 = \varepsilon/(2B)$. Since (a_i) is Cauchy, there is an integer N_1 such that $i, j \geq N_1$ implies $|a_i - a_j| < \varepsilon_1$. Similarly, if $\varepsilon_2 = \varepsilon/(2A)$, there is an integer N_2 such that $i, j \geq N_2$ implies $|b_i - b_j| < \varepsilon_2$. Let N be the maximum of N_1 and N_2. If $i, j \geq N$, then

$$\begin{aligned} |a_i b_i - a_j b_j| &= |a_i b_i - a_i b_j + a_i b_j - a_j b_j| \\ &\leq |a_i b_i - a_i b_j| + |a_i b_j - a_j b_j|. \end{aligned}$$

Observe that

$$|a_i b_i - a_i b_j| = |a_i||b_i - b_j| \leq A|b_i - b_j| < A\varepsilon_2 = \varepsilon/2.$$

Similarly,

$$|a_i b_j - a_j b_j| = |a_i - a_j||b_j| \leq |a_i - a_j|B < \varepsilon_1 B = \varepsilon/2.$$

Thus $|a_i b_i - a_j b_j| < \varepsilon/2 + \varepsilon/2 = \varepsilon$ as desired. □

Exercise 1. Finish the above proof for the case of $(a_i + b_i)_{i \geq l_0}$. Hint: for ε' just use $\varepsilon/2$.

Lemma 2. *Let $(a_i), (a_i'), (b_i), (b_i')$ be Cauchy sequences with values in an ordered field F. If $(a_i) \sim (a_i')$ then*

$$(a_i + b_i) \sim (a_i' + b_i) \qquad \text{and} \qquad (a_i b_i) \sim (a_i' b_i).$$

If $(b_i) \sim (b_i')$ then

$$(a_i + b_i) \sim (a_i + b_i') \qquad \text{and} \qquad (a_i b_i) \sim (a_i b_i').$$

Proof. We leave the (easier) case of sums to the reader. We prove the first statement for products; the second statement is similar. So we assume that $(a_i) \sim (a_j')$, and we aim to prove that $(a_i b_i) \sim (a_i' b_i)$. In other words, for each $\varepsilon > 0$ in F we aim to find a $N \in \mathbb{N}$ such that

$$i \geq N \;\Rightarrow\; |a_i b_i - a_i' b_i| < \varepsilon.$$

To find N, we use the fact that (b_i) is Cauchy, so is bounded (Chapter 9). So there is a $B \in F$ such that $|b_i| \leq B$ for all terms of the sequence. Clearly we can choose B to be positive. Let $\varepsilon' = \varepsilon/B$. Since $(a_i) \sim (a_j')$, there is an $N \in \mathbb{N}$ such that

$$i \geq N \implies |a_i - a_i'| < \varepsilon'.$$

Observe that if $i \geq N$ then

$$|a_i b_i - a_i' b_i| = |a_i - a_i'| \cdot |b_i| \;\leq\; |a_i - a_i'|B \;<\; \varepsilon' B = \varepsilon.$$

So N is as desired. □

Exercise 2. Complete the proof by proving the case of $(a_i + b_i) \sim (a'_i + b_i)$. Hint: you do not need boundedness for Cauchy sequences in that case. In fact, it is possible to just use $\varepsilon' = \varepsilon$.

Remark 1. Because of the above lemmas, we now know that addition and multiplication are well-defined operations on $\mathbb{R}$.

10.2 The finite modification lemma for sequences

By definition, a real number $x \in \mathbb{R}$ can be designated by giving a Cauchy sequence (a_i) of rational numbers. More precisely, x is the equivalence class $[(a_i)]$, but informally it is good to think of x more as the real number determined by (a_i) "in the limit". We think of each a_i as a rational approximation which gets closer to x. All of this will be made precise later in the chapter, and we will see that in some sense x is really the limit of (a_i).

This idea leads to the following useful lemma, which was already discussed in Chapter 8.

Lemma 3 (Finite modification lemma)**.** *Suppose $(a_i)_{i \geq n_0}$ and $(b_i)_{i \geq m_0}$ differ in only in a finite number of terms. In other words, suppose that there is an integer k greater than or equal to both n_0 and m_0 such that $a_i = b_i$ if $i \geq k$. Then $(a_i) \sim (b_i)$. In particular, (a_i) and (b_i) determine the same real number (in other words $[(a_i)]] = [(b_i)]$ in $\mathbb{R}$).*

Proof. Let $\varepsilon > 0$ be given. If $i \geq k$ then

$$|a_i - b_i| = |a_i - a_i| = |0| = 0 < \varepsilon.$$

This shows $(a_i) \sim (b_i)$, and so $[(a_i)] = [(b_i)]$ in $\mathbb{R}$. □

Remark 2. Because of this remark we will rarely indicate the starting point of sequences. What is important is what happens long-term after any finite number of terms. With this convention, we can define addition and multiplication without reference to the starting point:

$$[(a_i)] + [(b_i)] = [(a_i + b_i)], \qquad \text{and} \qquad [(a_i)][(b_i)] = [(a_i b_i)].$$

In the next two sections we will use constant sequences (c) where $c \in \mathbb{Q}$. This is a sequence where $c_i = c$ for all indices i. Because of the above lemma, we will not usually need to state the starting index i (but a convenient default is to start with $i = 0$).

10.3 The real numbers $\mathbb{R}$ as a commutative ring

Our next step is to prove that $\mathbb{R}$ is a commutative ring. It is a bit harder to show it is a field, and so we will postpone that for a later section.

Theorem 4. *Addition and multiplication on $\mathbb{R}$ are commutative and associative.*

Exercise 3. Prove the above theorem.

Recall that in Chapter 8 we proved that the constant sequences (c) converge to c. Since constant sequences converge, such sequences are Cauchy. So if $c \in \mathbb{Q}$, the constant sequence gives a real number $[(c)] \in \mathbb{R}$. We are particularly interest in $[(0)]$ and $[(1)]$:

Theorem 5. *An additive identity for $\mathbb{R}$ exists and is $[(0)]$. A multiplicative identity for $\mathbb{R}$ exists and is $[(1)]$.*

Remark 3. Identities, if they exist, are unique.[2] Thus we can say "*the* additive identity" and "*the* multiplicative identity" of $\mathbb{R}$.

Proof. Let $x = \big[(a_i)\big]$ be an arbitrary real number. By definition of $+$ in $\mathbb{R}$,

$$x + [(0)] = \big[(a_i)\big] + \big[(0)\big] = \big[(a_i + 0)\big] = \big[(a_i)\big] = x$$

where the next-to-last equality is due to the fact that 0 is the additive identity of $\mathbb{Q}$ (Chapter 7). By the commutative law (Theorem 4) we get

$$[(0)] + x = x + [(0)] = x.$$

Thus $[(0)]$ is the additive identity.

The proof that $[(1)]$ is the multiplicative identity is similar. □

We now consider inverses.

Lemma 6. *If (a_i) is a Cauchy sequence in an ordered field F, then $(-a_i)$ is also Cauchy.*

Proof. Suppose (a_i) is Cauchy. Let $\varepsilon > 0$ be given. In order to show that $(-a_i)$ is Cauchy, we must find an $N \in \mathbb{N}$ such that $\big|(-a_i) - (-a_j)\big| < \varepsilon$ for all $i, j \geq N$. Since (a_i) is Cauchy, there is a $N \in \mathbb{N}$ such that $|a_i - a_j| < \varepsilon$ for all $i, j \geq N$. If $i, j \geq N$ then

$$\big|(-a_i) - (-a_j)\big| = \big|(-1)(a_i - a_j)\big| = |-1||a_i - a_j| = |a_i - a_j|.$$

But $|a_i - a_j| < \varepsilon$, so $\big|(-a_i) - (-a_j)\big| < \varepsilon$ as desired. Thus $(-a_i)$ is Cauchy. □

Theorem 7. *Every element of $\mathbb{R}$ has an additive inverse. More specifically, if $x = \big[(a_i)\big]$, then $-x = \big[(-a_i)\big]$.*

Proof. Let $x \in \mathbb{R}$. Write x as $\big[(a_i)\big]$ where (a_i) is Cauchy in $\mathbb{Q}$. By Lemma 6 the sequence $(-a_i)$ is also Cauchy, so $y = \big[(-a_i)\big]$ is a real number. We leave it to the reader to show that y is the additive inverse of x. □

[2] For any binary operation on a set S, one can show that if there is an identity, it must be unique. For example, if 0 and $0'$ are additive identities, $0 = 0 + 0' = 0'$.

Exercise 4. Complete the proof of the above theorem.

As we will see, *multiplicative* inverses are trickier. Fortunately we do not need multiplicative inverses to conclude that $\mathbb{R}$ is a ring:

Theorem 8. *The real numbers $\mathbb{R}$ form a commutative ring.*

Exercise 5. Prove the above. Hint: some steps have been proved above. What laws have not been proved yet?

Now that we know that $\mathbb{R}$ is a commutative ring, we can use all the familiar algebraic manipulations and laws valid in rings.

10.4 The canonical embedding of $\mathbb{Q}$ in $\mathbb{R}$

Now that we have constructed $\mathbb{R}$ we wish to regard $\mathbb{Q}$ as a subset of $\mathbb{R}$. To do so we need to embed $\mathbb{Q}$ into $\mathbb{R}$. This will require an injective map $\mathbb{Q} \to \mathbb{R}$. What we will do is send any $r \in \mathbb{Q}$ to the constant sequence (r).

Theorem 9. *Let $b, c \in F$ where F is an ordered field. Suppose $b \neq c$. Then (b) and (c) are not equivalent sequences.*

Proof. Observe that both sequences converge. If they were equivalent then they would have to have the same limit (Chapter 8). However, the first sequence converges to b and the second to c. A contradiction. □

Corollary 10. *Let $b, c \in \mathbb{Q}$ be distinct. Then $[(b)] \neq [(c)]$ in $\mathbb{R}$.*

Proof. By the above theorem, $(b) \not\sim (c)$. So by general properties of equivalence classes $[(b)] \neq [(c)]$. □

Definition 4 (Canonical embedding)**.** The *canonical embedding* $\mathbb{Q} \to \mathbb{R}$ is the function defined by the rule $c \mapsto [(c)]$.

Theorem 11. *The canonical embedding $\mathbb{Q} \to \mathbb{R}$ is injective.*

Exercise 6. Prove the above theorem using Corollary 10.

Now that we have a canonical embedding $\mathbb{Q} \to \mathbb{R}$, and have shown that it is injective, we can use this to identify elements of $\mathbb{Q}$ with their images in $\mathbb{R}$. Thus we can think of $\mathbb{Q}$ as being a subset of $\mathbb{R}$.

We can go further. We can think of $\mathbb{Q}$ as a *subfield* of $\mathbb{R}$ (as defined in Chapter 8). To do so, we need to check that the addition and multiplication of $\mathbb{R}$ extends the addition and multiplication of $\mathbb{Q}$ defined in Chapter 7. In other words, when we are working with addition and multiplication on $\mathbb{Q}$ we want to be assured that we get the same result whether we use the addition of $\mathbb{Q}$ (Chapter 7) or the addition of $\mathbb{R}$ (this chapter). This is demonstrated in the following lemma.

Lemma 12. *The definitions of addition and multiplication on $\mathbb{R}$ extend the definitions of addition and multiplication on $\mathbb{Q}$. So $\mathbb{Q}$ is a subfield of $\mathbb{R}$.*

Proof. We give the proof for addition; the proof for multiplication is similar. Let $a, b \in \mathbb{Q}$ be given and let $+_{\mathbb{Q}}$ be the addition defined in Chapter 7. Let $+_{\mathbb{R}}$ be the addition defined in the current chapter. We must show that $a +_{\mathbb{Q}} b$ is identified with the same real number as $a +_{\mathbb{R}} b$ (via the canonical embedding).

This is actually pretty trivial once we figure out what is involved. The canonical embedding maps the rational number $a +_{\mathbb{Q}} b$ to the equivalence class of the constant sequence $(a +_{\mathbb{Q}} b)$. Since a is identified with the equivalence class of the constant sequence (a) and b is identified with the equivalence class of (b), the sum $a +_{\mathbb{R}} b$ is equal to the sum $[(a)] +_{\mathbb{R}} [(b)]$. By the definition of $+_{\mathbb{R}}$

$$[(a)] +_{\mathbb{R}} [(b)] = \big[(a +_{\mathbb{Q}} b)\big].$$

The result follows. □

Remark 4. Since 0 in $\mathbb{Q}$ is identified with the equivalence class $[(0)]$ of the constant sequence (0), and since $[(0)]$ is the additive identity of $\mathbb{R}$, we usually write 0 for the additive identity of $\mathbb{R}$. This is consistent with the practice of writing 0 for the additive identity of any ring.

Similarly, we write 1 for the multiplicative identity of $\mathbb{R}$.

Remark 5. In a similar manner, we see that additive inverse in $\mathbb{R}$ extends additive inverse in $\mathbb{Q}$. This follows from the identity $-[(r)] = [(-r)]$ proved above.

Subtraction in $\mathbb{R}$ extends the subtraction in $\mathbb{Q}$. This follows from the definition of $r - s$ (in any ring) as $r + (-s)$, and the fact that addition and additive inverse in $\mathbb{R}$ extend the operations in $\mathbb{Q}$.

10.5 The real numbers $\mathbb{R}$ as a field

Our next main step is to show that $\mathbb{R}$ is a field by showing that every $x \neq 0$ has a multiplicative inverse. Since nonzero real numbers are represented by Cauchy sequences that are not equivalent to the zero sequence (0), we begin by considering such Cauchy sequences.

Lemma 13. *Suppose (a_i) is a Cauchy sequence in an ordered field F such that $(a_i) \not\sim (0)$. Then there is $k \in \mathbb{N}$ and a positive $d \in F$ such that $|a_i| \geq d$ for all $i \geq k$.*

Proof. When we negate the definition of equivalence in the case of $(a_i) \sim (0)$ we find that there exists a positive $\varepsilon \in F$ such that for all $N \in \mathbb{N}$ there is an integer $i \geq N$ with $|a_i - 0| \geq \varepsilon$. Fix such an $\varepsilon_0 > 0$ for what follows.

Let $\varepsilon' = \varepsilon_0/2$. Since (a_i) is Cauchy, there is a $N' \in \mathbb{N}$ such that

$$i, j \geq N' \implies |a_i - a_j| < \varepsilon'.$$

As above, there is an $i_0 \geq N'$ with $|a_{i_0}| \geq \varepsilon_0$.

If $i \geq N'$ then since both $i, i_0 \geq N'$ we have

$$\varepsilon_0 \leq |a_{i_0}| = |a_i + (a_{i_0} - a_i)| \leq |a_i| + |a_{i_0} - a_i| \leq |a_i| + \varepsilon'.$$

Thus

$$|a_i| \geq \varepsilon_0 - \varepsilon' = \varepsilon_0 - \frac{\varepsilon_0}{2} = \frac{\varepsilon_0}{2}.$$

In other words, if we set $d = \varepsilon_0/2$ and $k = N'$ then for all $i \geq k$ we have $|a_i| \geq d$. □

Lemma 14. *Suppose (a_i) is a Cauchy sequence in an ordered field F such that $(a_i) \not\sim (0)$. Then there is an integer k_0 such that $a_i \neq 0$ for all $i \geq k_0$ and such that $\left(a_i^{-1}\right)_{i \geq k_0}$ is a Cauchy sequence.*

Proof. By the previous lemma there is an integer k_0 and a positive $d \in F$ such that $|a_i| \geq d$ for all $i \geq k_0$. In particular, $a_i \neq 0$ if $i \geq k_0$. So a_i has a multiplicative inverse in F for all $i \geq k_0$ since F is a field. By properties of ordered fields and absolute values (Chapter 8)

$$0 < \left|a_i^{-1}\right| = |a_i|^{-1} \leq d^{-1}$$

for all $i \geq k_0$.

Our goal is to show $\left(a_i^{-1}\right)_{i \geq k_0}$ is Cauchy. So let $\varepsilon \in F$ be positive; we want an N such that if $i, j \geq N$ then $\left|a_i^{-1} - a_j^{-1}\right| < \varepsilon$.

Let $\varepsilon' = \varepsilon d^2$. Since d and ε are positive, so is ε'. Since (a_i) is Cauchy, there is a N' such that $|a_i - a_j| < \varepsilon'$ for all $i, j \geq N'$. Let N be the maximum of N' and k_0. If $i, j \geq N$ then

$$\begin{aligned}
|a_i^{-1} - a_j^{-1}| &= \left|(a_j - a_i)a_i^{-1}a_j^{-1}\right| \quad (F \text{ is a field}) \\
&= |a_j - a_i| \left|a_i^{-1}\right| \left|a_j^{-1}\right| \quad (F \text{ is an ordered field}) \\
&\leq |a_j - a_i| \left|a_i^{-1}\right| d^{-1} \quad (j \geq k_0, \text{ so } |a_j^{-1}| \leq d^{-1}) \\
&\leq |a_j - a_i| \, d^{-1} d^{-1} \quad (i \geq k_0, \text{ so } \left|a_i^{-1}\right| \leq d^{-1}) \\
&< \varepsilon' d^{-2} = \varepsilon. \qquad (i, j \geq N')
\end{aligned}$$

So $|a_i^{-1} - a_j^{-1}| < \varepsilon$ as desired. We conclude that $\left(a_i^{-1}\right)_{i \geq k_0}$ is Cauchy. □

Theorem 15. *Let $x \in \mathbb{R}$. If $x \neq 0$ then x has a multiplicative inverse.*

Proof. Write $x = [(a_i)]$ where (a_i) is a Cauchy sequence of rational numbers. By the previous lemma, there is a k_0 such that $\left(a_i^{-1}\right)_{i \geq k_0}$ is Cauchy. Thus

$$y = \left[(a_i^{-1})\right]$$

is a real number. By definition of multiplication in $\mathbb{R}$,

$$xy = [(a_i)]\left[(a_i^{-1})\right] = \left[(a_i a_i^{-1})\right] = [(1)].$$

Thus $xy = 1$. By the commutative law for multiplication, $yx = xy = 1$. We conclude that y is the multiplicative inverse of x. □

We now come to the next main theorems of this chapter.

Theorem 16. *The set of real numbers $\mathbb{R}$ is a field.*

Proof. We know that $\mathbb{R}$ is a commutative ring by Theorem 8. We know that $0 \neq 1$ by Corollary 10. Multiplicative inverses exist by Theorem 15. We conclude that $\mathbb{R}$ is a field. □

10.6 The real numbers $\mathbb{R}$ as an ordered field

In order to show that $\mathbb{R}$ is an ordered field, we need to define the set of positive real numbers P, and to show that this set has the required properties: closure and trichotomy.

Since each real number can be thought of as $[(a_i)]$ where (a_i) is Cauchy, we might be tempted to say that $x = [(a_i)]$ is positive if each a_i is positive. *This idea does not work.* For example, the sequence $(1/i^2)$ converges to 0, and so is equivalent to the constant sequence (0). Thus $[(1/i^2)]$ is zero even though all its terms are strictly positive.

Furthermore, if a sequence has a finite number of zero or negative terms, and the rest are positive, then the sequence could represent a positive number. Thus there are two ways in which the naive definition of positive is defective. The following definition corrects both deficiencies.

Definition 5 (Positive)**.** A *positive-type Cauchy sequence* in an ordered field F is a Cauchy sequence (a_i) with the following property: there is a positive $d \in F$ and an $N \in \mathbb{N}$ such that $a_i \geq d$ for all $i \geq N$.

A *positive* real number is a real number of the form $[(a_i)]$ where (a_i) is a positive-type Cauchy sequence with terms in $\mathbb{Q}$.

Exercise 7. Suppose (a_i) and (b_i) are Cauchy sequences where $(a_i) \sim (b_i)$. Show that (a_i) is positive-type Cauchy if and only if (b_i) is.

Hint: We will be assuming that (a_i) is positive-type and proving that (b_i) is. Suppose there is a $d_0 > 0$ and a $N_0 \in \mathbb{N}$ such that $a_i \geq d_0$ for all $i \geq N_0$. We must find d and N that work for (b_i). Let $\varepsilon = d_0/2$ and choose a N_1 so

that $|a_i - b_i| < \varepsilon$ for all $i \geq N_1$. Why does such a N_1 exist? Choose N as the maximum of N_0 and N_1. What do you think d should be? Prove that your choice of N and d work for (b_i).

Remark 6. The above exercise tells us that if we wish to decide if a real number x is positive, we can take *any* Cauchy sequence from the equivalence class defining x, and check the definition for that particular sequence.

For example, if $r \in \mathbb{Q}$ is thought of as a real number via the canonical embedding, then we can decide if r is positive in $\mathbb{R}$ just by looking at the constant sequence (r). From this we conclude that a rational number r is positive in $\mathbb{R}$ if and only if it is positive in $\mathbb{Q}$. Thus the present definition of positive for $\mathbb{R}$ is compatible with the definition of Chapter 7 for elements that happen to be in $\mathbb{Q}$.

Theorem 17 (Closure)**.** *If $x, y \in \mathbb{R}$ are positive then so is $x + y$ and xy.*

Proof. Let $x = [(a_i)]$ and $y = [(b_i)]$ where (a_i) and (b_i) are positive-type Cauchy sequences of rational numbers. By definition there is a positive number $d_1 \in \mathbb{Q}$ and an integer $N_1 \in \mathbb{N}$ such that $a_i \geq d_1$ for all $i \geq N_1$. Likewise, there is a positive $d_2 \in \mathbb{Q}$ and a $N_2 \in \mathbb{N}$ such that $b_i \geq d_2$ for all $i \geq N_2$. Let $d = d_1 + d_2$. We know that $d > 0$ since it is the sum of positive elements. Let N be the maximum of N_1 and N_2. If $i \geq N$, then

$$a_i + b_i \geq d_1 + b_i \geq d_1 + d_2 = d.$$

Thus $x + y = [(a_i + b_i)]$ is positive.

The proof for xy is similar. □

Exercise 8. Prove the above for the case of multiplication.

We now want to prove a trichotomy law: for all $x \in \mathbb{R}$ exactly one of the following occurs (i) x is positive, (ii) $x = 0$, or (iii) $-x$ is positive. In the third case we also say that x is *negative*.

We divide the proof of this law into lemmas:

Lemma 18. *The real number 0 is neither positive nor negative.*

Proof. The real number 0 is defined by the constant sequence $(a_i) = (0)$. Since $a_i = 0$, there can be no $d > 0$ and N such that $a_i \geq d$ for all $i \geq N$. (Since if $a_i \geq d$ and $d > 0$ then $a_i > 0$, a contradiction). So 0 cannot be positive.

We now show that 0 cannot be negative. Suppose 0 is negative. Then -0 is positive (by definition of negative). But $-0 = 0$, and we have already shown that 0 is not positive. □

Lemma 19. *Let $x \in \mathbb{R}$. It is not possible for both x and $-x$ to be positive.*

Exercise 9. Prove the above. Hint: suppose not. Write $x = [(a_i)]$. Observe that $-x = [(-a_i)]$ by Theorem 7. Define a d_1 and N_1 for (a_i) and d_2 and N_2 for $(-a_i)$. Let i be the maximum of N_1 and N_2, and show that a_i is both positive and negative in $\mathbb{Q}$.

Remark 7. Notice how we use a trichotomy law for $\mathbb{Q}$ from an earlier chapter to help prove a trichotomy law for $\mathbb{R}$.

The above two lemmas show part of the trichotomy law: together they show that at most one of the trichotomy conditions hold. We still need to show that at least one condition holds. This follows from the next lemma.

Lemma 20. *If $x \neq 0$ is a real number, then either x or $-x$ is positive.*

Proof. Write $x = \big[(a_i)\big]$ where (a_i) is a Cauchy sequence in $\mathbb{Q}$. By Theorem 7, $-x = \big[(-a_i)\big]$.

By Lemma 13, there is a $k_1 \in \mathbb{N}$ and a positive rational number $d \in \mathbb{Q}$ such that $|a_i| \geq d$ for all $i \geq k$. Since (a_i) is Cauchy, there is a $k_2 \in \mathbb{N}$ such that $|a_i - a_j| \leq d/2$ for all $i, j \geq k_2$. Let N be the maximum of k_1 and k_2.

In particular, we have $|a_N| \geq d > 0$, so either $a_N \geq d$ or $a_N \leq -d$. This gives us two cases.

We begin with the case $a_N \geq d$. For all $i \geq N$ we have $|a_i - a_N| \leq d/2$ since $i, N \geq k_2$. This means

$$-\frac{d}{2} \leq a_i - a_N \leq \frac{d}{2}.$$

So

$$a_i \geq a_N - \frac{d}{2} \geq d - \frac{d}{2} = \frac{d}{2}.$$

This shows the for all $i \geq N$ we have $a_i \geq d/2$. Thus $x = \big[(a_i)\big]$ is positive by Definition 5.

Finally consider the case $a_N \leq -d$. For all $i \geq N$ we have $|a_i - a_N| \leq d/2$ since $i, N \geq k_2$. This means

$$-\frac{d}{2} \leq a_N - a_i \leq \frac{d}{2}.$$

So

$$-a_i \geq -a_N - \frac{d}{2} \geq d - \frac{d}{2} = \frac{d}{2}.$$

This shows the for all $i \geq N$ we have $-a_i \geq d/2$. Note that since $d > 0$, it follows that $d/2 > 0$. Thus $x = \big[(-a_i)\big]$ is positive by Definition 5. □

Putting these lemmas together, we conclude the following:

Theorem 21. *For every $x \in \mathbb{R}$ exactly one of the following occurs: (i) x is positive, (ii) $x = 0$, or (iii) $-x$ is positive.*

We now come to the next main theorem of this chapter.

Theorem 22. *The set of real numbers $\mathbb{R}$ is an ordered field.*

Proof. We know that $\mathbb{R}$ is a field by Theorem 16. To show that $\mathbb{R}$ is an ordered field we need to check that (i) the positive elements are closed under addition and multiplication, and (ii) the positive elements satisfy the trichotomy law. Both these were done in Theorems 17 and 21 respectively. □

Note. Now that we know that $\mathbb{R}$ is an ordered field, we can use all the definitions and results about ordered fields F and limits from Chapter 8 including facts about $<$ and absolute values. This would be a good time to review these definitions and results from Chapter 8, which we know holds for $\mathbb{R}$.

Remark 8. As mentioned above, positivity defined for $\mathbb{R}$ is compatible with the earlier concept of positivity defined for $\mathbb{Q}$. Since $x < y$ means $y - x$ is positive, if follows that inequality in $\mathbb{R}$ is compatible with inequality in $\mathbb{Q}$. In other words, we can show that if $x, y \in \mathbb{Q}$ then $x < y$ holds for $\mathbb{Q}$ if and only if it holds for $\mathbb{R}$.

We end this section with a few lemmas concerning the order relation of $\mathbb{R}$.

Lemma 23. *Suppose (a_i) is a Cauchy sequence of rational numbers. Suppose there is a $k \in \mathbb{N}$ such that $a_i \geq 0$ for all $i \geq k$. Then $x \geq 0$ where $x = [(a_i)]$ is the corresponding real number.*

Proof. The only way for $x \geq 0$ to fail is for $-x > 0$. Suppose this happens. Since $-x = [(-a_i)]$ the sequence $(-a_i)$ is of positive type. So there is an $N \in \mathbb{N}$ and a positive $d \in \mathbb{Q}$ such that $-a_i \geq d$ for all $i \geq N$. Let i be the maximum of N and k. Then

$$d \leq -a_i \leq 0,$$

which is a contradiction since d is positive. □

Lemma 24. *Suppose (a_i) and (b_i) are two Cauchy sequences of rational numbers, and let $x = [(a_i)]$ and $y = [(b_i)]$ be the corresponding real numbers. If there is a $k \in \mathbb{N}$ such that $a_i \leq b_i$ for all $i \geq k$, then $x \leq y$.*

Proof. Observe that $y - x = [(b_i - a_i)]$. For all $i \geq k$, we have $b_i - a_i \geq 0$. By the previous lemma, $y - x \geq 0$. The result follows. □

Remark 9. If we have $a_i < b_i$ instead, we cannot necessarily conclude that $x < y$. Without extra information, we can only conclude that $x \leq y$.

10.7 Relationship between $\mathbb{R}$ and $\mathbb{Q}$

In this section we will consider a few useful results relating $\mathbb{R}$ and $\mathbb{Q}$. For example, we will see that Cauchy sequences of rational numbers always converge to real numbers, and that all real numbers are limits of rational sequences. We will also see that $\mathbb{R}$ is an Archimedean ordered field. Recall from Chapter 8 that this implies that $\mathbb{Q}$ is dense in $\mathbb{R}$; in other words, between any two distinct real numbers we can always find a rational number.

We begin with a lemma that can be used to compare rational numbers to real numbers.

Lemma 25. *Suppose $x \in \mathbb{R}$ is given by $x = [(a_i)]$. Suppose that b is a rational number. If there is a $k \in \mathbb{N}$ such that $a_i \leq b$ for all $i \geq k$, then $x \leq b$ (where here we are thinking of b as a real number). If, instead, there is a $k \in \mathbb{N}$ such that $a_i \geq b$ for all $i \geq k$, then $x \geq b$.*

Proof. For the first statement, apply Lemma 24 to the sequences (a_i) and (b). For the second statement, switch the order and apply Lemma 24 again. □

Theorem 26. *Suppose $y > 0$ is a real number. Then there is a positive integer n such that $1/n \leq y$.*

Proof. Write y as $[(a_i)]$ where (a_i) is a positive-type Cauchy sequence of rational numbers. By Definition 5, there is a $N \in \mathbb{N}$ and a positive $d \in \mathbb{Q}$ such that $a_i \geq d$ for all $i \geq N$. Write $d = m/n$ where $m, n \in \mathbb{N}$ are positive. Thus $a_i \geq d \geq 1/n$ for all $i \geq N$.

Since $1/n \leq a_i$ for all $i \geq N$, we get $1/n \leq y$ by the above Lemma. □

Theorem 27. *The real numbers $\mathbb{R}$ form an Archimedean ordered field.*

Proof. This follows from the previous theorem (by a result in Chapter 8). □

Corollary 28. *The field $\mathbb{Q}$ is dense in $\mathbb{R}$. In other words, if $x, y \in \mathbb{R}$ are such that $x < y$, there is a $r \in \mathbb{Q}$ with $x < r < y$.*

Exercise 10. Which theorem in Chapter 8 yields the above corollary?

The following theorem says that if a Cauchy sequence of rational numbers represents a certain real number, then the Cauchy sequence (as a sequence in $\mathbb{R}$) converges to the real number.

Theorem 29. *Let (a_i) be a Cauchy sequence of rational numbers. Then (a_i) considered as a sequence of real numbers converges to the real number x where $x = [(a_i)]$.*

Proof. Let ε be an arbitrary positive real number. We must find a $N \in \mathbb{N}$ such that $|a_i - x| < \varepsilon$ for all $i \geq N$. It seems like we should be able to

use the definition of Cauchy sequence to find such a N. There is a slight problem: (a_i) is a Cauchy sequence in $\mathbb{Q}$, but $\varepsilon > 0$ is in $\mathbb{R}$.

We solve the problem by choosing a positive integer n such that $1/n \leq \varepsilon$ (Theorem 26). Let $\varepsilon' = 1/n$, and note that ε' is a positive element of $\mathbb{Q}$ such that $\varepsilon' < \varepsilon$. By definition of Cauchy sequence in $\mathbb{Q}$, we have an $N \in \mathbb{N}$ such that $|a_i - a_j| < \varepsilon'$ for all $i, j \geq N$. We will show that this N has the desired property for convergence: that $|a_i - x| < \varepsilon$ for all $i \geq N$. (We will actually show $|a_i - x| \leq \varepsilon'$, which is even stronger.)

So fix $i \geq N$. Recall that a_i is thought of as both a rational number and a real number via the canonical embedding $\mathbb{Q} \to \mathbb{R}$. More precisely, a_i as a real number is defined by $[(c_j)]$ where (c_j) is the constant sequence whose terms are all equal to the rational number a_i. Recall also that $x = [(a_j)]$.

So let (c_j) be the constant sequence whose terms are equal to a_i, and assume $j \geq N$. By our choice of N we have $|c_j - a_j| < \varepsilon'$ since $c_j = a_i$. By properties of absolute values (in $F = \mathbb{Q}$),

$$-\varepsilon' < c_j - a_j < \varepsilon'.$$

The above holds for all $j \geq N$, so we can use Lemma 25 to conclude that

$$-\varepsilon' \leq [(c_j - a_j)] \leq \varepsilon'.$$

In other words (using properties of absolute values in $F = \mathbb{R}$),

$$\big|[(c_j - a_j)]\big| \leq \varepsilon'.$$

This implies that

$$|a_i - x| = \big|[(c_j)] - [(a_j)]\big| = \big|[(c - a_j)]\big| \leq \varepsilon' < \varepsilon.$$

This completes the proof that (a_i) converges to x. □

Corollary 30. *Every Cauchy sequence of rational numbers converges in* $\mathbb{R}$*..*

Proof. Let (a_i) be a Cauchy sequence of rational numbers. Let $x = [(a_i)]$. By Theorem 29, (a_i) has limit x. □

Note. Our goal is to show all Cauchy sequences in $\mathbb{R}$ converge. Corollary 30 is a nice step in this direction, but we still more to show.

Remark 10. Let (a_i) be a sequence of rational numbers. There is some ambiguity of what *Cauchy* means for (a_i) when we embed $\mathbb{Q}$ into $\mathbb{R}$. We can mean the Cauchy condition holds for all positive $\varepsilon \in \mathbb{Q}$. Call this $\mathbb{Q}$-Cauchy. Or we can mean that the Cauchy condition holds for all positive $\varepsilon \in \mathbb{R}$. Call this $\mathbb{R}$-Cauchy.

In the above theorem and corollary we are thinking of $\mathbb{Q}$-Cauchy. We proved that any $\mathbb{Q}$-Cauchy sequence gives a convergent sequence in $\mathbb{R}$. But

convergent sequences are automatically Cauchy (Chapter 8). Thus any $\mathbb{Q}$-Cauchy sequence is automatically a $\mathbb{R}$-Cauchy sequence.

Conversely, any $\mathbb{R}$-Cauchy sequence whose terms are in $\mathbb{Q}$ is a $\mathbb{Q}$-Cauchy sequence. (If a condition holds for all $\varepsilon > 0$ in $\mathbb{R}$ then it will certainly hold for all $\varepsilon > 0$ in $\mathbb{Q}$ since $\mathbb{Q} \subseteq \mathbb{R}$). We conclude that if (a_i) is a sequence of rational numbers, there is no difference between being $\mathbb{Q}$-Cauchy or $\mathbb{R}$-Cauchy.

Corollary 31. *Every real number is the limit of a sequence of rational numbers*

Proof. Let $x = [(a_i)]$ be a real number. By Theorem 29, (a_i) has limit x. □

Corollary 32. *If $x \in \mathbb{R}$ and if $\varepsilon \in \mathbb{R}$ is positive, then there is a rational number $r \in \mathbb{Q}$ with $|x - r| < \varepsilon$.*

Proof. Since x is the limit of a sequence (a_i) of rational numbers, there is a $N \in \mathbb{N}$ such that $|a_i - x| < \varepsilon$ for all $i \geq N$. Let $r = a_N$. □

Remark 11. The proceeding two corollaries can also be proved as consequences of the Archimedean property of $\mathbb{R}$: they hold for all Archimedean ordered fields.

10.8 $\mathbb{R}$ is complete

As we have discussed before, $\mathbb{Q}$ has "holes". For example, $\mathbb{Q}$ is missing a square root for 2. Because of this, $\mathbb{Q}$ has Cauchy sequences that do not converge. We will now show that the real numbers $\mathbb{R}$ do not have Cauchy sequences that fail to converge. So $\mathbb{R}$ is complete, and does not have "holes".

We begin by showing that every Cauchy sequence of real numbers converges. We already know, from the previous section, that every Cauchy sequence of rational numbers converges in $\mathbb{R}$. But this is not enough for our current needs. We need to extend the result to Cauchy sequences with terms in $\mathbb{R}$. We begin with a lemma.

Lemma 33. *If (a_i) is a sequence of real numbers, then there is a sequence (b_i) of rational numbers such that $(a_i) \sim (b_i)$. (Equivalence is taken with $F = \mathbb{R}$.)*

Proof. For each a_i, we know by Corollary 32 that there is a rational number b_i such that $|a_i - b_i| < 1/i$. Consider the sequence (b_i) formed from such rational numbers.[3] We must show that $(a_i) \sim (b_i)$.

[3]In order to avoid using the axiom of choice, we can select b_i to have the smallest possible positive denominator, and among fractions with the smallest possible denominator we choose the smallest possible numerator.

Let $\varepsilon \in \mathbb{R}$ be an arbitrary positive real number. We must find a $N \in \mathbb{N}$ such that $|a_i - b_i| < \varepsilon$ for all $i \geq N$. By Theorem 26 we can find a positive n such that $1/n < \varepsilon$. Let $N = n$. If $i \geq N$ then

$$\begin{aligned} |a_i - b_i| &< 1/i \qquad \text{(choice of } b_i\text{)} \\ &\leq 1/n \qquad (i \geq N \text{ and } N = n) \\ &< \varepsilon \qquad \text{(choice of } n\text{)} \end{aligned}$$

Thus $(a_i) \sim (b_i)$ as desired. □

Theorem 34. *Every Cauchy sequences in* $\mathbb{R}$ *converges.*

Proof. Let (a_i) be a Cauchy sequence of real numbers. By Lemma 33 there is a sequence (b_i) of rational numbers such that $(b_i) \sim (a_i)$.

In Chapter 9 we proved that if two sequences are equivalent, and if one is Cauchy, then the other is. Since (a_i) is Cauchy, we conclude that (b_i) is also a Cauchy sequence. By Corollary 30 we conclude that (b_i) has a limit.

In Chapter 8 we proved that if two sequences are equivalent, and if one has a limit, then the other does as well. Since (b_i) has a limit, we conclude that (a_i) must have a limit. □

Now for the main theorem.

Theorem 35 (Main theorem)**.** *The field* $\mathbb{R}$ *is a complete ordered field.*

Proof. In Chapter 9 we proved that if an ordered field is Archimedean and if every Cauchy sequences converges in that field, then that field must be complete. So, since $\mathbb{R}$ is Archimedean, and since every Cauchy sequence in $\mathbb{R}$ converges, $\mathbb{R}$ is a complete ordered field. □

Chapter 11

Exploring $\mathbb{R}$

In this chapter we investigate some important properties of $\mathbb{R}$ that are a consequence of its completeness, and which fail for $\mathbb{Q}$. For example, every decimal expansion defines a real number, but not always a rational number. Also, for every positive integer n and every nonnegative real number x, there is a unique nonnegative nth root $x^{1/n}$. The existence of such roots often fails for rational numbers. We end the chapter by showing that $\mathbb{Q}$ is countable, but $\mathbb{R}$ is uncountable.

11.1 Review of properties of $\mathbb{R}$

We begin by collecting together for convenience some results about $\mathbb{R}$ that have already been proved.

Theorem 1. *The field $\mathbb{R}$ is a complete ordered field. In particular, every nonempty subset $S \subseteq \mathbb{R}$ which is bounded from above has a supremum (least upper bound). Likewise, every nonempty subset $S \subseteq \mathbb{R}$ which is bounded from below has a infimum (greatest lower bound).*

Proof. See Chapter 10 for the proof. The second statement is our definition of completeness from Chapter 9, and the third statement was proved to be a consequence of this definition. □

Theorem 2. *Every Cauchy sequence in $\mathbb{R}$ converges.*

Proof. See Chapter 10 for the proof (and Chapter 9 for the definition of Cauchy). □

Theorem 3. *The field $\mathbb{R}$ is an Archimedean ordered field, and $\mathbb{Q}$ is a dense subfield of $\mathbb{R}$.*

Proof. In Chapter 9 we showed that complete ordered fields are Archimedean. (We also proved this in Chapter 10 for $\mathbb{R}$ in particular). In Chapter 8 we proved that $\mathbb{Q}$ must be a dense subfield of any Archimedean ordered field. □

Theorem 4 (Intermediate Value Theorem)**.** *Let $[a, b]$ be a closed interval in $\mathbb{R}$ where $a < b$ are elements of $\mathbb{R}$. Suppose $f\colon [a, b] \to \mathbb{R}$ is continuous. If $C \in \mathbb{R}$ is any value between $A = f(a)$ and $B = f(b)$ then there is an element $c \in [a, b]$ such that $f(c) = C$.*

Proof. We proved this in Chapter 9 for complete ordered fields containing $\mathbb{Q}$ as a subfield. □

Theorem 5. *Suppose $C \in \mathbb{R}$ and that $C \geq 0$. Then there is a $c \in \mathbb{R}$ such that $c^2 = C$.*

Proof. We proved this in Chapter 9 as a corollary to the intermediate value theorem. □

Theorem 6. *Let (a_i) be a bounded monotonic sequence in $\mathbb{R}$. Then (a_i) converges.*

Proof. We proved this in Chapter 9 for complete ordered fields. □

11.2 More results about sequences

In Chapter 8 we proved some limit laws. Here we add a few more.

Theorem 7. *Let $x \in \mathbb{R}$. If $x > 1$ then the sequence $(x^i)_{i\geq 1}$ of powers is an unbounded strictly increasing sequence of positive terms.*

Proof. By induction we can show $x^i > 1$ for all $i \geq 1$. This induction uses the following inequality:

$$x^{i+1} = x^i x > x^i 1 = x^i$$

This inequality also shows that (x_i) is a strictly increasing sequence.

Now we show that the sequence is unbounded. Suppose instead that (x^i) is bounded, so it is a monotonic bounded sequence. Since $\mathbb{R}$ is complete, this sequence would then converge (Theorem 6). All convergent sequences are Cauchy, so (x^i) would have to be Cauchy. Now observe that, for all $i \in \mathbb{N}$,

$$|x^{i+1} - x^i| = x^{i+1} - x^i = x^i(x - 1) > 1(x - 1) = x - 1.$$

This implies that (x^i) is not Cauchy (take $\varepsilon = x - 1$). This gives a contradiction. Thus the sequence (x^i) cannot be bounded. □

Exercise 1. Complete the above proof. (1) Show that if $i \geq 1$ then $x^i > 1$ by induction. (2) Explain why the sequence is not Cauchy.

Theorem 8. *If (x_i) is an unbounded increasing sequence of positive terms in $\mathbb{R}$ (or in any ordered field F), then the sequence $\left(x_i^{-1}\right)$ converges to 0.*

Proof. Let $\varepsilon > 0$ be given. Since (x_i) is unbounded, ε^{-1} cannot be an upper bound of (x_i). So there is a $k \in \mathbb{N}$ such that $x_k > \varepsilon^{-1}$. Hence $x_k^{-1} < \varepsilon$. If $i \geq k$ then $x_i \geq x_k$ since the sequence is increasing. So

$$\left|x_i^{-1} - 0\right| = x_i^{-1} \leq x_k^{-1} < \varepsilon.$$

We conclude that $\left(x_i^{-1}\right)$ converges to 0. □

Exercise 2. Combine the above theorems to show the following for $x \in \mathbb{R}$: (1) if $x > 1$ then the sequence (x^{-i}) converges to 0. (2) If $0 < x < 1$ then the sequence (x^i) converges to 0. Hint: the second follows from the first.

11.3 Decimal sequences

It is common to think of a real number as something that can be written as an infinite decimal, such as $3.14159\ldots$ or $1.41421\ldots$. Even rational numbers can be written in this way: $3/2 = 1.5000\ldots$ or $2/3 = 0.666666\ldots$. Our goal in this and the next few sections is to formally justify this view of real numbers. For convenience, we typically restrict our attention to nonnegative real numbers.

Definition 1 (Decimal sequence)**.** Suppose $n \in \mathbb{N}$ and $(d_i)_{i\geq 1}$ is a sequence where $d_i \in \{0, \ldots, 9\}$ for all $i \geq 1$. Then the sequence (s_i) whose ith term is

$$s_i = n + \sum_{j=1}^{i} \frac{d_j}{10^j}$$

is called a *decimal sequence*. It is a sequence of rational numbers.

Remark 1. The sequence (s_i) in the above definition is an example of a type of sequence called a *series*. Series are sequences defined in terms of summation. Each s_i is called a *partial sum* of the series, and the limit, if it exists, is called the *value* of the series.

Remark 2. The above definition describes the mathematical definition of a decimal sequence. Now we discuss notation used in practice. Let n, s_i, d_i be as in the above definition. Let N be the base 10 numeral representing n, and let D_i be the standard digit symbol representing d_i. Then the term s_1 is written $N.D_1$, the term s_2 is written $N.D_1D_2$, the term s_3 is written $N.D_1D_2D_3$ and so on. Notation such as

$$N.D_1D_2D_3D_4\ldots$$

is used to represent *the limit* of the above decimal sequence (s_i), which we will show always exists. It is the "..." at the end indicates that we are referring to the limit (without the "...", the express $N.D_1D_2D_3D_4$ would refer to s_4).

For example, $3.22222\ldots$ denotes the limit of (s_i) where

$$s_i = 3 + \sum_{j=1}^{i} \frac{2}{10^j}.$$

So $3.22222\ldots$ is the limit of the sequence with terms 3, 3.2, 3.22, 3.222, ... Given facts about geometric series, one can show that $\sum_{j=1}^{i} \frac{2}{10^j}$ has limit equal $2/9$, so (s_i) has limit $3 + 2/9$. Thus $3.22222\ldots$ is $29/9$.

Obviously expressions such as $N.D_1D_2D_3D_4\ldots$ do not itself give full information about the decimal sequences or limits they represent. If there is an obvious pattern in the digits given, then the reader is expected to assume that the pattern continues. For example, the expression $3.1454545\ldots$ would suggest to the reader that $d_i = 4$ for even $i \geq 2$ and $d_i = 5$ for odd $i \geq 3$. Even if there is no such pattern, the the convention is that is that the number of digits expressed is enough to approximate the number at hand, and that further digits are not important (or not known) for the discussion.

If there is a $k \in \mathbb{N}$ such that $d_i = 0$ for all $i \geq k$ then s_i is constant for $i \geq k$ and we say that the decimal sequence "terminates". The limit is then equal to the kth term s_k, and we can write the number with a terminating decimal. For example, 7.3450 can be used to represent the limit when $n = 7, d_1 = 3, d_2 = 4, d_3 = 5, d_4 = 0$ and where we implicitly assume $d_k = 0$ for $k \geq 5$. Of course that same number could be written as 7.345 or 7.34500 or even $7.34500\ldots$. As we will see later, this number can also be written as $7.3449999\ldots$.

Remark 3. There is nothing sacred about base 10. We can easily replace 10 with another integer $B > 1$ in Definition 1, and insist that

$$d_i \in \{0, \ldots, B-1\}.$$

This would give us base B expansions of real numbers.

Informal Exercise 3. What rational number is represented by $3.22000\ldots$? If we are using base $B = 4$ notation, which rational number does $3.22000\ldots$ represent? (Write your answers as fractions in terms of two natural numbers given in base 10).

Informal Exercise 4. What rational number is represented by $2.2111\ldots$ in base 10. If we are using base $B = 4$ notation, which rational number is expressed by $2.2111\ldots$. (Write your answers as fractions in terms of two natural numbers given in base 10).

Now we establish that any decimal sequence is bounded and monotonic.

Theorem 9. *Suppose $n \in \mathbb{N}$ and $(d_i)_{i\geq 1}$ is a sequence where $d_i \in \{0, \ldots, 9\}$ for all $i \geq 1$. Then the sequence (s_i) whose ith term is given by*

$$s_i = n + \sum_{j=1}^{i} \frac{d_j}{10^j}$$

is increasing with upper bound $n+1$ and lower bound n. In particular, it is is bounded and monotonic.

Proof. Since $d_i \leq 9$ we have

$$\sum_{j=1}^{i} \frac{d_j}{10^j} \quad \leq \quad \sum_{j=1}^{i} \frac{9}{10^j}.$$

(This can be rigorously shown using induction). So, by Lemma 10 (below),

$$\sum_{j=1}^{i} \frac{d_j}{10^j} \quad \leq \quad 1.$$

Adding n gives $s_i \leq n+1$. So $n+1$ is an upper bound for the sequence.

Since $d_{i+1} \geq 0$, we can show that n is a lower bound and that

$$s_{i+1} = s_i + \frac{d_{i+1}}{10^{i+1}} \geq s_i$$

for all i. Thus (s_i) is increasing. □

The following is a special case of the formula for geometric series.

Lemma 10. *For all i,*

$$\sum_{j=1}^{i} \frac{9}{10^j} = 1 - \frac{1}{10^i}.$$

Proof. Let $s = \sum_{j=1}^{i} \frac{9}{10^j}$. By properties of summations and powers,

$$\frac{s}{10} = \frac{1}{10} \sum_{j=1}^{i} \frac{9}{10^j} = \sum_{j=1}^{i} \frac{9}{10^{j+1}} = \sum_{j=2}^{i+1} \frac{9}{10^j} = \sum_{j=2}^{i} \frac{9}{10^j} + \frac{1}{10} \frac{9}{10^i}.$$

This is similar to the expression for s. In fact, we can write s as follows:

$$s \quad = \quad \frac{9}{10} + \sum_{j=2}^{i} \frac{9}{10^j}.$$

When we take the difference, the term $\sum_{j=2}^{i} \frac{9}{10^j}$ cancels giving us

$$s - \frac{s}{10} = \frac{9}{10} - \frac{1}{10} \frac{9}{10^i} = \frac{9}{10} \left(1 - \frac{1}{10^i}\right).$$

Multiply both sides by 10, then divide by 9. The result follows. □

Corollary 11. *Suppose $n \in \mathbb{N}$ and $(d_i)_{i\geq 1}$ is a sequence where $d_i \in \{0, \ldots, 9\}$ for all $i \geq 1$. Then the sequence (s_i) whose ith term is given by*

$$s_i = n + \sum_{j=1}^{i} \frac{d_j}{10^j}$$

converges to a real number $\mathbb{R}$. More specifically, it converges to a real number x with $n \leq x \leq n+1$.

Proof. By Theorem 9, (s_i) is bounded and monotonic. Since $\mathbb{R}$ is a complete field, this implies that (s_i) has a limit x (Theorem 6). Since

$$n \leq s_i \leq n+1$$

we have $n \leq x \leq n+1$ by basic limit laws (Chapter 8). □

Remark 4. This shows that every decimal sequence defines a real number. For example, $3.17117111711117\ldots$ defines a real number between 3 and 4.

11.4 Decimal expansions

In this section we consider the converse to the problem in the previous section. We establish that every nonnegative real number is the limit of a decimal sequence. The decimal sequence giving x as its limit is called the *decimal expansion* of x.

Theorem 12. *Every nonnegative real number is the limit of a decimal sequence. In other words, every nonnegative real number has a decimal expansion.*

Proof. Let x be a nonnegative real number. We divide the proof into three steps. First we define a sequence (a_i) of rational numbers recursively in terms of the given x. Next we show that (a_i) is a decimal sequence. Finally we show that (a_i) converges to x.

We know $\mathbb{R}$ is archimedean, so by a result of Chapter 8 there is a unique integer n such that $n \leq x < n+1$. We define a_0 to be n.

Now suppose that $a_i \in \mathbb{Q}$ has been defined. We will now define a_{i+1} in terms of a_i. Consider

$$y = 10^{i+1}(x - a_i).$$

By the aforementioned result of Chapter 8 there is a unique integer d such that $d \leq y < d+1$. Now define a_{i+1}:

$$a_{i+1} \stackrel{\text{def}}{=} a_i + \frac{d}{10^{i+1}}.$$

Next multiply the terms occurring in the inequality $d \leq y < d+1$ by $1/10^{i+1}$ and simplify to observe that

$$a_{i+1} \leq x < a_{i+1} + \frac{1}{10^{i+1}}.$$

This process recursively defines a sequence $(a_i)_{i\geq 0}$ of rational numbers. We did so in such a way that $a_i \leq x < a_i + 1/10^i$ holds for all $i \geq 0$. In other words, we have

$$0 \leq x - a_i < \frac{1}{10^i}, \qquad \text{so} \quad 0 \leq 10^{i+1}(x - a_i) < 10.$$

In order to identify the digits, we select d_{i+1} to be the integer d that arises in the above definition. So for $i \geq 0$, let d_{i+1} be the unique integer such that

$$d_{i+1} \leq 10^{i+1}(x - a_i) < d_{i+1} + 1.$$

Equivalently, d_{i+1} is the largest integer less than or equal to $10^{i+1}(x - a_i)$, and, as established above, $0 \leq 10^{i+1}(x - a_i) < 10$. In other words, for each $i \geq 1$, we have $d_i \in \{0, \ldots, 9\}$. This completes the first part of the proof: we have defined (a_i) for $i \geq 0$ and the associated digits d_i for $i \geq 1$.

Our next step is to show that (a_i) is a decimal sequence. We do so by establishing that

$$a_i = a_0 + \sum_{j=1}^{i} \frac{d_j}{10^j}.$$

Observe that this can be shown by induction. With this established, we see that (a_i) is a decimal sequence. This completes the second part of the proof.

Finally, we show that (a_i) converges to x. By Corollary 11, (a_i) has a real limit, call it s. Above we established

$$a_i \leq x < a_i + 1/10^i$$

for all $i \geq 1$. By previously established limit laws, we get $s \leq x \leq s + 0$. Thus $x = s$. So (a_i) has limit x as desired. □

Exercise 5. Which limit laws were used to show that $s \leq x \leq s + 0$ follows from $a_i \leq x < a_i + 1/10^i$? (Hint: one law used was an exercise from this Chapter).

11.5 Uniqueness of decimal expansions

Some numbers have two distinct decimal expansions, but in other cases the expansions are unique. For example, $0.13999999\ldots$ and $0.1400000\ldots$ are two representations for the same number $7/50$. However, $0.22222222\ldots$ represents the unique expansion of $2/9$.

In what sense is the decimal expansion of a nonnegative $x \in \mathbb{R}$ unique? In this section we will consider this question of uniqueness of decimal expansions.

Definition 2. Suppose $n \in \mathbb{N}$ and $(d_i)_{i\geq 1}$ is a sequence where $d_i \in \{0, \ldots, 9\}$ for all $i \geq 1$, and that (s_i) is the decimal sequence whose ith term is given by

$$s_i = n + \sum_{j=1}^{i} \frac{d_j}{10^j}.$$

We call this a *nine-sequence* if there is a k such that $d_i = 9$ for all $i \geq k$. We call this a *zero-sequence* if there is a k such that $d_i = 0$ for all $i \geq k$.

Remark 5. The number $14.563599999\ldots$ is the limit of a nine-sequence, but (by the uniqueness result below) the number $14.599999999111\ldots$ is not, nor is $14.999999000\ldots$.

The number $14.56360000\ldots$ is the limit of a zero-sequence. As mentioned above we can write 14.5636 or 14.56360 for this number. For such a terminating epression, it is assumed that the digits beyond the termination point are 0, and that the associated decimal sequence is then a zero-sequence.

Theorem 13 (Non-uniqueness)**.** *Suppose the sequence with ith term*

$$s_i = d_0 + \sum_{j=1}^{i} \frac{d_j}{10^j}$$

is a nine-sequence, and let k be the least integer such that $d_i = 9$ for all $i > k$. Then (s_i) converges to the same number as the zero-sequence (s'_i) where

$$s'_i = d'_0 + \sum_{j=1}^{i} \frac{d'_j}{10^j}$$

and where (d_i) is defined as follows: $d'_i = d_i$ if $i < k$, but $d'_k = 1 + d_k$, and $d'_i = 0$ if $i > k$.

Proof. Let x be the limit of (s_i), and let x' be the limit of (s'_i). Observe

that for $i > k$,

$$\begin{aligned}
s_i &= \sum_{j=0}^{k-1} \frac{d_j}{10^j} + \frac{d_k}{10^k} + \sum_{j=k+1}^{i} \frac{9}{10^j} \\
&= \sum_{j=0}^{k-1} \frac{d'_j}{10^j} + \frac{d'_k - 1}{10^k} + \sum_{j=1}^{i-k} \frac{9}{10^{j+k}} \\
&= \left(\sum_{j=0}^{k-1} \frac{d'_j}{10^j} + \frac{d'_k}{10^k} \right) - \frac{1}{10^k} + \frac{1}{10^k} \sum_{j=1}^{i-k} \frac{9}{10^j} \\
&= \left(\sum_{j=0}^{k} \frac{d'_j}{10^j} \right) - \frac{1}{10^k} + \frac{1}{10^k} \left(1 - \frac{1}{10^{i-k}} \right) \\
&= s'_k - \frac{1}{10^k} + \frac{1}{10^k} - \frac{1}{10^k} \frac{1}{10^{i-k}} \\
&= x' - \frac{1}{10^i}.
\end{aligned}$$

The limit of (10^{-i}) is 0 by an earlier exercise. So $x = x'$. □

Remark 6. The above shows that a nine-sequence can be replaced by a zero-sequence representing the same real number. For example, $11.34999\ldots$ gives the same real number as $11.35000\ldots$ (where $k = 2$) and $9.99999\ldots$ gives the same result as $10.0000\ldots$ (where $k = 0$).

Theorem 14. *Consider two convergent sequences (s_i) and (t_i) defined in terms of sums:*

$$s_i = \sum_{j=l}^{i} a_j, \qquad t_i = \sum_{j=l}^{i} b_j.$$

Write A for the limit of (s_i) and B for the limit of (t_i). If $a_j \leq b_j$ for all $j \geq l$, then $A \leq B$. If in addition some $a_j < b_j$, then $A < B$.

Proof. Observe that $s_k \leq t_k$ for all $k \geq l$ by induction. So in the limit $A \leq B$ (See Chapter 8).

Now assume in addition that $a_j < b_j$ for a specific integer j. If $j > l$ then

$$s_j + (b_j - a_j) = (s_{j-1} + a_j) + (b_j - a_j) = s_{j-1} + b_j \leq t_{j-1} + b_j = t_j.$$

If, on the other hand, $j = l$, then

$$s_j + (b_j - a_j) = a_j + (b_j - a_j) = b_j = t_j.$$

In any case, $s_j + (b_j - a_j) \leq t_j$. By induction, this extends to $j \geq k$:

$$s_k + (b_j - a_j) \leq t_k.$$

So in the limit. $A + (b_j - a_j) \leq B$. Thus $A < B$ since $b_j - a_j > 0$. □

Theorem 15 (Comparison)**.** *Suppose we have two decimal sequences that differ starting in the kth digit. More precisely, suppose*

$$s_i = d_0 + \sum_{j=1}^{i} \frac{d_j}{10^j}, \qquad s'_i = d'_0 + \sum_{j=1}^{i} \frac{d'_j}{10^j}$$

such that there is a $k \in \mathbb{N}$ where $d_i = d'_i$ if $i < k$ but $d_k > d'_k$. Let S and S' be the respective limits of (s_i) and (s'_i). Then $S \geq S'$.

Equality $S = S'$ holds if and only if (i) $d_k = d'_k + 1$, (ii) (s_i) is a zero-sequence with $d_i = 0$ for all $i > k$ and (iii) (s'_i) is a nine-sequence with $d'_i = 9$ for all $i > k$.

Proof. Define (t_i) by

$$t_i = e_0 + \sum_{j=1}^{i} \frac{e_j}{10^j}$$

where $e_i = d_i$ if $i \leq k$, but where $e_i = 0$ if $i > k$. Let T be the limit of (t_i). Then $S \geq T$ by Theorem 14 where equality holds if and only if $(s_i) = (t_i)$ as sequences.

Define (t'_i) by

$$t'_i = e'_0 + \sum_{j=1}^{i} \frac{e'_j}{10^j}$$

where $e_i = d'_i$ if $i < k$, but where $e'_k = d_k - 1$ and $e'_i = 9$ if $i > k$. Let T' be the limit of (t'_i). Then $T' \geq S'$ by Theorem 14 where equality holds if and only if $(s'_i) = (t'_i)$ as sequences.

Finally, $T = T'$ by Theorem 13. Thus

$$S \geq T = T' \geq S'$$

with equality $S = T = T' = S'$ if and only if $(s_i) = (t_i)$ and $(s'_i) = (t'_i)$. The result follows. □

This leads to the main uniqueness results described in the following three corollaries.

Corollary 16. *Every nonnegative real number has a unique non nine-sequence decimal expansion.*

Corollary 17. *Every nonnegative real number has a unique non zero-sequence decimal expansion.*

Corollary 18. *Suppose x is a nonnegative real number with a decimal expansion that is neither a nine-sequence nor a zero-sequence. Then the decimal expansion of x is unique.*

Informal Exercise 6. Write 3/2 using two different decimal expansions. What is the unique non nine-sequence representing 3/2?

Informal Exercise 7. Write $14.999999000\ldots$ in terms of a nine-sequence.

11.6 Basic inequalities for kth powers

We now establish a tool-kit of useful results used later in the chapter.

Lemma 19. *Let x, y be positive elements of an ordered field F. Then x^n and y^n are also positive for all $n \in \mathbb{N}$. Furthermore, if $n \geq 1$ then*

$$x^n \leq y^n \iff x \leq y$$

and

$$x^n < y^n \iff x < y.$$

Proof. If $x \in F$ is positive, then observe that x^n is positive for $n \geq 0$ (by induction using closure of the positive subset $P \subseteq F$, and using the fact that 1 is positive in the base case).

Observe next that if $x \leq y$, then $x^n \leq y^n$ for $n \geq 0$ (also by induction). Similarly, if $x < y$ then $x^n < y^n$ for $n \geq 1$ (by induction starting at $n = 1$).

Suppose $x^n \leq y^n$ and $n \geq 1$. If $y < x$ then $y^n < x^n$ by the above. This is a contradiction to trichotomy. Thus $x \leq y$.

Suppose $x^n < y^n$ and $n \geq 1$. If $y \leq x$ then $y^n \leq x^n$ by the above. This is a contradiction to trichotomy. Thus $x < y$. □

Lemma 20. *Let x, y be nonnegative elements of an ordered field F. Let $n \in \mathbb{N}$. Then x^n and y^n are also nonnegative. Furthermore, if $n \geq 1$, then*

$$x^n \leq y^n \iff x \leq y.$$

Proof. The case where x and y are both positive is covered by Lemma 19. Observe that if one or both of x, y is zero then previously established facts about 0 give the conclusion. □

Exercise 8. Give details in the above proof where (i) $x = 0$, and (ii) $y = 0$.

Lemma 21. *Let x, y be nonnegative elements of an ordered field F. If n is a positive integer then*

$$x^n = y^n \iff x = y.$$

Proof. The direction $\Leftarrow$ follows from properties of equality: if $x = y$ then we have $x^n = y^n$ by substitution.

So suppose $x^n = y^n$. Then $x^n \leq y^n$, hence $x \leq y$ by the previous lemma. Likewise, $y \leq x$. So $x = y$. □

11.7 Existence of nth roots (nonnegative case)

In this section, we will establish that every nonnegative $x \in \mathbb{R}$ has a unique nonnegative nth root in $\mathbb{R}$.

Definition 3 (nth root)**.** Let F be a field, and let n be a positive integer. If $x^n = X$ where $x, X \in F$, then we say that x is an *nth root* of X.

In the special case where $n = 2$, then x is called a *square root* of X. In the special case where $n = 3$, then x is called a *cube root* of X.

First we show existence of an nth root for $C \geq 1$.

Lemma 22. *Let n be a positive integer. If $C \geq 1$ is a real number, then there is a positive real number c such that $c^n = C$.*

Proof. Let $b = C$. Observe, using induction for $n \geq 1$, that $f(x) = x^n$ defines a continuous function $[0, b] \to \mathbb{R}$ and that since $C \geq 1$,

$$C \leq C^n.$$

The above is for all $n \geq 1$. Now fix n, and let $f(x) = x^n$. Let $A = f(0)$ and let $B = f(b)$. Observe that $A = f(0) = 0$, so $A \leq C$. Since $b = C$, we have $B = f(b) = C^n$. So $C \leq B$, since $C \leq C^n$. Since $A \leq C \leq B$, there is a real $c \in [0, b]$ such that $f(c) = C$ by the Intermediate Value Theorem (Chapter 9). Since $f(c) = c^n$, the number c has the desired property. □

Exercise 9. Give details for two steps in the first paragraph of the above proof: (1) Show that $f(x) = x^n$ defines a continuous function (hint: consider $g(x) = x$ the identity function. What is g^n *in the ring of continuous functions, i.e. using exponentiation as defined in Chapter 6*?) (2) Show that if $C \geq 1$ then $C^n \geq C$ for all $n \geq 1$.

We need another lemma to handle roots of real numbers less than one.

Lemma 23. *Let n be a positive integer. If $0 < X < 1$ is a real number, then there is a positive real number x such that $x^n = X$.*

Proof. Let $C = 1/X$. Observe that $C > 1$. So there is a positive real number c with $c^n = C$ (by the previous lemma). Let $x = 1/c$. Observe that $x^n = 1^n/c^n = 1/C = X$ as desired. □

Theorem 24 (nth roots)**.** *Let $X \in \mathbb{R}$ be nonnegative and let $n \in \mathbb{N}$ be postive. Then X has a unique nonnegative nth root x. If X is positive then so is x.*

Proof. If $X > 0$ then the existence of a positive nth root follows from the previous two lemmas. If $X = 0$ then $x = 0$ is an nth root.

To show uniqueness, suppose x_1 and x_2 are two nth roots. Then

$$x_1^n = X = x_2^n.$$

By Lemma 21, $x_1 = x_2$ as desired. So uniqueness holds. □

Definition 4. If x is a nonnegative real number and if $n \in \mathbb{N}$ is a positive integer, then $x^{1/n}$ is defined to be the unique nonnegative nth root of x. We sometimes write $x^{1/n}$ as $\sqrt[n]{x}$.

If x is a nonnegative real number, then $\sqrt{x}$ is defined to be the unique nonnegative square root of x. In other words, $\sqrt{x}$ is $x^{1/2}$.

Definition 5. An *irrational* real number is an element of $\mathbb{R}$ that is not in $\mathbb{Q}$.

Remark 7. Suppose that b is a natural number that is not of the form a^n for some $a \in \mathbb{N}$. Then one can show that $b^{1/n}$ is irrational. We will not prove this here, but it can be easily proved using basic number theory.

For example 6 is not of the form a^3 with $a \in \mathbb{N}$, in other words 6 is not a cube. So $6^{1/3}$ can be shown to be irrational.

11.8 Fractional powers

Here we give a few properties of fractional powers.

Theorem 25. *Let $x, y \in \mathbb{R}$ be nonnegative, and let $n \in \mathbb{N}$ positive. Then*

$$(xy)^{1/n} = x^{1/n} y^{1/n}.$$

Proof. Let $v = x^{1/n}$ and $w = y^{1/n}$. By Definition 4, $v^n = x$ and $w^n = y$, and v and w are nonnegative. By closure properties, vw is nonnegative, and by properties of commutative rings,

$$(vw)^n = v^n w^n = xy.$$

Thus vw is the nonnegative nth root of xy. So $(xy)^{1/n} = vw = x^{1/n} y^{1/n}$. □

Exercise 10. Prove the following.

Theorem 26. *Let $x \in \mathbb{R}$ be nonnegative, and let $n \in \mathbb{N}$ positive. Then*

$$\left(x^{1/n}\right)^n = x, \qquad \text{and} \qquad (x^n)^{1/n} = x.$$

Definition 6 (Fractional powers)**.** Suppose x is a nonnegative real number, and p/q is a positive rational number with p, q positive integers. Then

$$x^{p/q} \stackrel{\text{def}}{=} \left(x^p\right)^{1/q}.$$

Lemma 27. *The above definition is well-defined: it does not depend on the choice of numerator and denominator used to represent the given rational number.*

Proof. Suppose that $p/q = r/s$ where p, q, r, s are positive integers. We must show that

$$(x^p)^{1/q} = (x^r)^{1/s}.$$

Let $v = (x^p)^{1/q}$ and $w = (x^r)^{1/s}$. Observe that

$$v^{qs} = x^{ps} \qquad \text{and} \qquad w^{qs} = x^{rq}.$$

Since $p/q = r/s$, we have $ps = qr$. Thus $v^{qs} = w^{qs}$. By Lemma 21, $v = w$ as desired. □

Theorem 28. *Suppose x is a nonnegative real number, and p/q is a positive rational number (with p, q positive integers). Then*

$$x^{p/q} = (x^p)^{1/q} = \left(x^{1/q}\right)^p$$

Proof. The first equality is true by definition. To establish $(x^p)^{1/q} = \left(x^{1/q}\right)^p$, raise both sides to the same power q. Both sides simplify to give the same answer, namely x^p. Now use Lemma 21. □

Remark 8. For example, you can compute $8^{2/3}$ in two ways. The first method starts with $8^2 = 64$. Then you take the cube root, which is 4. In the second method you take the cube root of 8. This is 2. Next square it. This gives 4. Of course, both methods give the same answer.

11.9 Roots in $\mathbb{R}$ (general case)

Above we considered only nonnegative nth roots of nonnegative real numbers. In this case we have existence and uniqueness. When we look at *all* real numbers we experience problems with existence and uniqueness when n is even. The case when n is odd works out better.

Lemma 29. *Suppose n is a positive integer. Then 0 has exactly one nth root. That root is 0.*

Exercise 11. Use the fact that $\mathbb{R}$ is a field to show that if $x \neq 0$ then $x^n \neq 0$ for all $n \geq 1$. Now prove the above lemma.

Theorem 30 (Roots for even exponents)**.** *Suppose n is a positive even integer, and that $x \in \mathbb{R}$.*

If $x > 0$, then x has exactly two nth roots: $x^{1/n}$ and $-x^{1/n}$.

If $x = 0$ then x has exactly one nth root. That root is 0.

If $x < 0$ then x has no nth roots.

Proof. Write n as $2m$.

First suppose that x is positive. Then $x^{1/n}$ is a positive nth root. Consider the negative real number $-x^{1/n}$. Then

$$\left(-x^{1/n}\right)^n = (-1)^n \left(x^{1/n}\right)^n = \left((-1)^2\right)^m x = 1^m \cdot x = x.$$

So $-x^{1/n}$ is a second nth root. Suppose y is a third nth root. Observe that y cannot be positive by the uniqueness claim of Theorem 24. $y \neq 0$ since $x \neq 0$. So y is negative. This implies that $-y$ is positive. Observe that

$$(-y)^n = y^n = x.$$

Thus $-y$ is the positive nth root $x^{1/n}$. This implies that $y = -x^{1/n}$. So there is no distinct third nth root.

The case of $x = 0$ is covered by the previous lemma.

Finally, suppose $x < 0$. If $y \in \mathbb{R}$, then y^2 is nonnegative. So $y^n = \left(y^2\right)^m$ is nonnegative. So $y^n \neq x$. Thus x has no nth roots. □

Lemma 31. *Suppose n is a positive odd integer. If $x < 0$ then $x^n < 0$.*

Exercise 12. Prove the above. Hint: write $n = 2m + 1$. (There is no need to use induction).

Theorem 32 (Roots for odd exponents)**.** *Suppose n is an odd integer and $x \in \mathbb{R}$. Then x has a unique nth root.*

If $x < 0$ then the unique nth root of x is $-|x|^{1/n}$.

Proof. As in the above lemma, if $y < 0$ then y^n is negative. This shows that if $x \geq 0$, then x cannot have any negative nth roots. But we know x has a unique nonnegative square root. Thus the theorem holds for $x \geq 0$.

If $x < 0$, then the nth power of $-|x|^{1/n}$ is equal to $-|x|$. But $-|x| = x$ in this case. So $-|x|^{1/n}$ is an nth root. Suppose y is another nth root. Then $(-y)^n = -x = |x|$. By the uniqueness claim for nonnegative reals demonstrated above, $-y = |x|^{1/n}$. Thus $y = -|x|^{1/n}$. So the nth root is unique. □

11.10 Countability and uncountability

We have encountered several differences between $\mathbb{Q}$ and $\mathbb{R}$. We consider one more very important difference: $\mathbb{Q}$ is countable, but $\mathbb{R}$ is not.

Definition 7. A nonempty set S is *countable* if there is a surjection $\mathbb{N} \to S$. The empty set is also considered to be countable. If S is nonempty and no such surjection exists, then S is said to be *uncountable*.

Remark 9. The surjective function $\mathbb{N} \to S$ described above is called a *counting function*. We do not require it to be a bijective since we want to consider finite S where a function $\mathbb{N} \to S$ cannot be a bijection.

However, it turns out that if S is infinite, one can show that the existence of a surjective $f\colon \mathbb{N} \to S$ implies the existence of a bijective $f'\colon \mathbb{N} \to S$. We will not need this result here, though.

Exercise 13. Show that every finite set is countable. Show that $\mathbb{N}$ and $\mathbb{Z}$ are countable. Hint: for $\mathbb{Z}$, define a function $\mathbb{N} \to \mathbb{Z}$ that sends each even $2k \in \mathbb{N}$ to k and sends each odd $2k-1 \in \mathbb{N}$ to $-k$.

Theorem 33. *The set of real numbers $\mathbb{R}$ is uncountable.*

Proof. We show that no function $f\colon \mathbb{N} \to \mathbb{R}$ can be surjective. We do so by taking any given $f\colon \mathbb{N} \to \mathbb{R}$, and finding a real number that is not in the image. So let $f : \mathbb{N} \to \mathbb{R}$ be given. We can think of f as providing a list or sequence of real numbers: $f(0), f(1), f(2), \ldots$. The goal is to construct a decimal sequence giving a new real number not on the list.

Let n be a positive integer greater than $f(0)$. For each $i \geq 1$ let d_i be the ith digit in the decimal expansion of $|f(i)|$. In other words d_i serves as the digit for the 10^{-i} place of the expansion. (For definiteness, we choose the decimal expansion to be the unique non nine-sequence).

Now define $d'_i = 5$ if $d_i \neq 5$, but choose $d'_i = 7$ if $d_i = 5$. Consider the decimal sequence defined by

$$s_i = n + \sum_{j=1}^{i} \frac{d'_j}{10^j}.$$

As we proved above, this defines a real number x between n and $n+1$. Note that x is positive since $x > n > 0$.

Observe that $x \neq f(0)$ since $x \geq n > f(0)$. If $i \geq 1$ and $f(i) \geq 0$ then the decimal expansion of x and $f(i)$ differ in the 10^{-i} position: the first has coefficient d'_i and the second d_i. By uniqueness of decimal expansions for non nine-sequences, $x \neq f(i)$. If $f(i)$ is negative then $x \neq f(i)$ since x is positive.

In any case, $x \neq f(i)$ for each $i \geq 0$. Thus x is not in the image of f. Thus f cannot be surjective. □

In contrast $\mathbb{Q}$ is countable. This is surprising at first since $\mathbb{Q}$ is dense in $\mathbb{R}$. The first step is to recall that each $\mathbb{Q}$ can be written as a fraction a/b where $b \geq 0$ and where $a, b \in \mathbb{Z}$. It is somewhat surprising at first that $\mathbb{Z} \times \mathbb{Z}$ is countable. For our needs, we can focus on the subset Q, and show it is countable:

$$Q \stackrel{\text{def}}{=} \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid b \geq 1\}.$$

Lemma 34. *There is a bijection $\mathbb{N} \to Q$. In particular, the set Q is countable.*

Proof. (Informal) We show this by constructing a sequence of points in Q in such a way that every element of Q occurs eventually in the sequence. There are several reasonable ways of doing this. One way is to proceed as follows: start with

$$(-1,1),(0,1),(1,1)$$

then continue with

$$(-2,1),(-2,2),(-1,2),(0,2),(1,2),(2,2),(2,1)$$

then continue with

$$(-3,1),(-3,2),(-3,3),(-2,3),\ldots,(2,3),(3,3),(3,2),(3,1)$$

and so on. In the nth group we consider the subset of pairs (a,b) where the max of $|a|$ and b is equal to n. One can show that there are $4n-1$ terms in each group, but what is important in this proof is that each has a finite number of terms. Combine these finite sequences into one infinite sequence:

$$(-1,1),(0,1),(1,1),(-2,1),(-2,2),(-1,2),(0,2),(1,2),(2,2),\ldots,$$

If we define $f\colon \mathbb{N} \to \mathbb{Q}$ by sending $f(n)$ to the nth term of the sequence (where the 0th term is the start of the sequence), then this gives a bijection $\mathbb{N} \to \mathbb{Q}$. □

Theorem 35. *The set $\mathbb{Q}$ is countable.*

Proof. Start with any surjection $f\colon \mathbb{N} \to Q$, for example the map from the previous lemma. Define a function $g\colon \mathbb{N} \to \mathbb{Q}$ by the rule that $g(n) = a/b$ where $(a,b) = f(n)$.

We show that every $r \in \mathbb{Q}$ is in the image of g, so g is surjective. Since r is rational, it can be written as a/b for some $(a,b) \in Q$. Since f is surjective, there is an $n \in \mathbb{N}$ such that $f(n) = (a,b)$. Thus

$$g(n) = a/b = r.$$

In other words, r is in the image of g. □

A rigorous proof of Lemma 34 (optional)

In this optional section we outline a rigorous proof of the existence of a bijection $\mathbb{N} \to Q$ where

$$Q \stackrel{\text{def}}{=} \{(a,b) \in \mathbb{Z} \times \mathbb{Z} \mid b \geq 1\}.$$

First we define an order on Q. It will be a bit different than the order described informally above, but it is easier to describe and work with. Given pairs $(a_1, b_1), (a_2, b_2) \in Q$ we define $(a_1, b_1) < (a_2, b_2)$ to hold if and only if one of the following occurs

(i) $\max(|a_1|, b_1) < \max(|a_2|, b_2)$.

(ii) $\max(|a_1|, b_1) = \max(|a_2|, b_2)$ and $a_1 < a_2$.

(iii) $\max(|a_1|, b_1) = \max(|a_2|, b_2)$ and $a_1 = a_2$ and $b_1 < b_2$.

Observe that $(-1, 1)$ is the minimum element of Q, but that Q has no maximum with this order. We show that $<$ is indeed an order relation:

Lemma 36. *The relation $<$ is a strict linear order on Q.*

Proof. For transitivity, assume $(a_1, b_1) < (a_2, b_2)$ and $(a_2, b_2) < (a_3, b_3)$. The condition $(a_1, b_1) < (a_2, b_2)$ divides into three cases, and the condition $(a_2, b_2) < (a_3, b_3)$ divides into three cases. For each of the nine combined possibilities, it is immediate that $(a_1, b_1) < (a_3, b_3)$.

In order to prove the trichotomy property for $(a_1, b_1), (a_2, b_2) \in Q$ divide into cases: $\max(|a_1|, b_1) \neq \max(|a_2|, b_2)$ or $\max(|a_1|, b_1) = \max(|a_2|, b_2)$. In the later case divide further into subcases: $a_1 \neq a_2$ or $a_1 = a_2$. □

By the above lemma, Q is an ordered set with $<$.

Lemma 37. *With the order $<$ defined above, Q is well-ordered.*

Proof. Let S be a nonempty subset of Q. We will show that S has a minimum. First consider the set T_1 of natural numbers that can be written as $\max(|a|, b)$ for some $(a, b) \in S$. Note that T_1 has a minimum t since $\mathbb{N}$ is well-ordered (Chapter 2). Let S_1 be the subset of S consisting of pairs $(a, b) \in S$ such that $t = \max(|a|, b)$. Observe that if $(a, b) \in S_1$ then $(a, b) < (a', b')$ for each $(a', b') \in S - S_1$ since $\max(|a'|, b')$ must be strictly greater than t.

Now let A be the following set:

$$A \stackrel{\text{def}}{=} \{a \in \mathbb{Z} \mid \text{there is a } b \text{ with } (a, b) \in S_1\}.$$

Observe that A is a nonempty subset of $\mathbb{Z}$ and has lower bound $-t$. Thus it must have a minimum a_0 (by a theorem of Chapter 4). Let S_2 be the set of pairs $(a_0, b) \in S_1$. Observe that if $(a_0, b) \in S_2$ then $(a_0, b) < (a', b')$ for each $(a', b') \in S_1 - S_2$ since $a_0 < a$. Combining with a previous result we get, in fact, that $(a_0, b) < (a', b')$ for each $(a_0, b) \in S_2$ and $(a', b') \in S - S_2$.

Observe also that S_2 is nonempty, so the set

$$B \stackrel{\text{def}}{=} \{b \in \mathbb{N} \mid (a_0, b) \in S_2\}$$

is nonempty. Let b_0 be its minimum. Observe that if (a_0, b_0) is the minimum of S_2. Conclude that (a_0, b_0) is actually the minimum of all of S. □

Now we are ready to define a bijection $f\colon \mathbb{N} \to Q$. We define this recursively by the conditions (i) $f(0) = (-1, 1)$ and (ii) $f(n+1)$ is the smallest element of Q strictly greater than $f(n)$. Condition (ii) is well-defined since Q is well-ordered and has no maximum. So by the principle of recursive definition, f is defined.

We still need to show that f is bijective. By induction we can show, for any fixed $n \in \mathbb{N}$, that $f(n+k) > f(n)$ for all $k \geq 1$. A corollary of this is that f is injective.

Suppose that $(a, b) \in Q$ is not in the image of f. By induction we can show that $f(n) < (a, b)$ for all $n \in \mathbb{N}$. Since f is injective, and since every element in the image is less than (a, b), this shows (as seen in Chapter 3) that the set of elements S less than (a, b) is infinite. However, if $t = \max(|a|, b)$

$$S \subseteq \{-t, \ldots, t\} \times \{1, \ldots, t\}$$

which is finite. A contradiction. We conclude that every element of Q is in the image of f.

Thus $f\colon \mathbb{N} \to Q$ is a bijection as desired.

More on countability (optional)

We will give yet another argument that

$$Q \stackrel{\text{def}}{=} \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid b \geq 1\}.$$

is countable. While doing so we will establish some additional important properties of countable sets.

Theorem 38. *Every subset of a countable set is countable.*

Proof. Suppose $S \subseteq T$ and T is countable. Our goal is to show S is countable, and if S is empty, it is countable by definition. So we will assume S is nonempty, and fix an element $s_0 \in S$. In this case T is nonempty as well.

Since, in this case, T is countable and nonempty, there is a surjective function $f\colon \mathbb{N} \to T$. Define a function $g\colon \mathbb{N} \to S$ as follows:

$$g(n) = \begin{cases} f(n), & \text{if } f(n) \in S \\ s_0 & \text{otherwise} \end{cases}$$

Given $s \in S$ we have $s \in T$, so there is an $n \in \mathbb{N}$ so that $f(n) = s$ since f is surjective. Observe that $g(n) = f(n) = s$. We have established that g is surjective, and so S is countable. □

Theorem 39. *Let $f\colon S \to T$ be a function. If f is surjective and S is countable then T is countable. If f is injective and T is countable then S is countable. Hence, if f is bijective then one of S and T is countable if and only if both are countable.*

Proof. First assume that f is surjective. If S is empty, T is empty so countable. Otherwise there is a surjection $\mathbb{N} \to S$ by definition of countable. Thus there is a surjection $\mathbb{N} \to T$ by composition.

Now assume that $f\colon S \to T$ is injective, and let T_0 be the image of f in T. From f we get a bijection $S \to T_0$. Let $g\colon T_0 \to S$ be the inverse. By the previous theorem, T_0 is countable, and note that $g\colon T_0 \to S$ is surjective. Thus S must also be countable by the first part of the current theorem. □

Now we give a third proof for the countability of Q:

Theorem 40. *The set $Q \stackrel{\text{def}}{=} \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid b \geq 1\}$ is countable.*

Proof. Let $f\colon Q \to \mathbb{Z}$ be defined by the rule

$$f(a, b) = 2^b(2a + 1).$$

We begin by showing that f is injective. Assume

$$2^{b_1}(2a_1 + 1) = 2^{b_2}(2a_2 + 1)$$

with $b_2 \geq b_1$. Then

$$(2a_1 + 1) = 2^{b_2 - b_1}(2a_2 + 1).$$

The left side is odd, so the right must be as well. Thus $b_2 = b_1$ and so we also have $(2a_1 + 1) = (2a_2 + 1)$. From the later equation we get $2a_1 = 2a_2$, and by cancelling we get $a_1 = a_2$. Thus $(a_1, b_1) = (a_2, b_2)$.

We have established the injectivity of the function $f\colon Q \to \mathbb{Z}$. Since $\mathbb{Z}$ is countable, we use the previous theorem to conclude Q is as well. □

Chapter 12

The Complex Numbers $\mathbb{C}$

12.1 Introduction

Although $\mathbb{R}$ is a complete ordered field, mathematicians do not stop at real numbers. The real numbers are limited in various ways. For example, not every polynomial with real coefficients factors into linear polynomials with real coefficients. This is related to the fact that there are real polynomials such as $x^2 + 1$ or $x^4 + 2x^2 + 5$ that have no real roots. The need to solve polynomial equations gave rise to the complex numbers.

As we know from basic algebra, when we work with quadratic equations sometimes the discriminant $b^2 - 4ac$ is negative, and in those cases we need to use complex numbers to find roots. The complex numbers allow us to use the quadratic equation successfully in all circumstances. However, the complex numbers did not arise first from quadratic equations. When a quadratic equation has no real solutions, why look for any solution at all? Wouldn't be easier to declare the problem unsolvable? This was the tactic used by early algebraists. It was later, in Renaissance Italy when the cubic and quartic equations were investigated, that square roots of negative numbers were first used. In fact, these so-called "imaginary numbers" are needed in the cubic equation even when looking for real solutions. Imaginary quantities arise in intermediate steps. For real solutions, the imaginary parts cancel out by the last step, but complex number arithmetic is required in intermediate computations. This means that you cannot avoid the complex numbers even when your goal is to find real solutions.

At first the complex numbers were viewed as fictitious numbers which were useful sometimes in finding "real" solutions. Later, about 1800, the idea arose of treating complex numbers as points in the plane. This made the

complex numbers into tangible, non-fictitious objects. We will follow this approach and define $\mathbb{C}$ as $\mathbb{R} \times \mathbb{R}$. After defining addition and multiplication on $\mathbb{C}$, our goal will be to establish that $\mathbb{C}$ is a field. However, this field cannot be made into an ordered field. Even though it is not an ordered field, we can still define an absolute value on $\mathbb{C}$. We consider some of the properties of this absolute value including the triangle inequality.

The topics of finding roots of complex numbers and roots of polynomials are treated in appendices. These topics are very important, but fall out of the regular scope of this book which aims to treat number systems axiomatically in a self-contained manner. These topics fall out of this scope since will need to rely on trigonometry and other subjects to make further progress in the complex numbers.

12.2 Basic definitions

Definition 1 (Complex numbers)**.** Define the set $\mathbb{C}$ of complex numbers as follows:

$$\mathbb{C} \stackrel{\text{def}}{=} \{(x, y) \mid x, y \in \mathbb{R}\}.$$

Remark 1. Recall that in set theory if S is a set then S^2 is defined to be $S \times S$ where $\times$ is the Cartesian product. So, as a set, $\mathbb{C}$ is just $\mathbb{R}^2$. There are differences between $\mathbb{C}$ and $\mathbb{R}^2$ when we start talking about binary operations. For example, the complex numbers $\mathbb{C}$ have a multiplication defined as a true binary operation, but $\mathbb{R}^2$ is typically given only a scalar multiplication.[1]

We now consider addition and a multiplication on $\mathbb{C}$.

Definition 2 (Addition and multiplication)**.** Suppose that (x_1, y_1) and (x_2, y_2) are in $\mathbb{C}$. Then

$$(x_1, y_1) + (x_2, y_2) \stackrel{\text{def}}{=} (x_1 + x_2,\ y_1 + y_2)$$

and

$$(x_1, y_1) \cdot (x_2, y_2) \stackrel{\text{def}}{=} (x_1 x_2 - y_1 y_2,\ x_1 y_2 + x_2 y_1).$$

The addition and multiplication on the right hand side of these equations are the addition and multiplication in $\mathbb{R}$ defined in Chapter 10.

Remark 2. Thus $\mathbb{C}$ has two binary operations: addition $+ : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and multiplication $\cdot : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$.

Exercise 1. Prove the following theorem.

Theorem 1. *Addition and multiplication on* $\mathbb{C}$ *are associative.*

[1] However, vector addition in $\mathbb{R}^2$ and addition in $\mathbb{C}$ do correspond.

Exercise 2. Prove the following two theorems.

Theorem 2. *Addition and multiplication on $\mathbb{C}$ are commutative.*

Theorem 3. *Addition and multiplication on $\mathbb{C}$ satisfy the distributive law.*

12.3 The canonical embedding of $\mathbb{R}$ in $\mathbb{C}$

We want to view the complex numbers as an extension of the real numbers. In other words, we want to think of $\mathbb{R}$ as a subset of $\mathbb{C}$. Our definition (Definition 1) does not define $\mathbb{C}$ in such a way to make $\mathbb{R}$ automatically a subset. In order to regard $\mathbb{R}$ as a subset of $\mathbb{C}$ we need an injective function that embeds $\mathbb{R}$ into $\mathbb{C}$.

Definition 3 (Canonical embedding)**.** The *canonical embedding* $\mathbb{R} \to \mathbb{C}$ is the function defined by the rule

$$x \mapsto (x, 0).$$

Theorem 4. *The canonical embedding $\mathbb{R} \to \mathbb{C}$ is injective.*

Proof. Call the canonical embedding F. To show that $F : \mathbb{R} \to \mathbb{C}$ is injective we must show that, for all $a, b \in \mathbb{R}$, if $F(a) = F(b)$ then $a = b$.

Suppose $F(a) = F(b)$ where $a, b \in \mathbb{R}$ are arbitrary. Then $(a, 0) = (b, 0)$. By the definition of ordered pair (in set theory), this implies that the first coordinates are equal and that the second coordinates are equal. Since the first coordinates are equal, $a = b$ as desired. □

If we identify $x \in \mathbb{R}$ with its image $(x, 0)$ in $\mathbb{C}$, then we can think of $\mathbb{R}$ as a subset of $\mathbb{C}$. So, from now on, if $x \in \mathbb{R}$ then we will think of x and $(x, 0)$ as being the same number.

In particular, $0 \in \mathbb{R}$ can be identified with $(0, 0)$, and $1 \in \mathbb{R}$ can be identified with $(1, 0)$.

Theorem 5. *The number 0 is an additive identity for $\mathbb{C}$ and 1 is a multiplicative identity for $\mathbb{C}$.*

Proof. Identify 0 with $(0, 0)$. We must show that $(0, 0)$ is the additive identity. This follows from Definition 2 and the fact that $\mathbb{R}$ is a ring (Ch. 10):

$$(x, y) + (0, 0) = (x + 0, y + 0) = (x, y)$$

for all $(x, y) \in \mathbb{C}$. Likewise, $(0, 0) + (x, y) = (x, y)$ by the commutative law. Checking that 1 is a multiplicative identity is left as an exercise. □

Exercise 3. Check that 1 is a multiplicative identity for $\mathbb{C}$.

Since we now think of $\mathbb{R}$ as a subset of $\mathbb{C}$ we have to be careful with $+$ and $\cdot$ in $\mathbb{R}$. We defined these operations for $\mathbb{R}$ in one way in Chapter 10, and then defined them for $\mathbb{C}$ in the current chapter. Do we get the same answer for real numbers $a, b \in \mathbb{R}$ as for the corresponding complex numbers $(a, 0)$ and $(b, 0)$? The answer is yes since, using Definition 2,

$$(a, 0) + (b, 0) = (a + b,\ 0 + 0) = (a + b,\ 0)$$

and

$$(a, 0) \cdot (b, 0) = (a \cdot b - 0 \cdot 0,\ a \cdot 0 + 0 \cdot b) = (ab,\ 0).$$

We summarize the above observations as follows.

Theorem 6. *Consider $\mathbb{R}$ as a subset of $\mathbb{C}$. Then the addition and multiplication operations on $\mathbb{C}$ extend the corresponding binary operations on $\mathbb{R}$.*

12.4 The square root of −1

The complex numbers possesses a number whose square is -1.

Definition 4. Let i be the complex number $(0, 1)$.

Remark 3. Observe that i is not in the image of the canonical embedding $\mathbb{R} \to \mathbb{C}$. In other words, it is not a real number.

We now show the key property of i.

Theorem 7. *The number $i \in \mathbb{C}$ satisfies the equation*

$$i^2 = -1.$$

Exercise 4. Use Definition 2 and the canonical embedding to prove the theorem.

Remark 4. Because of this theorem, we call i the *square root of* -1. However, this terminology is a bit deceptive: the square root of -1 is not unique since $(0, -1)$ is also a square root of -1.

Informal Exercise 5. Show that $i = (0, 1)$ and $(0, -1)$ are the only square roots of -1. Hint: suppose that $(x, y) \cdot (x, y) = (-1, 0)$ and show that $x = 0$ and $y = \pm 1$. You can use the fact that, for real numbers, the only square roots of positive 1 are ± 1. You also know that $x^2 \geq 0$.

Remark 5. The complex numbers $\mathbb{C}$ cannot be thought of as an ordered field. To see this, consider $i^2 = -1$. In an ordered field, all squares are nonnegative but -1 is always negative (since 1 must be positive).

12.5 Standard form of complex numbers

We do not typically write complex numbers as ordered pairs: we like to write $x + yi$ for (x, y). We now establish that (x, y) and $x + yi$ are indeed the same complex number:

Theorem 8. *Let $(x, y) \in \mathbb{C}$ be a complex number. Then*

$$(x, y) \;=\; x + yi.$$

Proof. Observe that

$$\begin{aligned}
x + yi &= (x, 0) + (y, 0)i && \text{(Canonical Embed.)} \\
&= (x, 0) + (y, 0) \cdot (0, 1) && \text{(Def. 4)} \\
&= (x, 0) + (y \cdot 0 - 0 \cdot 1,\ y \cdot 1 + 0 \cdot 0) && \text{(Def. 2)} \\
&= (x, 0) + (0, y) && \text{(Laws in Ch. 10: } \mathbb{R} \text{ is a ring)} \\
&= (x, y). && \text{(Def. 2)}
\end{aligned}$$

□

Remark 6. By the above theorem, we can think of the set $\mathbb{C}$ as follows:

$$\mathbb{C} \;=\; \{x + yi \mid x, y \in \mathbb{R}\}.$$

Theorem 9. *Let $x + yi$ and $v + wi$ be complex numbers where $x, y, v, w \in \mathbb{R}$. Then*

$$x + yi = v + wi \quad \iff \quad x = v \text{ and } y = w.$$

Proof. The ($\Leftarrow$) direction uses the substitution law (of equality).

We wish to prove the ($\Rightarrow$) direction, so suppose $x + yi = v + wi$. By Theorem 8,

$$(x, y) = x + yi \quad \text{and} \quad (v, w) = v + wi.$$

Thus $(x, y) = (v, w)$. By set theory, two ordered pairs are equal if and only if their components are equal. So $x = v$ and $y = w$. □

Remark 7. The above theorems shows that every complex number can be written uniquely in the form $x + yi$ where $x, y \in \mathbb{R}$.

12.6 The complex numbers $\mathbb{C}$ as a ring

We almost have everything we need to establish that $\mathbb{C}$ is a commutative ring: we have commutative, associative, distributive laws, and additive and multiplicative identities. We also need additive inverses:

Theorem 10. *Every complex number has an additive inverse. In fact, if $x + yi$ be a complex numbers with $x, y \in \mathbb{R}$, then $(-x) + (-y)i$ is the additive inverse of $x + yi$. In other words,*

$$-(x + yi) = (-x) + (-y)i$$

where the inverse on the left denotes additive inverse in $\mathbb{C}$, while the inverses on the right denote additive inverses in $\mathbb{R}$.

Proof. Observe that

$$\begin{aligned} \big((-x) + (-y)i\big) + (x + yi) &= (-x, -y) + (x, y) \quad \text{(Thm. 8)} \\ &= (-x + x, -y + y) \quad \text{(Def. 2)} \\ &= (0, 0) \qquad \text{(Chapter 10 laws about } \mathbb{R}\text{)} \\ &= 0. \qquad \text{(Use of canonical embed.)} \end{aligned}$$

Since $\big((-x) + (-y)i\big) + (x + yi) = 0$, and since addition is commutative, the result follows. □

We now have everything we need for the following:

Theorem 11. *The set of complex numbers $\mathbb{C}$ is a commutative ring.*

Exercise 6. Review the definition of commutative ring, and verify that we have indeed proved everything we need for the above theorem. Cite where each was done.

Remark 8. Now we can use all the laws that hold in general rings. For example, we know that if z is a complex number, then $0 \cdot z = 0$. Of course we could verify this directly, but the point is we do not have to: such a law holds in all rings. Likewise, $-(-z) = z$ for all $z \in \mathbb{C}$ since such an identity is true in all rings. Furthermore, if $z, w \in \mathbb{C}$ then

$$-(zw) = (-z)w = z(-w)$$

since such a law holds in all rings. Also $(-1)z = -z$ and $(-z)(-w) = zw$ and $-(z + w) = (-z) + (-w)$ for all $z, w \in \mathbb{C}$ since such laws holds in all rings.

Remark 9. Theorem 10 says that

$$-(x + yi) = (-x) + (-y)i$$

where the left-hand use of $-$ is additive inverse in $\mathbb{C}$, and the right hand use of $-$ is additive inverse in $\mathbb{R}$. When $y = 0$ we get $-x = -x$ where left-hand use of $-$ is additive inverse in $\mathbb{C}$, and the right hand use of $-$ is additive inverse in $\mathbb{R}$. Thus the two definitions of inverse agree. In other words, the additive inverse of $\mathbb{C}$ extends that of $\mathbb{R}$.

As in any ring, we define $z - w$ as $z + (-w)$. Since both the additive inverse in $\mathbb{C}$ and the addition in $\mathbb{C}$ extends the corresponding operations in $\mathbb{R}$, we can conclude that subtraction in $\mathbb{C}$ extends subtraction in $\mathbb{R}$.

Since $\mathbb{C}$ is a ring, $-(z-w) = w - z$, $(z+w) - w = z$, $(z-w) + w = z$, etc. are automatically true.

Remark 10. Now that we know $\mathbb{C}$ is a ring, we can rederive and provide some motivation for Definition 2. In other words, if we forget Definition 2, we can rederive formulas for addition and multiplication. For addition:

$$\begin{aligned}(x+yi)+(v+wi) &= (x+v)+(yi+wi) && \text{(Assoc. and Comm.)}\\ &= (x+v)+(y+w)i. && \text{(Distributive Law)}\end{aligned}$$

(Here the first step combines several uses of the associative and commutative laws.) For multiplication:

$$\begin{aligned}(x+yi)(v+wi) &= x(v+wi)+(yi)(v+wi) && \text{(Distributive Law)}\\ &= xv+x(wi)+(yi)v+(yi)(wi) && \text{(Distributive Law)}\\ &= xv+(xw)i+(yv)i+(yw)i^2 && \text{(Assoc/comm. for mult)}\\ &= xv+(xw)i+(yv)i+(yw)(-1) && \text{(Thm. 7)}\\ &= xv-yw+(xw)i+(yv)i && \text{(Properties of rings)}\\ &= (xv-yw)+(xw+yv)i. && \text{(Distr. law)}\end{aligned}$$

Another way of saying this is that the formulas for addition and multiplication are a result of the fact that $i^2 = -1$ and that $\mathbb{C}$ is a ring. If someone constructed the complex numbers in another way such $i \in \mathbb{C}$ with $i^2 = -1$, such that this version of $\mathbb{C}$ is a ring, and such that every element is of the form $x + yi$ with $x, y \in \mathbb{R}$, then that person would have the same formulas for addition and multiplication as we do.[2]

12.7 Complex conjugation

We need to show that $\mathbb{C}$ is not just a ring, but is a field. To do this we need to show that every nonzero element has a multiplicative inverse. We will need complex conjugation to show how to form the multiplicative inverses.

Definition 5 (Complex conjugation)**.** Suppose $z \in \mathbb{C}$ where $z = x + yi$ with $x, y \in \mathbb{R}$. Then

$$\overline{z} \stackrel{\text{def}}{=} x - yi.$$

The complex number $\overline{z}$ is called the *complex conjugate* of z.

Theorem 12. *Let $z \in \mathbb{C}$. Then $\overline{z} = z$ if and only if z is a real number.*

[2] A fancy way of saying this is that all rings with these properties are canonically isomorphic.

Exercise 7. Show the above theorem. Hint: use Theorem 9 and properties of the real numbers.

Exercise 8. Prove the following two theorems.

Theorem 13. *Let $w, z \in \mathbb{C}$. Then*

$$\overline{z+w} = \overline{z} + \overline{w} \qquad \text{and} \qquad \overline{zw} = \overline{z}\,\overline{w}.$$

Theorem 14. *If $z \in \mathbb{C}$ then $\overline{\overline{z}} = z$.*

Theorem 15. *If $z \in \mathbb{C}$ then $-\overline{z} = \overline{-z}$.*

Proof. By Theorem 13

$$\overline{-z} + \overline{z} = \overline{(-z) + z} = \overline{0} = 0.$$

Now subtract $\overline{z}$ from both sides. □

Corollary 16. *If $z, w \in \mathbb{C}$ then $\overline{z-w} = \overline{z} - \overline{w}$.*

Proof. This follows from the definition of $z - w$ as $z + (-w)$ together with Theorems 13 and 15. □

Theorem 17. *Let $z \in \mathbb{C}$. If $z = x + yi$ with $x, y \in \mathbb{R}$ then*

$$z\overline{z} = x^2 + y^2.$$

Exercise 9. Prove the above theorem.

Theorem 18. *Let $z \in \mathbb{C}$. Then $z\overline{z}$ is a nonnegative real number. Furthermore, if $z \neq 0$ then $z\overline{z} > 0$.*

Proof. Write z as $x + yi$ with $x, y \in \mathbb{R}$. From Theorem 17 we know that

$$z\overline{z} = x^2 + y^2.$$

In particular, $z\overline{z}$ is a real number since $\mathbb{R}$ is closed under addition and multiplication.

If $z = 0$ the result is nonnegative since $0^2 + 0^2 = 0$.

If $z \neq 0$ then either x or y (or both) is nonzero. Suppose, for example, that x is nonzero. Since the product of two positive numbers is positive, and the product of two negative numbers is also positive, we have $x^2 > 0$ regardless of whether x is positive or negative (ordered fields, Chapter 8). Also $y^2 \geq 0$ (if $y > 0$ it follows as for x, if $y = 0$ then $y^2 = 0$). Thus

$$x^2 + y^2 > 0 + y^2 \geq 0 + 0 = 0$$

by properties of ordered fields (Chapter 8). Thus $z\overline{z} > 0$. A similar argument shows the result if $y \neq 0$. □

12.8 The complex numbers $\mathbb{C}$ as a field

The complex numbers form a field. To see this we need to to check that $1 \neq 0$ which is obvious (since $1 \neq 0$ in $\mathbb{R}$, and since $\mathbb{R}$ embeds into $\mathbb{C}$), and that every nonzero element has a multiplicative inverse. Suppose that $z \in \mathbb{C}$ is nonzero. Then $z\overline{z}$ is a positive real number by Theorem 18. Since $\mathbb{R}$ is a field, and since $z\overline{z} \neq 0$, the multiplicative inverse $(z\overline{z})^{-1}$ exists in $\mathbb{R}$ (and hence in $\mathbb{C}$). Consider

$$w \stackrel{\text{def}}{=} (z\overline{z})^{-1}\,\overline{z}.$$

Then multiplying both sides by z we get

$$wz = (z\overline{z})^{-1}\,z\overline{z} = 1.$$

So z has a multiplicative inverse. We can use this fact to prove the following.

Theorem 19. *The set of complex numbers $\mathbb{C}$ is a field.*

Exercise 10. Review the definition of *field*, and verify that we have proved everything we need for the above theorem.

Informal Exercise 11. Use the above formula for w to find the multiplicative inverse of $z = 2 + i$. Write your answer in the form $a + bi$ with $a, b \in \mathbb{R}$.

Informal Exercise 12. Find the multiplicative inverses of i and $3i$.

Informal Exercise 13. Convert

$$z = \frac{7 + 2i}{2 + 3i}$$

to the form $x + yi$ with $x, y \in \mathbb{R}$.

Exercise 14. Let $z \in \mathbb{C}$ be nonzero. Show that $\overline{z}^{-1} = \overline{z^{-1}}$. In addition, let $w \in \mathbb{C}$. Show that

$$\overline{\left(\frac{w}{z}\right)} = \frac{\overline{w}}{\overline{z}}.$$

12.9 Absolute values in $\mathbb{C}$ and the triangle inequality

In Chapters 9 and 11 we showed that every nonnegative real number x has a unique nonnegative square root $\sqrt{x}$. The square root is used in the definition of absolute value in $\mathbb{C}$.

Definition 6 (Absolute value)**.** Let $z \in \mathbb{C}$. Then the *absolute value* $|z|$ of z is defined as follows:

$$|z| \stackrel{\text{def}}{=} \sqrt{z\overline{z}}.$$

Remark 11. Observe that if z is the point (x, y) then the above definition is equivalent to defining $|z|$ as $\sqrt{x^2 + y^2}$. Informally, we recognize this as the distance from (x, y) to the origin (Pythagorean theorem). This is analogous to the absolute value in $\mathbb{R}$ where the absolute value of a number is (informally) the distance of the number to 0. Observe that $|z| \geq 0$ by definition of square root, and that if z is real then this absolute value gives the same value as the absolute value defined for $\mathbb{R}$.

Finally, In order for this to be a well-behaved absolute value, we would want it to satisfy such familiar properties as the identity $|zw| = |z||w|$ and the i nequality $|z + w| \leq |z| + |w|$. These will be proved below.

Theorem 20. *If $z, w \in \mathbb{C}$ then*

$$|zw| = |z| \cdot |w|.$$

Proof. Observe that

$$\begin{aligned}
(|z| \cdot |w|)^2 &= |z|^2|w|^2 && \text{(Expon. Law, Ch. 6)} \\
&= z\overline{z}\, w\overline{w} && \text{(Def 6)} \\
&= zw\,\overline{z}\,\overline{w} && \text{(Comm./Assoc. Laws)} \\
&= zw\,\overline{zw} && \text{(Thm. 13)} \\
&= |zw|^2. && \text{(Def 6)}
\end{aligned}$$

By a result of Chapter 11, this implies that $|z| \cdot |w| = |zw|$. □

Theorem 21. *If $z \in \mathbb{C}$ then $|-z| = |z|$.*

Proof. By Theorem 20,

$$|-z|\,|-z| = |(-z)(-z)| = |zz| = |z||z|.$$

Thus $|-z|^2 = |z|^2$. So $|-z| = |z|$ by a result of Chapter 11. □

Theorem 22. *Suppose $z \in \mathbb{C}$. Then $|z| = 0$ if and only if $z = 0$.*

Exercise 15. Prove the above theorem.

Theorem 23. *Suppose $z = x + yi$ where $z \in \mathbb{C}$ and $x, y \in \mathbb{R}$. Then*

$$|x| \leq |z| \qquad \textit{and} \qquad |y| \leq |z|.$$

Proof. (sketch) Since $x^2 \geq 0$, we have $x^2 + y^2 \geq x^2$. Observe that $|x|^2 = x^2$ and $|z|^2 = x^2 + y^2$. Thus $|z|^2 \geq |x|^2$. So $|z| \geq |x|$.

A similar argument shows $|z| \geq |y|$. □

Now we wish to show the triangle inequality.

Lemma 24. *If $z \in \mathbb{C}$ then*

$$|z+1| \leq |z|+1.$$

Proof. Let $z = x + yi$ where $x, y \in \mathbb{R}$. Then $z + 1 = (x + 1) + yi$. Thus

$$|z+1|^2 = (x+1)^2 + y^2 = x^2 + 2x + 1 + y^2 = (x^2 + y^2) + 2x + 1.$$

If $x \geq 0$ then $x \leq |z|$ by Theorem 23. If $x < 0$ then $x \leq |z|$ since $|z| \geq 0$. In either case $x \leq |z|$. So

$$\begin{aligned} |z+1|^2 &= (x^2+y^2) + 2x + 1 \\ &= |z|^2 + 2x + 1 \\ &\leq |z|^2 + 2|z| + 1 \\ &= (|z|+1)^2. \end{aligned}$$

By a result in Chapter 11, this implies $|z + 1| \leq |z| + 1$. □

Theorem 25 (Triangle Inequality in $\mathbb{C}$)**.** *Let $z, w \in \mathbb{C}$. Then*

$$|z+w| \leq |z| + |w|.$$

Proof. If $w = 0$ then the result is clear. So assume $w \neq 0$. Let $u = zw^{-1}$. By Lemma 24,

$$|u+1| \leq |u| + 1.$$

Multiply both sides by $|w|$: So

$$|u+1||w| \leq (|u|+1)|w| = |uw| + |w| = \left|zw^{-1}w\right| + |w| = |z| + |w|.$$

However,

$$|u+1||w| = |(u+1)w| = |uw + w| = \left|zw^{-1}w + w\right| = |z+w|.$$

So

$$|z+w| \leq |z| + |w|.$$

□

Informal Exercise 16. Illustrate the triangle inequality with a picture using two particular complex numbers z and w. This picture should be such that there is a triangle with side lengths $|z|$, $|w|$ and $|z + w|$.

Appendices

Appendix A

"Chapter 0". Basic logic and set theory

This appendix reviews basic logic and some of the other set theoretical background needed for the course.

A.1 The logical basis

The purpose of this book is to systematically develop the number systems commonly used in mathematics. A second purpose is to illustrate the axiomatic method through the development of these number systems. In the spirit of the axiomatic method, our development of the number systems will be rigorous and self-contained: we will give careful proofs for our results. There are, however, two exceptions where we will allow results without proof:

1. *Axioms.* These are fundamental statements that are accepted without the need for formal justification. Sometimes they are presented as "self-evident", but technically they do not need to be obvious. They are, however, accepted as true for the purpose of proving further results.

   In this course the only axioms are the Dedekind-Peano axioms and the iteration axiom. In an optional section near the end of Chapter 1 the iteration axiom will be shown to be a consequence of the other axioms, so the only axioms that are necessary for this course are the Dedekind-Peano axioms.[1] In more advanced mathematics, the axiom of choice,

[1] These axioms, coupled with some basic set theory, suffice for a large part of mathe-

and certain advanced set theoretic axioms are also sometimes needed.

2. *Principles of logic and elementary set theory.* From the axioms, we will derive other results using logic. So we will take as given the knowledge of classical deductive logic. This logic can be used freely to derive new results. For example, we assume the basic principles related to connectives ($\wedge$, $\vee$, $\implies$, $\neg$, $\iff$) quantifiers ($\forall$, $\exists$, $\exists!$), and equality ($=$). The principles of classical first-order logic will be reviewed below from the point of view of Gentzen-style natural deduction.

   We will regard elementary set theory as part of our logical background and toolkit. These includes concepts, rules, and facts that are in common use in modern mathematics. Included under the heading of set theory are principles concerning ordered pairs, functions, and relations as well as sets (because ordered pairs, functions, and relations can be modeled as certain types of sets). One purpose of this chapter is to outline these core principles of set theory. These principles can be developed axiomatically from a small set of axioms, but we will not do so here. We simply take them as given.[2]

Aside from the axioms, and the basic facts of logic and set theory, every statement we wish to establish or use must be proved. Even something as simple as the commutative law of addition, or even the equation $1 + 1 = 2$, will be proved.

Likewise, every *concept* not occurring in the axioms, logic, or elementary set theory must be defined before it can be used. Such a definition must use only set theoretical and logical concepts as well as previously established concepts. For example, we will define addition and multiplication using the concept of function from basic set theory. Similarly, will provide definitions for all the number systems except the natural numbers using various set-theoretical ideas applied to previously established number systems. The set of natural numbers is an exception; it will not be defined. Since the set $\mathbb{N}$ is

matics. Even geometry can be developed from these axioms. For example, once you have developed the real numbers $\mathbb{R}$, you can define the plane to be $\mathbb{R}^2$ and three-dimensional space to be $\mathbb{R}^3$. In this approach you develop all the theorems of Euclidean geometry using the coordinate point of view and no new geometric axioms are needed. This is in contrast to Euclid's original approach, updated by Hilbert, which developes geometry using geometric axioms that do not rely on the real numbers.

[2]The best known axiomatic development of set theory uses the Zermelo-Fraenkel axioms including the axiom of choice. This is a very powerful axiom system and is overkill for what we do here. The principles discussed in this chapter can be proved in a weaker axiom system akin to Zermelo's original system without the axioms of infinity, choice, replacement, or foundation. The axiom of infinity is not needed in the background set theory since the existence of infinite sets is a consequence of the Dedekind-Peano axioms introduced in Chapter 1.

part of the axioms, it does not actually need to be defined. In general, terms used in the axioms do not need to be defined, and such undefined terms are called *primitive terms*.

A.2 Proofs

We can view a proof as sequence of assertions, called "steps", each of which can be justified by appealing to previously established results, rules, previous steps of the current proof, and assumptions. The final step is the result you are trying to prove, or something that immediately yields the result.

A typical step is justified by two things (1) one or more established statements or assumptions that support the current claim, and (2) a rule of logic, set theory, or theorem of mathematics that connects these statements to the current claim. The established statements (1) can included previously proved results, prior definitions, axioms and principles taken as given, hypotheses from the statement of the theorem currently being proved, previous steps from the current proof (that are valid in the current context), and local assumptions made to specify the current context. Some of the rules of inference used for (2) will be discussed later in this chapter.

In practice, (1) the supporting facts, and (2) the rule used to justify a step are not always specifically mentioned if they are obvious from context. For example, expressions such as 'thus' or 'from this it follows' are used to indicate that the previous step or series of steps is being used as supporting facts. Often the rule (2) is clear from the claim itself. However, beginners should err on the side of supplying more details than is necessary rather than too few details. When every detail is not written down, it only be because supplying the missing details is easy to both the author of the proof and the careful reader.

There is another way to justify a step. A step can be justified by including a whole subproof for the step. For example, a step of the form $\neg P$ is often justified by including a subproof with the extra assumption P that ends with a contradiction. A step of the form $P \implies Q$ can be justified by including a subproof that starts with assumption P and ends with Q. A proof of the statement $\forall\, x \in A, P(x)$ can be proved by a subproof that starts with the assumption of $a \in A$, where a is an arbitrary but fixed element of A, and ends with $P(a)$. Another example is proof by cases: one can justify a step with a proof by cases by appealling to multiple subproofs, each involving a separate case. Subproofs can themselves have subproofs. (Warning: statements established in a subproof cannot be regarded as valid outside the subproof since they are typically proved in the context of additional assumptions that are not in force outside the subproof).

The use of subproofs in a proof is the main thing that sets proofs in real mathematics apart from the simple two column proofs of traditional high

school geometry courses. Care must be used in writing proofs to signal to the reader where the subproof ends and the main proof resumes. In other words, the reader needs to be alerted to any context change. There are several styles used to present a subproof. A subproof can be put before or after the step it justified. One might write 'Claim: P' followed by a subproof of P. A subproof can be removed from the main body of the proof and be proved separately as the proof of a *lemma*.[3] Such lemmas can be put before or after the main proof. If a lemma is put after a proof, care should be taken to make sure that the lemma is independent of the proof it is supposed to support.

A.3 Formal proofs and the axiomatic method

As discussed above, each step in a proof should be justified (sometimes with a short justification, sometimes with a long subproof). In an *informal proof* the justification is fairly open-ended. It can involve any fact or rule that is mutually accepted by the writer and intended readers. It can involve facts and rules learned in prior math courses, or, in geometry or topology for instance, facts that are obvious from one's intuition. Conclusions that the intended reader can justify without too much work on their own are often written without full justification. Many informal proofs are really proof outlines.

In a *formal proof*, on the other hand, only facts and rules that are explicitly established can be used. Appeals to prior math courses, intuition, or details for the reader to work out are not allowed. Formal proofs are particularly suited to illustrating the axiomatic method and so will be the main type of proof in this course, but from time to time informal proofs and arguments will be allowed. You can use informal proofs in your scratch work to help you develop ideas and work out examples, and in exercises that are clearly labeled as "informal". Writing a formal proof should only be done after you have a strong understanding of the statement and how it can be justified.

Tom Hales explains the distinction between informal proof and formal proof:

> Traditional mathematical proofs are written in a way to make them easily understood by mathematicians. Routine logical steps are omitted. ... Proofs, especially in topology and geometry, rely on intuitive arguments ...
>
> A formal proof is a proof in which every logical inference has been checked all the way back to the fundamental axioms of

[3] A *lemma* is a type of theorem that is not necessarily of independent interest, but is useful for establishing another result, or other results. In this class, I will use the term *claim* for a statement whose proof is embedded as a subproof in a larger proof, and the term *lemma* for a statement whose proof is separated from a larger proof.

> mathematics. All the intermediate logical steps are supplied, without exception. No appeal is made to intuition, even if the translation from intuition to logic is routine.[4]

Some go further and require that formal proofs be presented in a purely symbolic formal language, or insist that they be written in a way that a suitable computer proof-checking program could check each step in a mechanical manner. We will not go that far, but will adhere to fairly strict standards, especially in the early part of the course.

The *axiomatic method* is the technique of carefully developing a body of results from a small set of axioms. The development of an axiomatic theory ideally should be rigorous and self-contained. The historic inspiration for the axiomatic method was Euclid's *Elements of Geometry*. In this course we will illustrate the axiomatic method in the development of the basic number systems.

There is a major psychological difficulty in using the axiomatic method: one starts by proving facts that are already known or obvious. In our development of the number systems, we need to pretend ignorance of anything about the numbers except for what has been established in this course. This is hard to do since facts about number systems have been ingrained into our minds from such an early age.

When developing a formal proof of a results it helps to adopt a hyper-skeptical attitude. Do not accept a step until you can see the justification for the step. When constructing a proof, you want to be both creative and critical (in the good sense) until you are completely satisfied that you have a tight, rigorous proof. *One of the best skills you can develop is to know when you do and do not have a valid proof of a result.* Ideally you should be able to assess the quality of your work independent of an external reviewer.

In addition to learning to be a skeptic, in order to succeed in this course, you need to develop a strong attention to detail. Cultivate a habit of careful, slow reading. It is all right, and often advisable, to read a section quickly to get the main ideas, as long as you follow it up with a second and third careful reading. If you do so, you will develop a thorough and lasting understanding of the material, and you will find it much easier to correctly complete the exercises.

A.4 Basic rules of logic

As mentioned above, a step in a proof is typically justified by appealing to a rule of inference applied to previously established statements. Many of the rules of inference come from logic, some come from set theory, and

[4] Notices of the AMS, December 2008, page 1371.

some rules of inference will be based on theorems proved in the course. We now present some of the logical rules of inference and logical identities that are commonly used in proofs. The reader is assumed to be already familiar with most of these, and these are stated mainly for reference. Many of these rules are taken from Genzen's natural deduction approach to proof where, for each logical operator, rules will be given for establishing a statement of a certain form (often called 'introduction rules') and other rules will be given for using a statement of that form (often called 'elimination rules') to prove other statements.

Statements of the form $P \wedge Q$

("Conjunctions"). Statements of the form "P and Q" (written symbolically as $P \wedge Q$) are very well-behaved. To establish $P \wedge Q$ one can proceed by first establishing P then establishing Q. In other words, you can justify $P \wedge Q$ by citing the earlier result P and the earlier result Q. This rule is represented schematically as follows:

$$\frac{\begin{array}{c}P\\Q\end{array}}{P \wedge Q} \quad (\wedge \text{ introduction rule})$$

You can use a prior result of the form $P \wedge Q$ to justify P or justify Q:

$$\frac{P \wedge Q}{P} \qquad \frac{P \wedge Q}{Q} \qquad (\wedge \text{ elimination rules})$$

These rules extend in the obvious way to conjuncts of three or more statements:

$$P_1 \wedge P_2 \wedge P_3, \quad \text{et cetera.}$$

Statements of the form $P \vee Q$

("Disjunction" or "inclusive or"). The simplest way to establish "P or Q" (symbolically $P \vee Q$) is to first establish P or, alternatively, first establish Q, This is not the only way of doing so, but it is conceptually the simplest. These inference rules are written schematically as follows:

$$\frac{P}{P \vee Q} \qquad \frac{Q}{P \vee Q} \qquad (\vee \text{ introduction rules})$$

In practice when you want to prove $P \vee Q$ it might not be possible to prove P or to prove Q directly. But there are other techniques. Another strategy is to assume one of the two is false, and deduce the other is true. For example, suppose you *assume* P is false, and you deduce Q from this assumption (in a subproof). This does not prove Q by itself, but it does

prove $P \vee Q$. The reason why is that there are two cases: either P is true, and you are done, or P is false, in which case you showed that Q is true. In either case one of the two is true. This technique is indicated schematically as follows:

$$\frac{\neg P \Rightarrow Q}{P \vee Q} \qquad \frac{\neg Q \Rightarrow P}{P \vee Q}$$

An established statement of the form $P \vee Q$ can be used to justify a later result R via proof by cases. In such a proof you (1) prove R in a subproof (called a "case") where P is assumed to be true, and (2) prove R in a subproof where Q is assumed. Using $P \vee Q$ together with the two subproofs, you can then conclude R. In other words, from $P \vee Q$ and $P \Rightarrow R$ and $Q \Rightarrow R$, you can conclude R:

$$\frac{\begin{array}{c} P \vee Q \\ P \Rightarrow R \\ Q \Rightarrow R \end{array}}{R} \quad (\vee \text{ elimination "proof by cases"})$$

These rules extend in the obvious way to disjuncts of three or more statements:

$$P_1 \vee P_2 \vee P_3, \quad \text{et cetera.}$$

Statements of the form $P \Rightarrow Q$

("Conditionals"). A common way to establish a claim of the form "if P then Q" (symbolically $P \Rightarrow Q$) is to supply a subproof. One assumes P and derives Q in the subproof. From the existence of this subproof one is entitled to assert $P \Rightarrow Q$. This is not the only way to prove $P \Rightarrow Q$. There are other rules such as the transitive rule for $\Rightarrow$ that we will discuss below, but it is the most basic way.

You can later use a result of the form $P \Rightarrow Q$ by applying it to an established statement P to derive a statement Q. This rule is called *modus ponens*, and can be schematically indicated as follows.

$$\frac{\begin{array}{c} P \Rightarrow Q \\ P \end{array}}{Q} \quad \text{(modus ponens)}$$

(This rule is also called $\Rightarrow$ elimination).

Statements of the form $\neg P$

("Negations"). The negation of P can be justified by showing $P \Rightarrow \mathcal{F}$ (typically with a subproof) where $\mathcal{F}$ is any contradiction of a previously

established result, or any obviously false statement (for example the statement $1 \neq 1$).[5] This rule is represented as follows:

$$\frac{P \implies \mathcal{F}}{\neg P} \quad (\neg \text{ introduction rule})$$

Once you have $\neg P$, you can use it to eliminate a case. For example, if you have $P \vee Q$ and you also have $\neg P$, you can conclude Q.

$$\frac{\begin{array}{c} P \vee Q \\ \neg P \end{array}}{Q} \quad \text{(elimination of case)}$$

In classical logic, we automatically accept any statement of the form

$$P \vee \neg P,$$

but we automatically reject

$$P \wedge \neg P.$$

(Such a conjunction is considered obviously false, and is sometimes called a "contradiction").

Statements of the form $P \iff Q$

("Biconditionals"). Statements of the form "P if and only if Q" can be established with the following rule:

$$\frac{\begin{array}{c} P \implies Q \\ Q \implies P \end{array}}{P \iff Q} \quad (\iff \text{ introduction rule})$$

These statements of the form $P \Leftrightarrow Q$ can be used to justify other statements with the following inference rules:

$$\frac{P \iff Q}{P \implies Q} \qquad \frac{P \iff Q}{Q \implies P} \qquad (\iff \text{ elimination rules})$$

The connective $\Leftrightarrow$ satisfies the reflexive, symmetric, and transitive laws:

$$P \iff P \qquad \frac{P \iff Q}{Q \iff P} \qquad \frac{\begin{array}{c} P \iff Q \\ Q \iff R \end{array}}{P \iff R}$$

It also satisfies a *substitution law*: if $P \iff Q$ then you can replace any occurrence of P with Q in a larger compound statement and the result

[5] For example, if Q has been established in the current context, then $\mathcal{F}$ can be $\neg Q$, or if $\neg Q$ has been established then $\mathcal{F}$ can be Q.

will be equivalent. This is sometimes written as follows: assume $\varphi(P)$ is a compound statement in which P occurs and assume $\varphi(Q)$ is the same statement but where one or more occurrences of P have been replaced by Q:

$$\frac{P \iff Q}{\varphi(P) \iff \varphi(Q)} \qquad (\iff \text{ substitution rule})$$

Contradictions and cases

From a contradiction $(Q \wedge \neg Q)$ or any other result that is known to be false (written $\mathcal{F}$) one can derive anything you want. This is written as follows:

$$\frac{\mathcal{F}}{P} \quad \text{(contradiction rule)}$$

This is a rather strange rule at first glance, but it is useful in proofs by cases. Suppose you want to justify a step R and you decide to prove it by cases based on a previous result $P_1 \vee \ldots \vee P_n$. In other words, you give a subproof for each case P_i. Your goal is to prove R in each of these cases. Some of these cases might turn out to be impossible. For example, P_i might imply an absurdity $\mathcal{F}$. The above rule will then allow you to conclude R in that case. In other words, if a case leads to a contradiction, you can automatically move on to the next case since everything is true in a contradictory case.

Other useful rules

There are several other rules that are useful to know. (These rules can be derived from the rules we have already considered).

$$\frac{P \implies Q \quad \neg Q}{\neg P} \qquad \frac{P \implies Q \quad Q \implies R}{P \implies R} \qquad \frac{P \iff Q \quad P}{Q} \qquad \frac{P \iff Q \quad Q}{P} \qquad \frac{P \iff Q \quad \neg Q}{\neg P}$$

$$\frac{P \implies Q}{P \wedge R \implies Q \wedge R} \qquad \frac{P \implies Q}{P \vee R \implies Q \vee R} \qquad \frac{P}{Q \implies P}$$

Useful identities

The following are important logical identities including commutative and associative laws. Each line below represents a type of statement that can

be accepted as automatically true.[6]

$$\begin{array}{rcl}
P \wedge Q & \iff & Q \wedge P \\
(P \wedge Q) \wedge R & \iff & P \wedge (Q \wedge R) \\
P \vee Q & \iff & Q \vee P \\
(P \vee Q) \vee R & \iff & P \vee (Q \vee R) \\
P \wedge P & \iff & P \\
P \vee P & \iff & P \\
\neg\neg P & \iff & P \\
(P \implies Q) & \iff & \neg P \vee Q \\
(P \implies Q) & \iff & (\neg Q \implies \neg P) \\
P \vee Q & \iff & (\neg Q \implies P) \\
P \vee Q & \iff & (\neg P \implies Q)
\end{array}$$

There are two distributive laws

$$\begin{array}{rcl}
(P \vee Q) \wedge R & \iff & (P \wedge R) \vee (Q \wedge R) \\
(P \wedge Q) \vee R & \iff & (P \vee R) \wedge (Q \vee R)
\end{array}$$

and two De Morgan laws.

$$\begin{array}{rcl}
\neg(P \vee Q) & \iff & (\neg P) \wedge (\neg Q) \\
\neg(P \wedge Q) & \iff & (\neg P) \vee (\neg Q)
\end{array}$$

A.5 Quantifiers

The above illustrates "propositional logic". Logic becomes more sophisticated and powerful when we introduce *quantifiers*. This results in "predicate logic" or "quantificational logic".

We begin with the universal quantifier $\forall$. The most direct way to justify the assertion $\forall x,\ Px$ is through a subproof where a is taken to be an arbitrary but fixed object and where Pa is proved. Here Px is a predicate with variable x. (Here a should be a new term that is not being used for any other purpose in the current context).

If you already have $\forall x,\ Px$ you can use it to justify special cases of the predicate Px using the following rule which is valid for any desired a:

$$\frac{\forall x,\ Px}{Pa} \qquad \forall \text{ elimination rule}$$

The other type of quantifier is the existential quantifier $\exists$. The most direct way to justify the assertion $\exists x, Px$ is to appeal to a statement of the form Pa. Here a can be any term making the predicate Px true, it does not

[6] Such identities are sometimes called "tautologies" in logic textbooks.

have to be arbitrary in any sense. This introduction rule can be represented schematically as follows:

$$\frac{Pa}{\exists x, Px} \qquad \exists \text{ introduction rule}$$

If you have $\exists x, Px$ already, you can use it to define a new constant a representing a choice of object such that Pa is true.[7]

In addition to the two basic quantifiers $\forall$ and $\exists$, we have a third quantifier $\exists!$ ("there exists a unique") which can be defined in terms of the other two. Here are two (equivalent) definitions:

$$\begin{aligned} \exists!\, x,\, Px \quad &\overset{\text{def}}{\iff} \quad \exists x \Big(Px \ \wedge \ \forall\, y\, (\, Py \ \Rightarrow y = x) \Big) \\ &\overset{\text{def}}{\iff} \quad \Big(\exists x,\, Px \Big) \ \wedge \ \Big(\forall\, y\, \forall\, z\, \big(Py \wedge Pz \ \Rightarrow y = z \big) \Big) \end{aligned}$$

Note: '!' stands for "unique" here, but it only means "unique" when it is used after '$\exists$'. Also, be careful of the term "unique": an object cannot be unique by itself. Uniqueness only makes sense in the context of a predicate Px that such an object satisifies.

Here are some identities involving quantifiers:

$$\begin{aligned} \neg\big(\, \exists x,\, Px \,\big) \quad &\iff \quad \forall x,\, \neg Px \\ \neg\big(\, \forall x,\, Px \,\big) \quad &\iff \quad \exists x,\, \neg Px \\ \forall x\, \forall y, P(x,y) \quad &\iff \quad \forall y\, \forall x, P(x,y) \\ \exists x\, \exists y, P(x,y) \quad &\iff \quad \exists y\, \exists x, P(x,y) \\ \forall x \big(Px \ \wedge \ Qy \big) \quad &\iff \quad \big(\forall x,\ Px \big) \wedge \big(\forall x,\ Qx \big) \\ \exists x \ \big(Px \ \vee \ Qy \big) \quad &\iff \quad \big(\exists x,\ Px \big) \vee \big(\exists x,\ Qx \big) \end{aligned}$$

Warning: $\forall x\, \exists y, P(x,y)$ is not logically equivalent to $\exists y\, \forall x, P(x,y)$. The order of the quantifiers makes a big difference in meaning in this case.

A.6 Equality

Equality $=$ satisfies the reflexive, symmetric, and transitive laws:

$$a = a \qquad \frac{a = b}{b = a} \qquad \frac{\begin{array}{c} a = b \\ b = c \end{array}}{a = c}$$

The symmetry law is also written

$$a = b \iff b = a.$$

[7] Related to this is an elimination rule that allows you to justify a statement R as follows: if you have $\exists x, Px$ and if you know that $Pa \implies R$ for arbitrary a (for example if you have a subproof with assumption Pa that proves R where a is arbitrary), then you can conclude R. Here R is a statement that does not involve a.

Equality also satisfies a *substitution law*: if $a = b$ then you can replace any occurrence of a with b in a larger compound term to form an equivalent term.[8] This is sometimes written as follows: assume $\tau(a)$ is a term in which a occurs and assume $\tau(b)$ is the same term but where one or more occurrences of a have been replaced by b:

$$\frac{a = b}{\tau(a) = \tau(b)} \qquad (= \text{substitution rule})$$

There is also a second substitution rule[9] for statements: assume $\varphi(a)$ is a statement in which a occurs and assume $\varphi(b)$ is the same statement but where one or more occurrences of a have been replaced by b:

$$\frac{a = b}{\varphi(a) \iff \varphi(b)} \qquad (= \text{substitution rule 2})$$

A.7 Elementary set theory

The basics concepts, rules, and facts of set theory will be used extensively in this course.

Equality and inclusion

Two sets are equal if and only if they have the same elements:

$$A = B \iff \forall x\, (x \in A \iff x \in B)$$

The set A is a subset of B if and only if every element of A is in B:

$$A \subseteq B \iff \forall x\, (x \in A \implies x \in B)$$

Two sets are equal if and only if each is a subset of the other. This gives rise to the following rules:

$$\frac{\begin{array}{c} A \subseteq B \\ B \subseteq A \end{array}}{A = B} \qquad \frac{A = B}{A \subseteq B} \qquad \frac{A = B}{B \subseteq A}$$

We also have the following:

$$\frac{\begin{array}{c} A \subseteq B \\ x \in A \end{array}}{x \in B} \qquad \frac{\begin{array}{c} A \subseteq B \\ B \subseteq C \end{array}}{A \subseteq C} \qquad A \subseteq A$$

[8] A *term* is an expression denoting an object.

[9] Warning: in these substitution rules we assume that the bound variables are distinct from the variables occurring in a and b.

The empty set

The empty set $\varnothing$ is the set with no elements:

$$\neg\big(\exists\, x,\ x \in \varnothing\big)$$

The empty set is a subset of all sets:

$$\varnothing \subseteq A$$

Here are rules to show a set A is empty or nonempty:

$$\frac{\neg\big(\exists\, x,\ x \in A\big)}{A = \varnothing} \qquad \frac{\exists\, x,\ x \in A}{A \neq \varnothing}$$

Small sets

Small sets can be denoted by denoting all the elements. For example, the expression $\{0, 1, 4, 9, 16, 25, 36, 49, 64, 81\}$ denotes the square integers less than 100.

Here are equivalences related to small sets:

$$\begin{array}{rcl} x \in \{a\} & \iff & x = a \\ x \in \{a, b\} & \iff & (x = a) \vee (x = b) \\ x \in \{a, b, c\} & \iff & (x = a) \vee (x = b) \vee (x = c) \\ & \text{etc.} & \end{array}$$

Here are some equalities:

$$\begin{array}{rcl} \{a, b\} & = & \{b, a\} \\ \{a, a\} & = & \{a\} \end{array}$$

Intersections, unions, and differences

These are governed by the following equivalences:

$$\begin{array}{rcl} x \in A \cap B & \iff & (x \in A) \wedge (x \in B) \\ x \in A \cup B & \iff & (x \in A) \vee (x \in B) \\ x \in A - B & \iff & (x \in A) \wedge (x \notin B) \end{array}$$

They satisfy the following rules

$$\frac{\begin{array}{c} A \subseteq C \\ B \subseteq C \end{array}}{A \cup B \subseteq C} \qquad \frac{\begin{array}{c} C \subseteq A \\ C \subseteq B \end{array}}{C \subseteq A \cap B} \qquad \frac{\begin{array}{c} C \subseteq A \\ C \cap B = \emptyset \end{array}}{C \subseteq A - B}$$

They satisfy the following inclusions:

$$\begin{array}{rcl} A & \subseteq & A \cup B \\ B & \subseteq & A \cup B \\ A \cap B & \subseteq & A \\ A \cap B & \subseteq & B \\ A - B & \subseteq & A \end{array}$$

And they satisfy the following equalities:

$$\begin{array}{rcl} A \cap B & = & B \cap A \\ A \cup B & = & B \cup A \\ (A \cap B) \cap C & = & A \cap (B \cap C) \\ (A \cup B) \cup C & = & A \cup (B \cup C) \\ A \cap A & = & A \\ A \cup A & = & A \\ A \cap \varnothing & = & \varnothing \\ A \cup \varnothing & = & A \\ (A \cup B) \cap C & = & (A \cap C) \cup (B \cap C) \\ (A \cap B) \cup C & = & (A \cup C) \cap (B \cup C) \\ (A - B) \cup B & = & A \cup B \\ (A - B) \cap B & = & \varnothing \end{array}$$

Quantification over a set

The quantifier $\forall\, x \in A$ is defined by the following

$$\forall\, x \in A,\ Px \iff^{\text{def}} \forall\, x\, \big(x \in A \implies Px\big)$$

The quantifier $\exists\, x \in A$ is defined by the following

$$\exists\, x \in A,\ Px \iff^{\text{def}} \exists\, x\, \big(x \in A \wedge Px\big)$$

The most direct way to justify the assertion $(\forall\ x \in A,\ Px)$ is through a subproof where a is taken to be an arbitrary but fixed element of A and where Pa is proved.

We have the following elimination rule:

$$\frac{\begin{array}{c}\forall\, x \in A,\ Px \\ a \in A\end{array}}{Pa}$$

Observe that we can use this type of quantifier to show the subset relation:

$$A \subseteq B \iff \forall\, x \in A, (x \in B)$$

To justify $(\exists\, x \in A,\ Px)$ we have the following introduction rule:

$$\frac{\begin{array}{c} Pa \\ a \in A \end{array}}{\exists\, x \in A,\ Px}$$

If you have $(\exists\, x \in A,\ Px)$ already, you can use it to define a new constant a representing a choice of element of A such that Pa.[10]

We have a third quantifier $\exists!\ x \in A$. Here are two (equivalent) definitions:[11]

$$\begin{aligned} \exists!\, x \in A,\ Px &\overset{\text{def}}{\iff} \exists\, x \in A, \Big(Px \ \wedge\ \forall\, y \in A\ (\, Py \ \Rightarrow y = x) \Big) \\ &\overset{\text{def}}{\iff} \Big(\exists\, x \in A,\ Px \Big) \wedge \Big(\forall\, y, z \in A,\ \big(Py \wedge Pz \ \Rightarrow y = z \big) \Big) \end{aligned}$$

Here are some rules associated to these concepts:

$$\frac{\begin{array}{c} \exists\, x \in A,\ Px \\ A \subseteq B \end{array}}{\exists\, x \in B,\ Px} \qquad \frac{\begin{array}{c} \forall\, x \in B,\ Px \\ A \subseteq B \end{array}}{\forall\, x \in A,\ Px}$$

Here are some equivalences:

$$\begin{aligned} \neg\big(\exists\, x \in A,\ Px \big) &\iff \forall\, x \in A,\ \neg Px \\ \neg\big(\forall\, x \in A,\ Px \big) &\iff \exists\, x \in A,\ \neg Px \\ \forall\, x \in A,\ \forall\, y \in A,\ P(x,y) &\iff \forall\, y \in A,\ \forall\, x \in A,\ P(x,y) \\ \exists\, x \in A,\ \exists\, y \in A,\ P(x,y) &\iff \exists\, y \in A,\ \exists\, x \in A,\ P(x,y) \\ \forall\, x \in A, \big(Px \wedge Qy \big) &\iff \big(\forall\, x \in A,\ Px \big) \wedge \big(\forall\, x \in A,\ Qx \big) \\ \exists\, x \in A, \big(Px \vee Qy \big) &\iff \big(\exists\, x \in A,\ Px \big) \vee \big(\exists\, x \in A,\ Qx \big) \end{aligned}$$

Warning: as we see above, the negation of $\exists\, x \in A, Px$ is $\forall\, x \in A,\ \neg Px$ not $\forall\, x \notin A,\ \neg Px$. Remember that the negation should be a statement about elements of A, not about elements outside of A.

General unions and intersections

Let Z be a set of sets (for intersections we require that Z is nonempty). Then we have the following types of unions and intersections:

$$\bigcup Z \ = \ \bigcup_{X \in Z} X \ = \ \big\{ u \ \mid\ \exists\, X \in Z,\ u \in X \big\}$$

$$\bigcap Z \ = \ \bigcap_{X \in Z} X \ = \ \big\{ u \ \mid\ \forall\, X \in Z,\ u \in X \big\}$$

[10] Related to this is an elimination rule that allows you to justify a statement R as follows: if you have $\exists\, x \in A,\ Px$ and if you know that $Pa \implies R$ for arbitrary $a \in A$, then you can conclude R. Here R is a statement that does not involve a.

[11] The notation $\forall y, x \in A$ is short for $\forall y \in A, \forall x \in A$.

We have the following special cases:

$$\bigcup\{A\} = A, \quad \bigcup\{A, B\} = A \cup B, \quad \bigcup\{A, B, C\} = A \cup B \cup C, \quad \text{etc.}$$

$$\bigcap\{A\} = A \quad \bigcap\{A, B\} = A \cap B, \quad \bigcap\{A, B, C\} = A \cap B \cap C, \quad \text{etc.}$$

The general union and intersection are especially useful for cases where Z is an infinite set of sets.

They satisfy the following rules

$$\frac{\forall X \in Z,\ X \subseteq C}{\bigcup Z \subseteq C} \qquad \frac{\forall X \in Z,\ C \subseteq X}{C \subseteq \bigcap Z}$$

$$\frac{X \in Z}{X \subseteq \bigcup Z} \qquad \frac{X \in Z}{\bigcap Z \subseteq X}$$

A.8 Ordered pairs

An *unordered pair* is a set $\{a, b\}$. Here $\{a, b\} = \{b, a\}$. When we we want the order to be significant for equality, we use *ordered pairs*.[12]

We use (a, b) to denote the ordered pair with first coordinate a and second coordinate b. We have the following:

$$(a, b) = (c, d) \iff (a = c) \wedge (b = d)$$

The *Cartesian product* $A \times B$ of sets A and B is the set of ordered pairs with first coordinate in A and second coordinate in B:

$$A \times B = \Big\{(a, b) \mid\ (a \in A) \wedge (b \in B)\Big\}$$

We sometimes write $A \times A$ as A^2.

A.9 Functions

Modern set theory does not concern itself only with basic properties and operations on sets, but it also concerns itself with functions and their properties. In fact, functions are considered to be a special type of set: a type of set of ordered pairs. Each function has a *domain* and *codomain*.

If A and B are sets, then we write $f : A \to B$ to indicate that f is a function with domain A and codomain B. Such a function f maps each element $a \in A$ to an element $fa \in B$. We sometimes write fa as $f(a)$,

[12] In some set theory books, ordered pairs are defined in terms of unordered pairs. Sometimes (a, b) is defined as $\{\{a\}, \{a, b\}\}$, but sometimes it is defined differently. You do not need to worry about how ordered pairs are defined, but instead concentrate on the key identity $(a, b) = (c, d) \iff (a = c) \wedge (b = d)$.

especially when grouping needs to be indicated. Schematically we have the following:

$$\frac{\begin{array}{c} f : A \to B \\ a \in A \end{array}}{fa \in B}$$

We call fa the *value*, or the *image of* a. (Warning: there may be elements of B that are not of the form fa. However, if f is surjective then every element of B is indeed of the form fa.)

When are two functions equal? If $f\colon A \to B$ and $g\colon A \to B$ are functions with matching domain and codomain then

$$f = g \iff \forall\, x \in A,\ (fx = gx).$$

We will use a notation that distinguishes between two kinds of arrows, written as $\to$ and $\mapsto$. We use $\to$ as above to indicate domain and codomain. We use $\mapsto$ to illustrate a definition or description of a function. More specifically, if we want to define $f : A \to B$ by a rule, we sometimes indicate the rule by writing an expression of the form $x \mapsto \varphi(x)$. Here x stands for an arbitrary element of the domain, and $\varphi(x)$ is an expression for the value of the function.

Composition

Suppose $f\colon A \to B$ and $g\colon B \to C$ are functions such that the codomain of f is equal to the domain of g. We define the composition $g \circ f : A \to C$ to be the function given by the rule

$$x \mapsto g(fx).$$

In other words, if $x \in A$ then $(g \circ f)(x) = g(f(x))$. Schematically:

$$\frac{\begin{array}{c} f : A \to B \\ g : B \to C \\ a \in A \end{array}}{(g \circ f)(a) = g(f(a))} \qquad \frac{\begin{array}{c} f : A \to B \\ g : B \to C \end{array}}{(g \circ f) : A \to C}$$

Composition satisfies the associative law:

$$\frac{\begin{array}{c} f : A \to B \\ g : B \to C \\ h : C \to D \end{array}}{h \circ (g \circ f) = (h \circ g) \circ f}$$

Images and inverse images of sets

Suppose S is a subset of the domain of $f : A \to B$. Then

$$f[S] \stackrel{\text{def}}{=} \{fx \mid x \in S\}$$

so

$$y \in f[S] \iff \exists\, x \in S,\ y = fx$$

and

$$f[S] \subseteq B$$

The set $f[S]$ is called the *image* of $A \subseteq S$. Warning: the term *image* is ambiguous: it can refer to elements $fa \in B$ or subsets $f[S] \subseteq B$. It is usually clear what is meant based on context, but if there is a chance of confusion, we use the phrase "image of the *set* A" to indicate that we mean a subset of the codomain B and not an element of B.

The *image of the function* $f : A \to B$ is the image of the whole domain A. Thus the image of f is $f[A]$.

Suppose S is a subset of the codomain of $f : A \to B$. Then

$$f^{-1}[S] \stackrel{\text{def}}{=} \{x \in A \mid fx \in S\}.$$

This set, called the *inverse image* or *preimage*, is defined even if the inverse function f^{-1} is not defined. We have

$$x \in f^{-1}[S] \iff fx \in S$$

and

$$f^{-1}[S] \subseteq A.$$

Also

$$f^{-1}[B] = A.$$

Identity functions

If A is a set, then the identity function

$$id_A : A \to A$$

is the function defined by the rule $x \mapsto x$. Thus we have the simple law

$$\frac{a \in A}{id_A(a) = a}$$

We also have composition laws:

$$\frac{f : A \to B}{f \circ id_A = f} \qquad \frac{g : B \to A}{id_A \circ g = g}$$

Injective and surjective functions

There are two types of functions that arise often in mathematics: injective functions (also called one-to-one functions) and surjective functions (also called onto functions). These are important in this course, so *students are advised to review these concepts until they have a good understanding of these concepts.*

An *injective function* or *injection* $f : A \to B$ is a function that sends distinct elements of the domain to distinct elements of the codomain. More formally:

$$f : A \to B \text{ is injective } \iff \forall x, y \in A,\ (x \neq y \implies fx \neq fy).$$

This is more commonly expressed in the following equivalent form:

$$f : A \to B \text{ is injective } \iff \forall x, y \in A,\ (fx = fy \implies x = y).$$

So, to prove a function is injective you will often check that the equality $fx = fy$ implies the equality $x = y$ for all x, y in the domain A, and then use the following rule:

$$\frac{\begin{array}{c} f : A \to B \\ \forall x, y \in A,\ (fx = fy \implies x = y) \end{array}}{f : A \to B \text{ is injective}}$$

An *surjective function* or *surjection* $f : A \to B$ is a function whose image is equal to the codomain:

$$f : A \to B \text{ surjective } \iff f[A] = B.$$

In other words,

$$f : A \to B \text{ surjective } \iff \forall\, b \in B,\ \exists\, a \in A,\ \ fa = b.$$

So one common way to check that a function is surjective is to take an arbitrary element b in the codomain B, and show that you can find an element a in the domain A that maps to B.

Another way to show a function is injective or surjective is to find an inverse function. (It is enough to have a left inverse for injective, and a right inverse for surjective. We will discuss inverses in more detail below):

$$\frac{\begin{array}{c} f : A \to B \\ g : B \to A \\ g \circ f = id_A \end{array}}{f \text{ injective}} \qquad \frac{\begin{array}{c} f : A \to B \\ g : B \to A \\ g \circ f = id_A \end{array}}{g \text{ surjective}}$$

Composition of functions behaves well:

$$\frac{\begin{array}{c} f : A \to B \text{ injective} \\ g : B \to C \text{ injective} \end{array}}{g \circ f : A \to C \text{ injective}} \qquad \frac{\begin{array}{c} f : A \to B \text{ surjective} \\ g : B \to C \text{ surjective} \end{array}}{g \circ f : A \to C \text{ surjective}}$$

Bijective functions

A *bijective function* or *bijection* $f : A \to B$ is a function that is both injective and surjective. We have

$$f : A \to B \text{ is bijective } \iff \forall\, b \in B,\ \exists!\, a \in A,\ fa = b$$

and

Identity maps are bijections.

We also have the following laws:

$$\frac{\begin{array}{c} f : A \to B \\ g : B \to A \\ \forall\, a \in A.\ g(fa) = a \\ \forall\, b \in B.\ f(g\,b) = b \end{array}}{f \text{ and } g \text{ bijective}} \qquad \frac{\begin{array}{c} f : A \to B \\ g : B \to A \\ g \circ f = id_A \\ f \circ g = id_B \end{array}}{f \text{ and } g \text{ bijective}} \qquad \frac{\begin{array}{c} f : A \to B \text{ bijective} \\ g : B \to C \text{ bijective} \end{array}}{g \circ f : A \to C \text{ bijective}}$$

Inverse functions

If $f : A \to B$, then an inverse to f is a function $f^{-1} : B \to A$ such that

$$f^{-1} \circ f = id_A \qquad \text{and} \qquad f \circ f^{-1} = id_B.$$

If an inverse exists, it is unique, but not every function has an inverse. In fact

$$f : A \to B \text{ bijective } \iff f \text{ has an inverse.}$$

This gives rise to the following:

$$\frac{f : A \to B \text{ bijective}}{f^{-1} \circ f = id_A} \qquad \frac{f : A \to B \text{ bijective}}{f \circ f^{-1} = id_B}$$

$$\frac{f : A \to B \text{ bijective}}{\forall\, a \in A.\ f^{-1}(f\,a) = a} \qquad \frac{f : A \to B \text{ bijective}}{\forall\, b \in B.\ f(f^{-1}\,b) = b}$$

We also have the following:

$$\frac{\begin{array}{c} f : A \to B \\ g : B \to A \\ \forall\, a \in A.\ g(f(a)) = a \\ \forall\, b \in B.\ f(g(b)) = b \end{array}}{f = g^{-1} \text{ and } g = f^{-1}} \qquad \frac{\begin{array}{c} f : A \to B \\ g : B \to A \\ g \circ f = id_A \\ f \circ g = id_B \end{array}}{f = g^{-1} \text{ and } g = f^{-1}}$$

Finally, we have the following:

$$\frac{\begin{array}{c} f : A \to B \text{ bijective} \\ g : B \to C \text{ bijective} \end{array}}{g \circ f : A \to C \text{ bijective}}$$

Restrictions of functions

Suppose that $f\colon A \to B$ is a function and $C \subseteq A$. Then we can *restrict* f to C. This results in a function with domain C. The restriction is written $f|_C\colon C \to B$.

The restriction is defined by the rule $f|_C(c) \stackrel{\text{def}}{=} f(c)$ for all $c \in C$. In other words, if $f\colon A \to B$ is defined by a certain rule $a \mapsto \varphi(a)$, then $f|_C\colon C \to B$ is defined by *the same rule*: $c \mapsto \varphi(c)$. The only difference is that this variable c has values only in the subset C.

The restriction of an injective function remains injective, but it is not necessarily true that the restriction of a surjective function remains a surjective.

There is another concept that is used from time to time in mathematics: *restriction of codomain*. Suppose that $f\colon A \to B$ and that $D \subseteq B$. If D is large enough to contain the image $f[A]$ (so that $f[A] \subseteq D$) then we can form a function $f'\colon A \to D$ defined by the same rule as f. In other words $fa = f'a$ for all $a \in A$. The only difference between f and f' is the codomain. (There is no standard notation for the restriction of codomain: we used a here prime). If $f[A]$ is not a subset of D, then the restriction of codomain is not allowed: it results in a function that is not well-defined.

Inclusions Functions

If $A \subseteq B$ then there is an inclusion function $\iota\colon A \to B$ that behaves similarly to an identity function (except the domain is not necessarily equal to the codomain). So $\iota(a) = a$ for all $a \in A$. The function has the effect of *including* A into B. The inclusion function is injective, but not surjective (unless $A = B$).

A.10 Binary relations and equivalence relations

A *binary relation* on a set A is a subset R of the Cartesian product $A \times A$. In other words, R is a set of ordered pairs with first and second coordinate in A.

If $(x, y) \in R$ where R is a relation on A, then we say that x and y are related by R. We often write this as xRy using *infix* notation. If $(x, y) \notin R$ the we say that x and y are not related by R, and write $\neg(xRy)$ or $x \not R\, y$.

We can think of $=$ as giving a binary relation on the set A by defining the relation R to be the set $\{(x, x) \mid x \in A\}$. This set is sometimes called the *diagonal* or the *graph of the identity function*.

Types of Binary Relations

We review the concept of reflexive, symmetric, and transitive relations.

A binary relation is *reflexive* if xRx for all $x \in R$.

A binary relation is *symmetric* means that for all $x, y \in A$ with xRy we have yRx.

A binary relation is *transitive* means that for all $x, y, z \in A$ with xRy and yRz, we have xRz.

Equivalence relations

A binary relation that is (i) reflexive, (ii) symmetric, and (iii) transitive is called an *equivalence relation* on R. For example, $=$ is an equivalence relation.

For the rest of this section we assume that R is an equivalence relation on A. We write xRy as $x \sim y$. In other words, since we have an equivalence relation we have $x \sim x$ for all $x \in A$, $x \sim y \implies y \sim x$ for all $x, y \in A$, and $(x \sim y) \wedge (y \sim z) \implies x \sim z$ for all $x, y, z \in A$.

We define the *equivalence class* $[x]$ to be the following set

$$[x] \stackrel{\text{def}}{=} \{y \in A \mid x \sim y\}$$

The only two equivalence classes can intersect is for them to be equal. In other words, if $[x] \cap [y]$ is not empty, then $[x] = [y]$. By the reflexive law $x \in [x]$, so every element is in an equivalence class, and each equivalence class has at least one element. Thus equivalence classes *partitions* all of A into disjoint, nonempty subsets.

We will need the following (related) laws:

$$[x] = [y] \iff x \sim y$$

$$x \in [x]$$

$$x \in [y] \iff y \in [x]$$

$$x \in [y] \iff [x] = [y]$$

Exercise 1. Prove that if $[x] \cap [y]$ is not empty, then $x \sim y$. Prove it using the definition of equivalence class, plus the symmetric and transitive laws.

Appendix B

Exploring $\mathbb{C}$

In this appendix we consider further properties of $\mathbb{C}$ which every mathematician should know. The reason why this material is in an appendix, and not the main text, is that it relies on basic trigonometry as well as the (real) exponential function. This is a departure from the main text which is a self-contained axiomatic development of the number systems.

Although we certainly could define and develop these important functions using our usual rigorous methodology, this would take us too far afield. So we take a few facts from precalculus as given. Consequently, for this appendix all the concepts and results can be thought of as "informal" since we draw on basic mathematical knowledge developed outside this course.

In this appendix we will consider the polar form of complex numbers and De Moivre's law. We will also consider the complex exponential function. We will establish that every nonzero element $z \in \mathbb{C}$ has n distinct nth roots for every positive integer n. In other words, if $z \neq 0$ then the polynomial $X^n - z$ has n roots in $\mathbb{C}$. In the next appendix, we will consider the important *fundamental theorem of algebra* concerning complex roots of more general polynomials.

B.1 Review of trigonometric and the real exponential functions

In order to study the complex numbers in more depth, we need some basic trigonometry and some basic facts about the real exponential function e^x. Actually the only trig functions we will need are the sine and cosine functions, $\sin x$ and $\cos x$. We will use radians, so will need to assume we have the positive real number π.

We will list here a few basic facts that will be used in this appendix. We take the following propositions as given, and assume the reader is familiar with them (but not necessarily their rigorous proofs).

Proposition 1 (Sine and cosine)**.** *The sine and cosine functions are both continuous functions* $\mathbb{R} \to [-1, 1]$*. The number* π *is a positive real number such that the sine and cosine functions are periodic with period* 2π *in the sense that*

$$\sin(x + 2\pi k) = \sin x, \qquad \cos(x + 2\pi k) = \cos x$$

for all $x \in \mathbb{R}$ *and* $k \in \mathbb{Z}$*. Some special values of the sine function are as follows:* $\sin(0) = 0$, $\sin(\pi/2) = 1$*, and* $\sin \pi = \sin 2\pi = 0$*. Some special values of the cosine function are as follows:* $\cos(0) = 1$, $\cos(\pi/2) = 0$*, and* $\cos \pi = -1$*. In fact,* $\cos x = 1$ *if and only if* $x = 2k\pi$ *for some* $k \in \mathbb{Z}$*.*

The sine function is odd and the cosine function is even in the sense that

$$\sin(-x) = -\sin x, \qquad \cos(-x) = \cos x$$

for all $x \in \mathbb{R}$*.*

Remark 1. Sometimes we will use Greek letters such as $\alpha, \beta, \theta \in \mathbb{R}$ for our inputs, where we think the inputs as angles in radians. So we will sometimes write expressions such as $\sin \theta$ or $\cos \alpha$. We adopt the usual convention that $\sin^2 \theta$ is $(\sin \theta)^2$ and $\cos^2 \theta$ is $(\cos \theta)^2$. So here these expressions do *not* refer to iteration.

Proposition 2 (Trig identities)**.** *For all* $\theta, \alpha, \beta \in \mathbb{R}$*, we have*

$$\sin^2 \theta + \cos^2 \theta = 1,$$

$$\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta,$$

$$\sin(\alpha + \beta) = \cos \alpha \sin \beta + \cos \beta \sin \alpha.$$

Proposition 3 (The real exponential function)**.** *The real exponential function* $\exp x$*, also written* e^x*, is a continuous function* $\mathbb{R} \to (0, \infty)$*. The exponential function has special value* $\exp 0 = 1$*. Also*

$$e^{x+y} = e^x e^y$$

for all $x, y \in \mathbb{R}$*.*

Proposition 4 (Polar coordinates)**.** *Let* $(x, y) \in \mathbb{R}^2$ *be a point on the real plane. Then there are real numbers* r *and* θ *such that*

$$(x, y) = (r \cos \theta, r \sin \theta).$$

Furthermore, if we require that r *be nonnegative, then* r *is unique. If we restrict* θ *in the range* $0 \leq \theta < 2\pi$*, or some other half-open interval of length* 2π*, and if* $(x, y) \neq (0, 0)$*, then* θ *is also unique.*

Remark 2. We call the r in the above proposition the *radius of* (x, y). When it is unique, we call the θ in the above proposition the *angle of* (x, y). We call (r, θ) *polar coordinates* of the point (x, y).

B.2 Polar form of complex numbers

From Proposition 4 it easily follows that every complex number z can be expressed as $r\cos\theta + r\sin\theta \cdot i$ where r and θ are real numbers, and where r is nonnegative. We call this a *polar form* of z.

Theorem 5 (Polar form of a complex number)**.** *Let $z \in \mathbb{C}$. Then there are numbers $r, \theta \in \mathbb{R}$ with $r \geq 0$ such that*

$$z = r\cos\theta + r\sin\theta \cdot i.$$

Proof. Write z as $x + yi$ where $x, y \in \mathbb{R}$. By Proposition 4, there is a $\theta \in \mathbb{R}$ and a nonnegative $r \in \mathbb{R}$ such that $x = r\cos\theta$ and $y = r\sin\theta$. So

$$z = x + yi = r\cos\theta + r\sin\theta \cdot i.$$

□

Remark 3. To help with readability, the polar form is sometimes written with the i before $r\sin\theta$:

$$z = x + yi = r\cos\theta + i\,r\sin\theta.$$

The radius of a complex number is just the absolute value:

Theorem 6 (Radius formula)**.** *If $z = r\cos\theta + ir\sin\theta$ then $r = |z|$.*

Proof. Observe

$$|z|^2 = (r\cos\theta)^2 + (r\sin\theta)^2 = r^2(\cos^2\theta + \sin^2\theta) = r^2.$$

The first equality uses the formula for absolute value of complex numbers from Chapter 12. The last equality uses Proposition 2. □

Informal Exercise 1. Let $z = r\cos\theta + ir\sin\theta$ be a complex number written in polar form. Show that $\overline{z} = r\cos(-\theta) + ir\sin(-\theta)$.

The above gives a nice description of complex conjugation in polar form. There is also a nice description of multiplication for complex numbers in polar form:

Theorem 7 (Product formula)**.** *Let $z, w \in \mathbb{C}$. If z, w are written in polar coordinates as*

$$z = r_1 \cos\theta_1 + r_1 \sin\theta_1 \cdot i \qquad w = r_2 \cos\theta_2 + r_2 \sin\theta_2 \cdot i$$

then the product can be written as

$$zw = r\cos\theta + r\sin\theta \cdot i$$

where $r = r_1 r_2$ and $\theta = \theta_1 + \theta_2$.

Remark 4. In other words, when you multiply complex numbers, you multiply the radii and add the angles. So multiplication has a very nice geometric interpretation.

Addition also has a geometric interpretation. It is just vector addition.

Proof. For convenience, write $\cos\theta_1$ as c_1, $\cos\theta_2$ as c_2, $\sin\theta_1$ as s_1, and $\sin\theta_2$ as s_2. So

$$\begin{aligned}
zw &= (r_1c_1 + r_1s_1i)(r_2c_2 + r_2s_2i) \\
&= r_1(c_1 + s_1i)r_2(c_2 + s_2i) && \text{(Distr. Law)} \\
&= r_1r_2(c_1 + s_1i)(c_2 + s_2i) && \text{(Comm./Assoc. Laws)} \\
&= r_1r_2(c_1c_2 + c_1s_2i + s_1ic_2 + s_1is_2i) && \text{(Distr. Law)} \\
&= r_1r_2\big(c_1c_2 + c_1s_2i + s_1c_2i - s_1s_2\big) && (i^2 = -1) \\
&= r_1r_2\big((c_1c_2 - s_1s_2) + (c_1s_2 + s_1c_2)i\big) \\
&= r_1r_2\big(\cos(\theta_1 + \theta_2) + \sin(\theta_1 + \theta_2)i\big). && \text{(Proposition 2)}
\end{aligned}$$

□

Corollary 8 (Inverse formula)**.** *Let $z \in \mathbb{C}$. If $z \neq 0$ is*

$$z = r\cos\theta + r\sin\theta \cdot i$$

in polar coordinates, then

$$z^{-1} = r^{-1}\cos(-\theta) + r^{-1}\sin(-\theta)i$$

and

$$z^{-1} = r^{-1}\cos\theta - r^{-1}\sin\theta \cdot i.$$

Informal Exercise 2. Prove the above corollary.

B.3 De Moivre's theorem

In Chapter 6 we considered exponentiation in general rings and fields. So if $z \neq 0$ in $\mathbb{C}$ it has an power z^n for all $n \in \mathbb{Z}$. De Moivre's theorem gives a interesting polar form formula for these powers.

Theorem 9 (De Moivre's theorem)**.** *Let $z \in \mathbb{C}$ and $n \in \mathbb{Z}$. Suppose $z \neq 0$. If z is*

$$z = r\cos\theta + r\sin\theta \cdot i$$

in polar form, then

$$z^n = r^n\cos(n\theta) + r^n\sin(n\theta)i.$$

Proof. Fix $z \neq 0$. First we prove

$$z^n = r^n\cos(n\theta) + r^n\sin(n\theta)i.$$

for all nonnegative n using induction.

Observe the result holds for $n = 0$ since $z^0 = 1$ (exponentiation in rings) and

$$r^0\cos(0) + r^0\sin(0)i = 1 + 0i = 1$$

since $r^0 = 1$, $\cos 0 = 1$ and $\sin 0 = 0$.

Now suppose that the result holds for the natural number $n = u$. Then

$$z^{u+1} = z^u z = \big(r^u\cos(u\theta) + r^u\sin(u\theta)i\big)\big(r\cos(\theta) + r\sin(\theta)i\big)$$

By Theorem 7,

$$z^u z = r'\cos\theta' + r'\sin\theta' \cdot i$$

where $r' = r^u r = r^{u+1}$ and $\theta' = u\theta + \theta = (u+1)\theta$. Thus the result holds for $n = u + 1$.

By induction, the theorem holds for all $n \geq 0$. For $n < 0$, let $m = -n$. So the theorem holds for z^m. Now use Corollary 8 to show that $z^n = (z^m)^{-1}$ has radius $(r^m)^{-1}$ and angle $-(m\theta)$. Since $(r^m)^{-1} = r^n$ and $-(m\theta) = n\theta$, the result follows. □

Remark 5. This shows that, in polar coordinates, when one takes the nth power one takes the nth power of the radius and multiplies the angle by n.

B.4 The complex exponential function

We now consider the complex exponential function $z \mapsto \exp z$ as a function $\mathbb{C} \to \mathbb{C}$.

Definition 1 (Complex exponential function)**.** Let $z \in \mathbb{C}$. If $z = x + yi$ where $x, y \in \mathbb{R}$ then

$$\exp z \stackrel{\text{def}}{=} \exp(x)\,(\cos y + i \sin y).$$

Here $\exp(x)$ is the value of the real exponential function, which we take as given.

Remark 6. We often write $\exp z$ as e^z. The following results will help justify this notation.

Theorem 10. *When we restrict the complex exponential function to $\mathbb{R}$ then the values agree with the real exponential function. Thus the complex exponential function is an extension of the real exponential function to the larger domain $\mathbb{C}$.*

Proof. Observe that

$$\exp(x + 0i) = \exp(x)\,(\cos 0 + \sin 0 \cdot i) = \exp(x) \cdot (1 + 0 \cdot i) = \exp x.$$

□

Corollary 11. *If z is real, then the value e^z of the complex exponential function is real. Furthermore*

$$e^0 = 1.$$

Remark 7. If $y \in \mathbb{R}$,

$$e^{iy} = \cos y + i \sin y.$$

To see this, take $x = 0$ and observe $\exp x = e^0 = 1$. This formula for e^{iy} implies

$$e^z = e^x e^{iy}$$

as expected.

Informal Exercise 3. Let $z = x + yi$ where $x, y \in \mathbb{R}$. Show that $|e^z| = e^x$. Conclude then that $e^z \neq 0$.

Theorem 12. *Let $z \in \mathbb{C}$. If z is*

$$z = r\cos\theta + ir\sin\theta$$

in polar form, then

$$z = re^{\theta i}.$$

Proof. By definition, $e^{i\theta} = \cos\theta + i\sin\theta$. See Remark 7. □

Corollary 13. *Every complex number can be written in the form*

$$z = re^{\theta i}$$

where $r = |z|$.

Informal Exercise 4. Suppose z, w are complex numbers. Then show that

$$e^{z+w} = e^z e^w.$$

Hint: Use Theorem 7, and known properties of e^x when x is real.

Remark 8. A special case of this is when $w = -z$. Since $e^0 = 1$, we conclude that $e^z e^{-z} = 1$. In other words, e^{-z} is the multiplicative inverse of e^z.

Theorem 14. *Suppose z is a complex number, and n is an integer. Then*

$$(e^z)^n = e^{nz}.$$

Proof. This follows by induction using Informal Exercise 4. The case of negative n has to be established separately using the above remark. □

Informal Exercise 5. Show that $e^{\pi i} = -1$. This is Euler's formula, and is considered by many to be one of the most amazing formulas in mathematics since it puts e, π, i into one simple formula. Observe that $e^{\pi i} + 1 = 0$ combines $e, \pi, i, 0, 1$, and involves addition, multiplication, and exponentiation.

B.5 Nth roots of complex numbers

In Chapter 11 we established that for every positive integer n and for every nonnegative x in $\mathbb{R}$ there exists a nth root of x in $\mathbb{R}$. When n is odd, then we can extend this result for negative x as well. However, we cannot hope to have nth roots in $\mathbb{R}$ when n is even and $x < 0$.

In this section we establish that nth roots exist for any positive integer n and any $z \in \mathbb{C}$. In fact, if $z \neq 0$ there are exactly n distinct nth roots evenly distributed on a circle in the complex plane.

First we establish existence of nth roots:

Theorem 15 (Existence of roots)**.** *Suppose z is a complex number written in the form $z = re^{\theta i}$. Then $r^{1/n} e^{\theta i/n}$ is an nth root of z.*

Proof. This follows from Theorem 14 and other properties of exponentiation. □

Informal Exercise 6. Sketch the cube root of -1 in the complex plane. Use the formula from Theorem 15.

In order to classify nth roots of a complex number $z \neq 0$, it is convenient to start with $z = 1$.

Definition 2 (Roots of unity)**.** Let n be a fixed positive integer. Every complex number of the form $e^{2k\pi i/n}$ with $k \in \mathbb{Z}$ is called an *nth root of unity*.

Theorem 16. *If z is an nth root of unity then $z^n = 1$. Furthermore, there are n distinct nth roots of unity.*

Proof. Observe that $e^{2\pi i} = 1$. Thus

$$\left(e^{2k\pi i/n}\right)^n = e^{2k\pi i} = \left(e^{2\pi i}\right)^k = 1^k = 1.$$

Thus every nth root of unity is an nth root of 1.

If $0 \leq k_1 < k_2 < n$ then the angles of the corresponding roots of unity satisfy $0 \leq 2k_1\pi/n < 2k_2\pi/n < 2\pi$. By uniqueness of angle (Proposition 4) in the interval $[0, 2\pi)$, $e^{2k_1\pi i/n}$ and $e^{2k_2\pi i/n}$ are distinct. Thus there are at least n distinct nth roots of unity.

For general $k \in \mathbb{Z}$ the remainder upon dividing by n gives rise to the same root of unity. This can be seen as follows. Let $k = qn + r$ with $0 \leq r < n$. Then

$$e^{2k\pi i/n} = e^{2(qn+r)\pi i/n} = \left(e^{2\pi i}\right)^q e^{2r\pi i/n} = 1^q\, e^{2r\pi i/n}.$$

So there are no more nth roots of unity beyond the n discussed above. □

The converse is true: if z satisfies $z^n = 1$ then it is an nth root of unity:

Theorem 17. *Let n be a positive integer. Suppose $z \in \mathbb{C}$ is such that $z^n = 1$. Then z is an nth root of unity.*

Proof. Observe that $1 = |z^n| = |z|^n$. By uniqueness of nonnegative nth roots in $\mathbb{R}$, this implies that $|z| = 1$. Write z as $re^{i\theta}$ where r is the radius and $\theta \in \mathbb{R}$. Observe that $r = 1$ since $|z| = 1$. Also

$$1 = z^n = \left(e^{i\theta}\right)^n = e^{in\theta} = \cos(n\theta) + i\sin(n\theta).$$

This implies $\cos(n\theta) = 1$. This means that $n\theta = 2\pi k$ for some $k \in \mathbb{Z}$. The result follows. □

Corollary 18. *Let n be a positive integer. There are exactly n solutions to the equation $z^n = 1$ in $\mathbb{C}$. These are the nth roots of unity.*

Theorem 19. *Suppose z is a complex number written in the form $z = re^{\theta i}$. Let n be a positive integer, and let ζ be an nth root of unity. Then $r^{1/n}e^{\theta i/n}\zeta$ is an nth root of z. If $z \neq 0$ then every nth root of z can be written in this way.*

Proof. Observe that

$$\begin{aligned}\left(r^{1/n}e^{\theta i/n}\zeta\right)^n &= \left(r^{1/n}\right)^n \left(e^{\theta i/n}\right)^n \zeta^n \\ &= r \cdot e^{\theta i} \cdot 1 \\ &= z.\end{aligned}$$

Thus $r^{1/n}e^{\theta i/n}\zeta$ is an nth root of z.

This gives one root, now suppose w is any nth root of z where $z \neq 0$. Then $w \neq 0$. Observe that if $u = r^{1/n}e^{\theta i/n}w^{-1}$ then $u^n = 1$. By Theorem 17, we see that u it is an nth root of unity. Hence $\zeta' \stackrel{\text{def}}{=} u^{-1}$ is also a root of unity, and $w = r^{1/n}e^{\theta i/n}\zeta'$. □

Theorem 20 (Nth roots)**.** *Let n be a positive integer. Then every complex number $z \neq 0$ has exactly n distinct nth roots.*

Proof. There are at most n distinct nth roots of z. This follows from the previous theorem and the fact that there are only n distinct nth roots of unity.

Suppose ζ and ζ' are roots of unity such that

$$r^{1/n}e^{\theta i/n}\zeta = r^{1/n}e^{\theta i/n}\zeta'.$$

By cancelling, we get $\zeta = \zeta'$. So distinct nth roots of unity give distinct nth roots of z. So z has exactly n such roots. □

Remark 9. Of course 0 has 0 for an nth root. In fact, 0 is the only nth root of 0.

Informal Exercise 7. Find the general formula for the three cube roots of -1. The formula should be in terms of the polar form of these roots. Use polar coordinates to plot these three roots in the complex plane.

Informal Exercise 8. Describe and plot the four 4th roots of $1+i$. Describe and plot the five fifth roots of unity. Describe and plot the 6 distinct 6th roots of 2^6. (Hint: use the same approach as the previous exercise).

Appendix C

Polynomials

This appendix contains an informal survey of some key results concerning polynomials. These results illustrate some very important properties of various number systems. For example, one reason the complex numbers are so important in mathematics is that every polynomial with coefficients in $\mathbb{C}$ has a full set of roots.

Only some of the results in this appendix are proved. With or without proof they are included due to their importance in mathematics. They are important results to know, even if the proofs may have to wait for future courses.

C.1 Polynomial rings

Polynomials can be constructed in a rigorous manner in the style of our other constructions of the number systems, and the operations of addition and multiplication can be defined rigorously. However, to do so here would take us to far afield. So we will appeal to common (precalculus level) experience in our approach to polynomials.

Definition 1 (Set of polynomials)**.** Let R be a commutative ring, and x a variable. Then $R[x]$ is the set of polynomials $a_n x^n + \ldots + a_1 x + a_0$ with coefficients $a_i \in R$.

Remark 1. We adopt the usual conventions for identity, negation and subtraction used in algebra. So $x^2 - 2x - 1$ is short for $1x^2 + (-2)x + (-1)$, and $-x^3 + x$ is short for $(-1)x^3 + 0x^2 + 1x + 0$.

In the above x can be replaced by any given variable. The variable must be a "symbolic" variable. That is, it must be a variable not currently being

used to represent a fixed value. So if y is not being used to represent a fixed value, we can define $\mathbb{Z}[y]$, say, to be the set of polynomials with variable y and coefficient in $\mathbb{Z}$. This set would contain $3y^2-2$, but would not contain $3x^2-2$ or $(1/2)y^2$.

Example. Observe that $7x^3 - 3x^2 + 11$ is in $\mathbb{Z}[x]$. It is also in $\mathbb{Q}[x]$, in $\mathbb{R}[x]$, and in $\mathbb{C}[x]$ since $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. Observe that $\frac{7}{11}x^3 - 3x^2 + 11$ is in $\mathbb{Q}[x]$ but not in $\mathbb{Z}[x]$. Observe that $7T^3 - \sqrt{2}T^2 + T - 11$ is in $\mathbb{R}[T]$ but not in $\mathbb{Q}[T]$. Observe that $Z - i$ is in $\mathbb{C}[Z]$ but not in $\mathbb{C}[S]$.

If $a_n x^n + \ldots + a_1 x + a_0$ is a polynomial with coefficients a_i, we adopt the convention that $a_i = 0$ for all values of i not occurring in the expression $a_n x^n + \ldots + a_1 x + a_0$. For example, when writing $7x^3 + x - 11$ in the form $a_n x^n + \ldots + a_1 x + a_0$, we consider $a_2 = 0$ and $a_4 = 0$, but $a_3 = 7$, $a_1 = 1$ and $a_0 = -11$. Two polynomials $a_n x^n + \ldots + a_1 x + a_0$ and $b_k x^k + \ldots + b_1 x + b_0$ are defined to be equal if and only if $a_i = b_i$ for all $i \geq 0$.

Example. Observe that $\overline{6}x^3 + \overline{2}x^2 - x + \overline{1} = -x^2 + \overline{2}x + \overline{1}$ in $\mathbb{F}_3[x]$.

Among the polynomials in $R[x]$ are the *constant* polynomials a_0. In other words, $a_0 \in R$ can be thought of as both an element of R and as a constant polynomial in $R[x]$. Thus $R \subseteq R[x]$. (More formally, we define an injective canonical embedding $R \to R[x]$ which maps c to the constant polynomial c.)

Polynomials are added and multiplied in the usual way. For example, in $\mathbb{Z}_6[x]$ the product of $\overline{2}x^2 + \overline{3}x + \overline{1}$ with $\overline{3}x^2 + \overline{2}$ can be computed as follows

$$(\overline{2}x^2 + \overline{3}x + \overline{1})(\overline{3}x^2 + \overline{2}) = \overline{6}x^4 + \overline{4}x^2 + \overline{9}x^3 + \overline{6}x + \overline{3}x^2 + \overline{2} = \overline{3}x^3 + x^2 + \overline{2}.$$

Exercise 1. Multiply $\overline{2}x^2 + \overline{3}x + \overline{1}$ by $\overline{3}x^2 + x - \overline{2}$ in $\mathbb{F}_5[x]$.

The set $R[x]$ is closed under addition and multiplication. So $+$ and $\times$ give two binary operations $R[x] \times R[x] \to R[x]$. It turns out (but we skip the proofs), that these operations satisfy the expected associative, commutative, distributive, identity and inverse laws. More precisely, the following holds:

Theorem 1. *If R is a commutative ring, then $R[x]$ is also a commutative ring. The additive identity is the constant 0 polynomial, and the multiplicative identity is the constant 1 polynomial.*

C.2 Substitutions

Definition 2 (Substitution)**.** If $f \in R[x]$ then $f(a)$ denotes what we get when we substitute a for x in f. It is defined whenever the substitution makes sense (typically when a is in R, or when a is in a ring containing R).

Example. If $f = x^2 + \overline{1}$ in $\mathbb{Z}_8[x]$ then $f(\overline{3}) = \overline{2}$.

Example. If $f = x^3$ in $\mathbb{Z}_{12}[x]$ then $f(x + \overline{2}) = (x + \overline{2})^3 = x^3 + \overline{6}x^2 + \overline{8}$. (Did you see what happened to the linear term?).

Example. If $f \in R[x]$, and y is another variable, then $f(y)$ is in $R[y]$ and has the same coefficients. However, if x and y are different variables, then $f(x)$ is not considered to be equal to $f(y)$ unless f is a constant polynomial.

Example. Let $f \in R[x]$. Observe that $f(x)$ is just f itself since when we replace x with x we get what we started with. So $f(x)$ is another way of writing f. So we can write f as $f(x)$ when we want to emphasize that f is a polynomial in x.

Example. Here is an amusing example. Suppose $f = x^3 - x \in \mathbb{Z}_3[x]$. Then $f(\overline{0}) = \overline{0}$, $f(\overline{1}) = \overline{0}$, and $f(\overline{2}) = \overline{0}$. So $f(a) = \overline{0}$ for all $a \in \mathbb{Z}_3$ but $f \neq \overline{0}$. So polynomials cannot be treated as functions when R is finite: two distinct polynomials, for example f and $\overline{0}$ as above, can have identical values. (This shows that for finite fields, polynomials are not exactly the same thing as functions. The only function $\mathbb{Z}_3 \to \mathbb{Z}_3$ whose values are all zero is the zero function. In fact two functions are said to be equal if and only if they have the same values. In contrast, we have found two distinct polynomials whose values are zero.)

Definition 3 (Root of a polynomial)**.** Let $f \in R[x]$ and $a \in R$. If $f(a) = 0$ then a is called a *root* of $f \in R[x]$.

The above example (preceding the definition) shows that every element of $\mathbb{F}_3$ is a root of $x^3 - x \in \mathbb{F}_3[x]$.

Exercise 2. Find the roots of $x^3 - \overline{1}$ in $\mathbb{F}_7$. Find the roots of $x^3 - \overline{1}$ in $\mathbb{F}_5$.

C.3 The quotient-remainder theorem for polynomials

Let F be a field. The ring of polynomials $F[x]$ has a quotient-remainder theorem. To state this theorem we need to discuss a notion of size for $F[x]$, called the *degree*:

Definition 4 (Degree)**.** Let $f \in R[x]$ where R is a commutative ring. If f has the form $a_n x^n + \ldots + a_1 x + a_0$ with $a_n \neq 0$ then the *degree* of f is defined to be n and the *leading coefficient* is defined to be a_n.

If $f = 0$ then the degree of f is said to be *undefined* (some authors give it degree $-\infty$).

Be careful when using this definition in modular arithmetic. For example, the polynomial $6x^3+2x^2-x+1$ in $\mathbb{F}_3[x]$ has only degree 2, and $6x^3+2x^2-x+1$ in $\mathbb{F}_2[x]$ has degree 1. However, $6x^3 + 2x^2 - x + 1$ in $\mathbb{F}_5[x]$ has degree 3

You would hope that the degree of fg would be the sum of the degrees of f and g individually. However, examples such as

$$(2x^2 + 3x + 1)(3x^2 + 2) = 3x^3 + x^2 + 2.$$

in $\mathbb{Z}_6[x]$ spoil our optimism. However, if the coefficients are in a field F then it works.

Theorem 2 (Additivity of degree)**.** *If $f, g \in F[x]$ are non-zero polynomials where F is a field, then*

$$\deg(fg) = \deg f + \deg g.$$

Informal Exercise 3. Justify the above theorem. Explain why the proof does not work if the coefficients are in $\mathbb{Z}_m$ where m is composite. Hint: focus on the leading coefficients.

As mentioned above, the degree of a polynomial is a measure of size. When we divide we want the size of the remainder to be smaller than the size of the quotient. This leads to the following:

Theorem 3 (Quotient-remainder theorem)**.** *Let $f, g \in F[x]$ be polynomials where F is a field. Assume g is not zero. Then there are unique polynomials $q(x)$ and $r(x)$ such that (i) $f(x) = q(x)g(x) + r(x)$, and (ii) the polynomial $r(x)$ is either the zero polynomial or has degree strictly smaller than $g(x)$.*

Remark 2. The polynomial $q(x)$ in the above is called the *quotient* and the polynomial $r(x)$ is called the *remainder*.

Remark 3. This theorem extends to polynomials in $R[x]$ where R is a commutative ring that is not a field, as long as we add the extra assumption that the leading coefficient of g is a unit in R.

Remark 4. This theorem can be used a basis to prove theorems about GCDs and unique factorization in $F[x]$.

As an important special case of the above theorem, consider $g(x) = x - a$ where $a \in R$. Then the remainder $r(x)$ must be zero, or have degree zero. So $r = r(x)$ is a constant polynomial. What is this constant? To find out, write $f(x) = q(x)(x - a) + r$. When we substitute $x = a$ we get

$$f(a) = q(a)(a - a) + r = 0 + r = r.$$

In other words, $r = f(a)$. This gives the following:

Corollary 4. *Let $a \in F$ where F is a field, and let $f \in F[x]$. Then there is a unique polynomial $q \in F[x]$ such that*

$$f(x) = (x - a)q(x) + f(a).$$

Remark 5. This actually works for commutative rings as well as for fields F since the leading coefficient of $g(x) = x - a$ is 1 which is always a unit.

The following is a special case of the above corollary (where $f(a) = 0$).

Corollary 5. *Let $a \in F$ where F is a field, and let $f \in F[x]$. Then a is a root of f if and only if $(x - a)$ divides f.*

C.4 The number of roots

Theorem 6. *Let $f \in F[x]$ be a nonzero polynomial with coefficients in a field F. Then f has at most $n = \deg f$ roots in F.*

Proof. This is proved by induction. Let S be the set of natural numbers n such that every polynomial f that has degree n has at most n roots in F. Our goal is to show that $S = \mathbb{N}$.

Showing $0 \in S$ is easy. If f is a non-zero constant polynomial of degree 0, then it has 0 roots since it is a nonzero constant polynomial.

Suppose that $k \in S$. We want to show $k+1 \in S$. To do so, let f be a polynomial of degree $k+1$. If f has no roots, then the statement is trivially true. Suppose that f does have a root $a \in F$. Then, by Corollary 5,

$$f(x) = q(x)(x - a).$$

By Theorem 2, $\deg f = 1 + \deg q$. In other words, $\deg q = k$. By the inductive hypothesis $k \in S$, the polynomial q has at most k roots.

We will now show that the only possible root of f that is not a root of q is a (but a could also be a root of q). Suppose that f has a root $b \neq a$. Then $0 = f(b) = q(b)(b-a)$. Since $b - a \neq 0$, we can multiply both sides by the inverse: $0(b-a)^{-1} = q(b)(b-a)(b-a)^{-1}$. Thus $0 = q(b)$. So every root of f not equal to a must be a root of $q(x)$. Since $q(x)$ has at most k roots, it follows that $f(x)$ must have at most $k+1$ roots. So $k+1 \in S$.

By the principle of mathematical induction, $\mathbb{N} = S$. The result follows. □

Remark 6. Observe how this can fail if F is replaced by the ring $\mathbb{Z}_m$ where m is not a prime. The polynomial $x^2 - 1 \in \mathbb{Z}_8[x]$ has degree 2, yet it has four roots! (Can you find them?)

Exercise 4. Find all four roots of $x^2 - 1 \in \mathbb{Z}_8[x]$ in $\mathbb{Z}_8$.

Exercise 5. Show that if $f, g \in F[x]$ are non-zero polynomials where F is a field, then the set of roots of fg is the union of the set of roots of f with the set of roots of g.

Exercise 6. Show that the result of the above exercise does not hold in $\mathbb{Z}_8[x]$ by looking at a factorization of $x^2 - \overline{1}$.

Exercise 7. Although the result of Exercise 5 does not hold if F is replaced by a commutative ring with zero divisors (such as $\mathbb{Z}_m$ where m is composite), one of the two inclusions does hold. Which one and why?

C.5 Irreducible polynomials

One can prove unique factorization into irreducible polynomials for $F[x]$. A polynomial $f \in F[x]$ is said to be *irreducible* if it is not a constant and if it

has no divisors g with $0 < \deg g < \deg f$. These polynomials play the role of prime numbers in polynomial rings. One can use the methods of Chapter 5 to prove that every nonconstant polynomial is the product of a constant times one or more irreducible polynomials.

Finally, even if F is finite, one can prove that there are an infinite number of irreducible polynomials in $F[x]$ using a similar argument to that used in showing that there are an infinite number of primes.

Exercise 8. Show that every linear polynomial is irreducible. (We will see that in $\mathbb{C}$, these are the only irreducible polynomials).

Exercise 9. Show that a quadratic polynomial $f \in F[x]$ with no roots in F must be irreducible. Show that, because of this, $x^2+\overline{1}$ is irreducible in $\mathbb{F}_3[x]$.

C.6 Fundamental theorem of algebra

One of the great advantages of using the field $\mathbb{C}$ is that every nonconstant polynomial has a root. This is called the *fundamental theorem of algebra*.

Theorem 7 (Fundamental theorem of algebra, part 1)**.** *Every nonconstant polynomial in $\mathbb{C}[X]$ has a root in $\mathbb{C}$.*

Corollary 8. *Every non-constant polynomial with real or complex coefficients has a root in $\mathbb{C}$.*

Corollary 9 (Fundamental theorem of algebra, part 2)**.** *Every non-constant polynomial in $\mathbb{C}[x]$ is the product of linear polynomials in $\mathbb{C}[x]$.*

For real roots we get the following weaker results (which can be proved using the intermediate value theorem):

Theorem 10. *Every polynomial of odd degree in $\mathbb{R}[x]$ has a root in $\mathbb{R}$.*

Real polynomials do not always factor into linear real polynomials. The following weaker result is true:

Theorem 11. *Every non-constant polynomial in $\mathbb{R}[x]$ factors into a product of linear and irreducible quadratic polynomials in $\mathbb{R}[x]$*

In other words, we have to allow for the possibility of quadratic factors that have no real roots. The irreducible polynomials of $\mathbb{R}[x]$ are the linear polynomials and the quadratic polynomials with no real roots.[1] Contrast this with $\mathbb{C}[x]$ where the irreducible polynomials are just the linear polynomials.

[1]Irreducible quadratic polynomials in $\mathbb{R}[x]$ are those for which the quadratic formula requires square roots of negative numbers. In this case the polynomial has two complex roots, and the roots are complex conjugates of each other.

In $\mathbb{Q}[x]$ the situation is even worse. We can find polynomials of any degree that have no roots in $\mathbb{Q}$, and we can find polynomials of any degree that are irreducible, and do not factor into smaller degree factors.

Exercise 10. Show that the only irreducible polynomials in $\mathbb{C}[x]$ are the linear polynomials.

Exercise 11. Factor x^4-1 into irreducible polynomials in $\mathbb{C}[x]$. Factor x^4-1 into irreducible polynomials in $\mathbb{R}[x]$.

Exercise 12. Assume the fundamental theorem of algebra, Part 1. Prove from this the fundamental theorem of algebra, part 2. (Use induction based on degree. Start with degree 1).

Bibliography

[1] Richard Dedekind, *Stetigkeit und irrationale Zahlen (Continuity and Irrational Numbers)* (1872). Currently published by Dover University Press in an English translation as part of *Essays on the Theory of Numbers* (1963). A pioneering work on foundations of the real numbers $\mathbb{R}$. Introduces the construction of the real numbers using what we now call *Dedekind cuts*.

[2] Richard Dedekind, *Was sind und was sollen die Zahlen (What numbers are and what they should be)* (1888). Currently published by Dover University Press in an English translation as part of *Essays on the Theory of Numbers* (1963). Another pioneering work on foundations of number systems. This contains his approach to the natural numbers using essentially what we call "Peano's axioms".

[3] Solomon Feferman, *The Number Systems: Foundations of Algebra and Analysis*, Chelsea Publishing Company (1964). This book served as an important inspiration and source for the current book.

[4] Edmund Landau, *Foundations of Analysis: The arithmetic of whole, rational, irrational, and complex numbers. A supplement to Text-Books on the Differential and Integral Calculus*, Translated by F. Steinhardt Chelsea Publishing Company (1951—1966). Originally published as *Grundlagen der Analysis* in German (1930). This is in some sense *the* classic number systems textbook.

[5] Daniel J. Velleman, *How to Prove It: A Structured Approach*, Cambridge University Press (3rd edition, 2019). This book is an example of a textbook that gives the logical background needed for this book.