Rings With Inquiry

Rings With Inquiry

Michael Janssen

Dordt University

Melissa Lindsey

University of Wisconsin-Madison

Edition: Fall 2020 Beta Edition

Website: https://ringswithinquiry.org

Introduction

In [1], the author defines a *purely structural property* as one that "can be defined wholly in terms of the concepts same and different, and part and whole (along with purely logical concepts)." This definition and its reference to parts and wholes calls to mind the history of the word *algebra* itself, which comes from the Arabic *al-jabr*, literally meaning "the reunion of broken parts". One of the concepts fundamental to the historical development of algebra is the notion of *factorization*; closely related questions that have driven the development of algebra over the centuries are: when does a polynomial equation have solutions in a particular number system, and is there a systematic way to find them?

The goal of this book is to explore the idea of factorization from an abstract perspective. In Chapter 1, we develop our foundations in the integers. Much of this chapter will be familiar to students who have had a first course in number theory, but we are especially concerned with results that preview the structural questions we'll investigate in more abstract settings.

In Chapter 2, we begin the process that so defines modern mathematics: abstraction. We ask: from the point of view of algebra, what properties of the integers are really important? And then we study those. We find that they are also held by several other familiar collections of numbers and algebraic objects, and then study those objects in increasing depth.

Chapter 3 begins with an exploration of factorization properties of polynomials in particular. We then precisely describe what we mean by "unique factorization" before demonstrating that every Euclidean domain possesses the unique factorization property. We conclude with a brief exploration of the ways in which systems of numbers and polynomials can fail to possess the unique factorization property.

Finally, in Chapter 4, we explore the concept of ideals in general, and use

them to build new rings and study properties of homomorphisms.

Throughout this book, we will walk in the realms of abstraction, and catch glimpses of the beauty and incredible power of this perspective on mathematics. ReferencesReferencesg:references:idm180004608224 J. Franklin, *An Aristotelian Realist Philosophy of Mathematics: Mathematics as the Science of Quantity and Structure*, Palgrave Macmillan UK, 2014

A Note to Students

Welcome! We are so glad you are here.

This book was written in response to a bit of student feedback from one of the author's (Janssen) first time teaching his institution's introduction to abstract algebra. In short, the feedback was: *I am a future teacher, and I do not know why I had to take this course.* The criticism was fair; much of what we had covered that semester did not look like what one typically thinks of as "algebra", yet it definitely was. In the intervening years, we set out to rethink the way we introduce students to this beautiful subject, and subsequently developed this free resource to do so.

The focus of the first three chapters is *factorization*, the ability to write certain objects (e.g., numbers, polynomials) as a product of simpler objects (such as prime numbers). This is something you have likely been doing for several years, but we have generally taken it as an article of faith that this can be done in a unique way. The first three chapters of this work has the goal of uncovering structural conditions sufficient to guarantee something like the unique factorization into primes that we know and love from the integers. The fourth chapter is a coda that allows for a deeper exploration of our main objects of interest (rings), as well as functions that relate these objects to one another.

This book was written with the belief that the best way to learn mathematics is to *do* mathematics. Thus, there are vanishingly few worked examples or proved theorems. Instead, you will provide the proofs and solutions to exercises. It is possible, if you look hard enough, to find some of the answers elsewhere on the internet. I implore you to resist the urge to do so. You will learn much more by struggling with a problem, even if you do not ultimately solve it, than you will by searching for an answer after a few minutes of toying

with a problem. The rewards of your struggle will be deep and long-lasting.

Let's begin.

A Note to Instructors

Welcome! Thank you for considering this text; it won't be for everyone, as strong opinions informed its creation. The strongest are (a) that students learn math best by doing it, and (b) that students–especially pre-service teachers–more naturally learn modern algebra by encountering rings first.

Pedagogically, these notes fall under the big tent of *inquiry-based learning* (IBL). Broadly, there are several types of statements you'll find as you read these notes.

- *Theorems*: A *numbered* theorem is a statement that students are expected to prove for themselves. The authors generally assign 3–6 numbered theorems (or exercises, or lemmas) for each class meeting, with students expected to present their work during the next class. These presentations and the ensuing discussions form the regular work of the class. Students are *not* expected to prove *unnumbered* theorems. The unnumbered theorems unify nearby numbered theorems (such as stating an existence theorem and uniqueness theorem as a single result), or are otherwise too technical or complicated to be illuminating. Nonetheless, they are generally important results of which students should be aware.

- *Lemmas*: There are a few lemmas in the notes. As a rule, these lemmas pull out a step from nearby theorems that might be too big to reasonably expect students to take by themselves. If you would like to suggest additional lemmas, feel free to get in touch with the authors.

- *Exercises*: The exercises are generally computational in nature, and presage an upcoming generalization (or reinforce a recent theorem). They are generally labeled as *Activity*, *Exploration*, or *Investigation*. As such, more than a correct numerical answer is needed for a good solution to

an exercise.

- *Challenges*: There are a few (unnumbered) challenge problems in the text. These problems may be assigned or they may not, but they are generally difficult and their omission will not disrupt the flow of the text. Students may be interested merely in knowing their statements (e.g., $\mathbb{Z}[x]$ is not a PID).

We begin with a brief overview of some results from elementary number theory regarding divisibility and primes, and introduce modular arithmetic. Other than induction, no proof techniques are explicitly discussed. It is assumed that students using these notes have had an introduction to proofs.

Brief attention is paid to fields before we dive in to rings. Other than mentioning their existence, no attention is given to noncommutative rings. Rings and ideals are developed with an eye toward eventually proving that every Euclidean domain is a unique factorization domain. We briefly explore nonunique factorization (though this could be done in outside homework, if desired) before turning to an exploration of homomorphisms and ideals in general.

As of this writing (July 2020), groups are not covered in this book. Depending on personal preference, with the time left at the end of the semester (often approximately 1–3 weeks, depending on your class's pace), you could present an introduction to groups directly to your students, or use freely available IBL material from the *Journal of Inquiry-Based Learning in Mathematics*.

The book has been used to carry a full semester course at least three times: twice at Dordt University (Fall 2018 and 2020), and once at Morningside College (Fall 2019). Future plans for the text include:

- An expanded treatment of fields, with an emphasis on extensions of $\mathbb{Q}[x]$ (and possible introduction to groups via permutations of roots of polynomials).
- Additional optional end-of-section exercises.
- Integration of SageMath cells to aid computation where appropriate.
- PDF and HTML versions of both the student and instructor versions.
- Low-cost print copies of the student version.

There is no planned timeline for any of these projects. If you are interested in helping make one of these happen, please email me (Mike) and let me know! Or, if you just want to let me know that you've found the text useful, that would also be welcome news. And of course, if you find any typos or mistakes, I would love to know that as well.

Acknowledgements

There are several people we wish to thank. In no particular order:

David Farmer, for handling the initial conversion of the LaTeX code to PreTeXt.

Mitch Keller, for his help with PreTeXt (especially creating the Student Version of the book) and for being willing to take a chance using this book for his Fall 2019 modern algebra course at Morningside College.

The Network of IBL Communities (and NSF-DUE #1925188) and the UNO STEM TRAIL Center for supporting the collaboration with Mitch Keller.

Rob Beezer and the PreTeXt team for creating this wonderful authoring language.

Contents

Chapter 1

The Integers

As children we start exploring the properties and structure of the positive integers as soon as we learn to count and we extend our understanding throughout our schooling as we learn about new operations and collections of numbers. We begin our journey into abstract algebra with an overview of some familiar (and some possibly unfamiliar) properties of the integers that are relevant to our course of inquiry. With this foundation set, we will see in later chapters just how far we can extend these properties in more abstract setting.

1.1 Induction and Well-Ordering

Guiding Questions.

In this section, we'll seek to answer the questions:

- What is the Well-Ordering Principle?
- What is mathematical induction, and how can we use it to prove statements about $\mathbb{N}$?

In this section we will assume the basic algebraic/arithmetic properties of the integers such as closure under addition, subtraction, and multiplication, most of which we will formalize via axioms in subsequent sections. Axiom 1.1.2 formalizes the familiar notion that nonempty subsets of the positive integers have a smallest element, which will be used repeatedly throughout the text. We then explore a closely related idea, mathematical induction, via an example and exercises.

Definition 1.1.1 The collection of **natural numbers** is denoted by $\mathbb{N}$, and is the set

$$\mathbb{N} = \{1, 2, 3, \ldots\}.$$

By $\mathbb{N}_0$ we mean the set $\mathbb{N} \cup \{0\} = \{0, 1, 2, 3, \ldots\}$. $\Diamond$

In some sense, the fundamental properties of $\mathbb{N}$ are (a) there is a smallest natural number, and (b) there is always a next natural numberIn fact, one can build a model of $\mathbb{N}$ with set theory and the Peano axioms, which utilize the notion of a *successor*--the next natural number.). A consequence of the Peano postulates is the *well-ordering principle*, which we state as an axiom.

Axiom 1.1.2 Well-Ordering Principle. *Every nonempty subset of* $\mathbb{N}_0$ *contains a least (smallest) element under the usual ordering,* $\leqslant$[1].

The Well-Ordering Principle is useful for producing smallest elements of nonempty subsets defined to have certain properties, as the following example demonstrates.

Exploration 1.1.1 In this exploration, we investigate polynomials with real coefficients, as well as their degrees. We will define these terms more formally in Definition 2.2.1, but for now you may use your intuition from previous courses in algebra.

Let S be the set of all polynomials f in the variable x with real coefficients such that $f(2) = f(-2) = 0$ and $f(0) = -4$.

1. Give an example of an $f \in S$ and $g \notin S$.

2. Let $D = \{\deg f : f \in S\}$ be the set of possible degrees of polynomials in S. Show that $D \neq \emptyset$ and $D \subseteq \mathbb{N}_0$.

3. Apply the Well-Ordering Principle to argue that D has a least element. To what does this correspond in S?

Solution. TBD.

We will use this principle throughout the text, next in Theorem 1.2.5.

[1] Our word choice is suggestive. In fact, other orderings do exist, and while the set of nonnegative real numbers $\mathbb{R}$ does not satisfy the well-ordering principle under the usual ordering $\leqslant$, the Well Ordering Axiom asserts that there exists a well ordering on *any* set, including $\mathbb{R}$. Accepting this axiom is equivalent to accepting the axiom of choice.

Figure 1.1.3 A suspect use of the Well-Ordering Principle.

Definition 1.1.4 The set of integers consists of the positive and negative natural numbers, together with zero, and is denoted by $\mathbb{Z}$:

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

$\Diamond$

Mathematical Induction.

Let $P(m)$ be a statement about the natural number m[2]. Let $k_0 \in \mathbb{N}$ be such that the statement $P(k_0)$ is true (the *base case*), and suppose there is an $n \geqslant k_0$ such that for all k satisfying $k_0 \leqslant k \leqslant n$, $P(k)$ is true (the *inductive hypothesis*). Then $P(n+1)$ is true, and thus $P(m)$ is true for all $m \geqslant k_0$ (the *inductive step*).

Mathematical induction is like climbing an infinite staircase. The *base case* tells us that we can take a first step on the staircase (k_0). In the *inductive hypothesis*, we assume we can take all the steps up to a certain height (n). In the *inductive step*, we prove that this allows us to take the $(n+1)$st step.

Thus, if we can take step k_0, we can (by the inductive step) take step k_0+1. And since we can take step k_0+1, we can (again by the inductive step) take step k_0+2. And so on, forever (or, if the notion of actual infinity makes you uncomfortable, as far as we want to go).

[2]Sample statements could include "m is really interesting" or "$3m^2+m+2$ is even".

Example.

For all $n \geqslant 1$,

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}.$$

Proof. Base case: When $n = 1$, the equation $1 = \frac{1 \cdot (1+1)}{2}$ is true.

Inductive Hypothesis: Assume that there exists a n such that whenever $k \leqslant n$, the equation

$$1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2} \tag{1.1.1}$$

is true.

Inductive Step: Our goal is to show that $P(n+1)$ is true. That is, we wish to establish that

$$1 + 2 + 3 + \cdots + n + (n+1) = \frac{(n+1)((n+1)+1)}{2}. \tag{1.1.2}$$

We begin on the left-hand side of (1.1.1), where we may apply the inductive hypothesis to see that

$$1 + 2 + 3 + \cdots + n + (n+1) = \left[\frac{n(n+1)}{2}\right] + (n+1). \tag{1.1.3}$$

Through the use of straightforward algebra, the right-hand side becomes

$$\frac{n(n+1)}{2} + \frac{2(n+1)}{2} = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2}. \tag{1.1.4}$$

Putting (1.1.3) and (1.1.4) together, we obtain

$$1 + 2 + 3 + \cdots + n + (n+1) = \frac{(n+1)((n+1)+1)}{2},$$

which is exactly the goal we stated in (1.1.2).

We conclude with opportunities to practice induction.

Theorem 1.1.5 *For all $k \geqslant 1$, $3^k > k$.*
Proof. TBD. ■

Theorem 1.1.6 *Prove that the sum of the first n cubes is $\frac{n^2(n+1)^2}{4}$. That is,*

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}.$$

Proof. We wish to prove that

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}.$$

Base Case: When $n = 1$, the theorem clearly holds.

Inductive Hypothesis: Assume for all $k \leqslant n$ that we have

$$1^3 + 2^3 + 3^3 + \cdots + k^3 = \frac{k^2(k+1)^2}{4}. \tag{1.1.5}$$

Inductive Step: We wish to prove that

$$1^3 + 2^3 + 3^3 + \cdots + n^3 + (n+1)^3 = \frac{(n+1)^2((n+1)+1)^2}{4} = \frac{(n+1)^2(n+2)^2}{4}.$$

We apply (1.1.5) to obtain

$$\begin{aligned} 1^3 + 2^3 + 3^3 + \cdots + n^3 + (n+1)^3 &= \frac{n^2(n+1)^2}{4} + (n+1)^3 \\ &= \frac{n^2(n+1)^2}{4} + \frac{4(n+1)^3}{4} \\ &= \frac{n^2(n+1)^2 + 4(n+1)^3}{4} \\ &= \frac{(n+1)^2(n^2 + 4(n+1))}{4} \\ &= \frac{(n+1)^2(n^2 + 4n + 4)}{4} \\ &= \frac{(n+1)^2(n+2)^2}{4}. \end{aligned}$$

The theorem is proved. ■

Theorem 1.1.7 (Bernoulli's Inequality). *Given a real number $b > -1$, $(1+b)^n \geqslant 1 + bn$ for all $n \in \mathbb{N}_0$.*

Proof. TBD. ■

1.2 Divisibility and GCDs in the Integers

Guiding Questions.

In this section, we'll seek to answer the questions:

- What does it mean for one integer to divide another?
- What properties does divisibility enjoy in the integers?
- What is the greatest common divisor of two integers?
- How can we compute the greatest common divisor of two integers?

1.2.1 Divisibility and the Division Algorithm

In this section, we begin to explore some of the arithmetic and algebraic properties of $\mathbb{Z}$. We focus specifically on the divisibility and factorization properties

of the integers, as these are the main focus of the text as a whole. One of the primary goals of this section is to formalize definitions that you are likely already familiar with and of which you have an intuitive understanding. At first, this might seem to unnecessarily complicate matters. However, it will become clear as we move forward that formal mathematical language and notation are necessary to extend these properties to a more abstract setting. We begin with a familiar notion.

Definition 1.2.1 Let $a, b \in \mathbb{Z}$. We say that a **divides** b, and write $a \mid b$, if there is an integer c such that $ac = b$. In this case, say that a and c are **factors** of b. If no such $c \in \mathbb{Z}$ exists, we write $a \nmid b$. $\Diamond$

Note that the symbol $\mid$ is a *verb*; it is therefore correct to say, e.g., $2|4$, as 2 *does* divide 4. However, it is an abuse of notation to say that $2 \mid 4 = 2$. Instead, we likely mean $4 \div 2 = 2$ or $\frac{4}{2} = 2$ (though we will not deal in fractions just yet).

Investigation 1.2.1 Determine whether $a \mid b$ if:

1. $a = 3$, $b = -15$
2. $a = 4$, $b = 18$
3. $a = -7$, $b = 0$
4. $a = 0$, $b = 0$

Comment briefly on the results of this investigation. What did you notice? What do you still wonder?

Solution. TBD.

We next collect several standard results about divisibility in $\mathbb{Z}$ which will be used extensively in the remainder of this text.

Theorem 1.2.2 *Let $a, b, c \in \mathbb{Z}$. If $a \mid b$ and $a \mid c$, then $a \mid (b + c)$.*
Proof. Our hypothesis means that there exist integers k_1 and k_2 such that $b = ak_1$ and $c = ak_2$. Then $b + c = ak_1 + ak_2 = a(k_1 + k_2)$. Let $k' = k_1 + k_2$, and observe that k' is an integer; then $b + c = ak'$, so $a \mid (b + c)$. ■

Theorem 1.2.3 *Let $a, b, c \in \mathbb{Z}$. If $a \mid b$, then $a \mid bc$.*
Proof. Our hypotheses means there is an integer k for which $b = ak$. Then $bc = (ak)c = a(kc)$. We observe that $k' = kc$ is an integer, so $bc = ak'$, and therefore $a \mid bc$. ■

Investigation 1.2.2 Consider the following partial converse to Theorem 1.2.3: If $a, b, c \in \mathbb{Z}$ with $a|bc$, must $a|b$ or $a|c$? Supply a proof or give a counterexample.

Solution. The converse is false: $6|3 \cdot 4$, but $6 \nmid 3$ and $6 \nmid 4$.

Theorem 1.2.4 *Let $a, b, c, d \in \mathbb{Z}$. If $a = b + c$ and d divides any two of a, b, c, then d divides the third.*

Proof. We consider two cases.

Case 1: Assume without loss of generality that $d|a$ and $d|b$. Then there are integers k_1, k_2 such that $dk_1 = a$ and $dk_2 = b$. We observe that $c = a - b = dk_1 - dk_2 = d(k_1 - k_2)$, so $d|c$.

Case 2: Assume that $d|b$ and $d|c$, so there are $k_1, k_2 \in \mathbb{Z}$ such that $dk_1 = b$ and $dk_2 = c$. Then $a = b + c = dk_1 + dk_2 = d(k_1 + k_2)$, so $d|a$. ■

Investigation 1.2.3 Formulate a conjecture akin to the previous theorems about divisibility in $\mathbb{Z}$, and then prove it.

Solution. Answers vary.

As we saw above, not all pairs of integers a, b satisfy $a \mid b$ or $b \mid a$. However, our experience in elementary mathematics does apply: there is often something left over (a remainder). The following theorem formalizes this idea for $a, b \in \mathbb{N}$.

Theorem 1.2.5 The Division Algorithm for $\mathbb{N}$. *Let $a, b \in \mathbb{N}$. Then there exist unique integers q, r such that $a = bq + r$, where $0 \leqslant r < b$.* ***Hint 1****. There are two parts to this theorem. First, you must establish that q and r exist. This is best done via Axiom 1.1.2. If you're stuck on that, check the second hint.*

Once you have established that q and r exist, show that they are unique but assuming $a = bq + r$ and $a = bq' + r'$, where r, r' both satisfy the conditions of the theorem. Argue that $q = q'$ and $r = r'$.

Hint 2*. Let $S = \{a - bs : s \in \mathbb{N}_0, a - bs \geqslant 0\}$.*

Proof. First, assume that $a < b$. Then $q = 0$ and $r = a$ are sufficient. If $a = b$, then $q = 1$ and $r = 0$ will work.

Now, assume that $a > b$. Then the set $S = \{s \in \mathbb{N} : a - bs \geqslant 0\}$ is nonempty, as $a - b \in S$. By Axiom 1.1.2, S has a least element, which we will call r. Let $q \in \mathbb{N}_0$ be such that $r = a - bq$; then $a = bq + r$.

To finish the proof, it is enough to show that this choice of r satisfies $0 \leqslant r \leqslant b - 1$. Observe that $r \in S$, so $r \geqslant 0$. If $r \geqslant b$, then $a - b(q + 1) = a - bq - b = r - b \geqslant 0$. Since $q + 1 > q$, this is a contradiction to the assumption that q was the largest element of S. Thus, $r < b$, i.e., $r \leqslant b - 1$.

Let q, q', r, r' be such that $a = bq + r$ and $a = bq' + r'$. We see that $bq + r = bq' + r'$, so $b(q - q') = r' - r$, so $b|r' - r$. If $r' - r = 0$, we are done.

Otherwise, $r' - r$ is a nonzero multiple of n. If $r' \geqslant r$, then $0 \leqslant b(q - q') = r' - r \leqslant b - 1 - r$, a contradiction. Similarly, if $r' \leqslant r$, write $b(q' - q) = r - r' \geqslant 0$, and a symmetric argument leads to a contradiction.

Thus, $r' = r$, so $b(q - q') = 0$, and since $b > 0$, $q - q' = 0$, i.e., $q = q'$. ■

Warning! This theorem has two parts: existence and uniqueness. Do not try to prove them both at the same time.

Unsurprisingly, the Division Algorithm also holds in $\mathbb{Z}$, though the existence of negative integers requires a careful restatement.

Corollary 1.2.6 The Division Algorithm for $\mathbb{Z}$. *Let $a, b \in \mathbb{Z}$ with $b \neq 0$. Then there exist unique integers q, r such that $a = bq + r$, where $0 \leqslant r < |b|$.* ***Hint.*** *Consider cases, and apply Theorem 1.2.5 wherever possible.*

Proof. TBD. ■

1.2.1.1 Greatest Common Divisors

We next turn to another familiar property of the integers: the existence of greatest common divisors.

Definition 1.2.7 Let $a, b \in \mathbb{Z}$ such that a and b are not both 0. A **greatest common divisor** of a and b, denoted $\gcd(a, b)$, is a natural number d satisfying

1. $d \mid a$ and $d \mid b$
2. if $e \in \mathbb{N}$ and $e \mid a$ and $e \mid b$, then $e \mid d$.

If $\gcd(a, b) = 1$, we say that a and b are **relatively prime** or **coprime**.

◊

This definition may be different than the one you are used to, which likely stated that $d \geqslant e$ rather than condition 2 in Definition 1.2.7. It can be proved using the order relations of $\mathbb{Z}$ that the definition given here is equivalent to that one. However, we will prefer this definition, as it generalizes naturally to other number systems which do not have an order relation like $\mathbb{Z}$.

Activity 1.2.4 Compute $\gcd(a, b)$ if:

1. $a = 123$, $b = 141$
2. $a = 0$, $b = 169$
3. $a = 85$, $b = 48$

Now that you have had a bit of practice computing gcds, describe your method for finding them in a sentence or two.

Solution. TBD.

How did you answer the last question in Activity 1.2.4? If you are like the authors' classes, the answers probably varied, though you have referred

at some point to a "prime" (whatever those are), or possibly some other ad hoc method for finding the gcd. Most such methods rely in some form on our ability to factor integers. However, the problem of factoring arbitrary integers is actually surprisingly computationally intensive. Thankfully, there is another way to compute $\gcd(a, b)$, to which we now turn.

Theorem 1.2.8 *Let $a, b, c \in \mathbb{Z}$ such that $a = b + c$ with a and b not both zero. Then $\gcd(a, b) = \gcd(b, c)$.*

Proof. Let $d = \gcd(a, b)$ and $e = \gcd(b, c)$. Since $d|a$ and $d|b$, $d|c$, and so $d|e$ by definition.

Similarly, since $e|b$ and $e|c$, $e|a$, so $e|d$ by definition.

Thus, $e = d$. ■

Investigation 1.2.5 Suppose $a, b, c \in \mathbb{Z}$ such that there exists $q \in \mathbb{Z}$ with $a = bq + c$ and a and b not both zero. Prove or disprove: $\gcd(a, b) = \gcd(a, c)$.

Solution. TBD.

Investigation 1.2.6 (Euclidean Algorithm). Let $a, b \in \mathbb{N}$. Use Theorem 1.2.5 and Investigation 1.2.5 to determine an algorithm for computing $\gcd(a, b)$. How could your method be modified to compute $\gcd(a, b)$ for $a, b \in \mathbb{Z}$?

Solution. Let $a, b \in \mathbb{N}$. Write $a = bq_1 + r_1$. By Theorem 1.2.5, $\gcd(a, b) = \gcd(b, r_1)$.

Then, write $b = r_1q_2 + r_2$ using the Division algorithm. Observe $\gcd(b, r_1) = \gcd(r_1, r_2)$. Continue until there is a k for which $r_k = 0$, so $\gcd(r_k, r_{k-1}) = r_{k-1}$. Then $r_{k-1} = \gcd(r_k, r_{k-1}) = \gcd(r_{k-1}, r_{k-2}) = \cdots = \gcd(a, b)$.

Activity 1.2.7 Use the Euclidean algorithm to compute $\gcd(18489, 17304)$.

Solution. We write

$$\begin{aligned} 18489 &= 17304 \cdot 1 + 1185 \\ 17304 &= 1185 \cdot 14 + 714 \\ 1185 &= 714 \cdot 1 + 471 \\ 714 &= 471 \cdot 1 + 243 \\ 471 &= 243 \cdot 1 + 228 \\ 243 &= 228 \cdot 1 + 15 \\ 228 &= 15 \cdot 15 + 3 \\ 15 &= 3 \cdot 5 + 0. \end{aligned}$$

So $\gcd(18489, 17304) = 3$.

The following identity provides a useful characterization of the greatest common divisor of two integers, not both zero. We will return to this idea several times, even after we have left the familiar realm of the integers.

Theorem 1.2.9 Bézout's Identity. *For any integers a and b not both 0, there are integers x and y such that*

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

Hint 1. *Apply Axiom 1.1.2 to a well-chosen set.*

Hint 2. *Apply Axiom 1.1.2 to $S = \{as + bt : s, t \in \mathbb{Z},\ as + bt > 0\}$.*

Proof. Consider the set $S = \{as + bt : s, t \in \mathbb{Z},\ as + bt > 0\}$. We first show that $S \neq \emptyset$.

Suppose without loss of generality that $a \neq 0$, and consider $a > 0$; then $s = 1, t = 0$ is sufficient to guarantee $as + bt \in S$. If $a < 0$, then $s = -1, t = 0$ is sufficient. Regardless, $S \neq \emptyset$, and by the Well-Ordering Principle, S has a least element, $d = ax + by$. We claim $d = \gcd(a, b)$.

We first show that $d|a$ Use the division algorithm to write $a = dq + r$, where $r < d$. We find $r = a - dq = a - (ax + by)q = a(1 - xq) + by$, and if $r > 0$, $r \in S$, contradicting the status of d as the least element of S. Thus, $r = 0$, and $d|a$. By a symmetric proof, $d|b$.

Suppose then that e is a common divisor of a and b; write $a = ek_1$ and $b = ek_2$. Then $d = ax + by = (ek_1)x + (ek_2)y = e(k_1x + k_2y)$, whence $e|d$. Therefore, $d = \gcd(a, b)$. ■

We conclude with an answer to the questions raised by Investigation 1.2.2.

Theorem 1.2.10 *Let a, b, and c be integers. If $a|bc$ and $\gcd(a, b) = 1$, then $a|c$.*

Proof. Suppose $\gcd(a, b) = 1$. By Theorem 1.2.9, there are integers x and y such that

$$ax + by = 1.$$

Multiply by c to get

$$acx + bcy = c.$$

Since $a|bc$ by assumption, there is an integer k such that $ak = bc$, so we have

$$acx + bcy = acx + aky = a(cx + ky) = c,$$

so $a|c$. ■

In this section, we have collected some initial results about divisibility in the

integers. We'll next explore the multiplicative building blocks of the integers, the primes, in preparation for a deeper exploration of factorization.

1.3 Primes and Factorization

Guiding Questions.

In this section, we'll seek to answer the questions:

- What are primes? What properties do they have?
- What does the Fundamental Theorem of Arithmetic say?
- Why is the Fundamental Theorem of Arithmetic true?

As described in the Preface , our main goal is to build a deep structural understanding of the notion of *factorization*. That is, splitting objects (e.g., numbers, polynomials, matrices) into products of other objects. One of the most familiar examples of this process involves factoring integers into products of primes.

Definition 1.3.1 Let $p > 1$ be a natural number. We say p is **prime** if whenever $a, b \in \mathbb{Z}$ such that $p \mid ab$, either $p \mid a$ or $p \mid b$.

A natural number $m > 1$ is said to be **composite** if it is not prime. ◊

This is almost certainly not the definition of prime that you are familiar with from your school days, which likely said something to the effect that a prime $p > 1$ is a natural number only divisible by 1 and itself. However, Definition 1.3.1 is often more useful than the usual definition. And, as Lemma 1.3.2 demonstrates, the two notions are equivalent.

Lemma 1.3.2 Euclid's Lemma. *Given any $p \in \mathbb{N}$, $p > 1$, p is prime if and only if whenever $m \in \mathbb{N}$ divides p, either $m = p$ or $m = 1$.*

Proof. Let p be prime and suppose $m \in \mathbb{N}$ divides p, so $p = mk$ for some $k \in \mathbb{N}$. By definition, $p|m$ or $p|k$. If $p|m$, then $pj = m$, so $p = pjk$, and $jk = 1$, which means $j = k = 1$ and thus $m = p$. If $p|k$, then $pj = k$, so $p = mpj$, and $mj = 1$, so $m = j = 1$.

Conversely, assume $p \in \mathbb{N}$ has the property that whenever an integer m exists with $m|p$, then $m = p$ or $m = 1$. Suppose $p|ab$, so there exists a $k \in \mathbb{N}$ such that $pk = ab$. Note that if $\gcd(a, p) = d > 1$, then $d|p$, and so $d = p$, and then $p|a$. If $\gcd(a, p) = 1$, then by Bézout's Identity there are integers x and y such that $ax + py = 1$, so we may multiply by b to obtain $abx + pby = b$,

whence $p|b$. ■

Exploration 1.3.1 Using Lemma 1.3.2 as a guide, give a biconditional characterization for composite numbers. That is, finish the sentence: "A number $m \in \mathbb{N}$ is composite if and only if ...."

Answer. "A number $m \in \mathbb{N}$ is composite if and only if there exist natural numbers $a, b \neq 1$ such that $m = ab$."

Remark 1.3.3 How does your definition treat the number 1? The primality of 1 has been the subject of much debate stretching back to the Greeks (most of whom did not consider 1 to be a number). Throughout history, mathematicians have at times viewed 1 as prime, and at other times, not prime. The main argument for the non-primality of 1 is that if 1 were taken to be prime, we would need to word theorems like the Fundamental Theorem of Arithmetic (below) in such a way that only prime factorizations not including 1 can be considered. For, if 1 is prime, we would have to consider, e.g., $6 = 2 \cdot 3 = 1 \cdot 2 \cdot 3 = 1 \cdot 1 \cdot 2 \cdot 3$ as three different factorizations of 6 into primes.

However, neither is 1 composite (your definition should rule this out in some way). Instead, we call 1 a **unit**, which we'll explore more fully in Definition 2.2.7 and the following; consequently, the opposite of "prime" is not "composite", but "not prime".

Theorem 1.3.4 *Let $a \in \mathbb{N}$ such that $a > 1$. Then there is a prime p such that $p \mid a$.*

Proof. We proceed by mathematical induction. Note that when $a = 2$, the statement holds.

Assume that there is a $k \in \mathbb{N}$ such that for all $a \leqslant k$, the statement holds. Consider $k+1$. If $k+1$ is prime, we are done. If not, then $k+1$ is composite, and by the answer to Question 1.3.1, there are integers $a, b \neq 1$ such that $k+1 = ab$. By induction, there is a prime p such that $p|a$, so $p|k+1$. ■

Theorem 1.3.5 *Suppose p and q are primes with $p|q$. Then $p = q$.*

Proof. TBD. ■

Our first major theorem makes two claims: that positive integers greater than 1 *can* be factored into products of primes, and that this factorization can happen in only one way. As the semester progresses, we will see other theorems like this one, and catch glimpses of other ways to think about the *unique factorization property*.

Fundamental Theorem of Arithmetic.

Every natural number greater than 1 is either a prime number or it can be expressed as a finite product of prime numbers where the expression is unique up to the order of the factors.

The proof is broken into two parts: existence (Theorem 1.3.6) and uniqueness (Theorem 1.3.8).

Theorem 1.3.6 Fundamental Theorem of Arithmetic–Existence Part[1]**.** *Every natural number $n > 1$ is either a prime number or it can be expressed as a finite product of prime numbers. That is, for every natural number $n > 1$, there exist primes $p_1, p_2, \ldots, p_m$ and natural numbers $r_1, r_2, \ldots, r_m$ such that*

$$n = p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m}.$$

Hint. *Induction!*

Proof. It is enough to show that, if $n > 1$, we may write $n = p_1 p_2 \cdots p_k$, where the p_i's are not necessarily distinct primes. Then we can collect the common primes and write them with exponent notation.

Note that when $n = 2$, there is such an expression, so let $k > 2$ be the least positive integer that fails to be expressed as above. We note that k cannot be prime, or it could be expressed as above. Thus, by Theorem 2.1, there is a prime p such that $k = pb$, where $1 < p, b < k$. Since k was the least positive integer that cannot be factored as a product of primes, b has a prime factorization. But then so does k, a contradiction. ■

Lemma 1.3.7 *Let p and $q_1, q_2, \ldots, q_n$ all be primes and let k be a natural number such that $pk = q_1 q_2 \ldots q_n$. Then $p = q_i$ for some i.*

Proof. We first state a claim that will be useful.

Claim 1: If p, q are primes such that $p|q$, then $p = q$.

Proof of Claim 1: Since $p, q > 1$ and both p and q are prime, $p = q$ by the definition of primality. ✓

Now assume that $pk = q_1 q_2 \cdots q_n = (q_1 q_2 \cdots q_{n-1}) q_n$. If $p|q_n$, we're done by Claim 1. If not, the definition of a prime guarantees that $p|q_1 q_2 \cdots q_{n-1}$. If $p|q_{n-1}$, we're done by Claim 1. Otherwise, $p|q_1 q_2 \cdots q_{n-2}$.

Repeating this process, we see that either $p|q_i$ for some $i > 2$, or $p|q_1 q_2$. Now the definition of a prime guarantees that $p|q_1$ or $p|q_2$, in which case Claim 1 requires that $p = q_1$ or $p = q_2$. ■

[1] This approach to the Fundamental Theorem of Arithmetic is adapted from [1.3.1].

Theorem 1.3.8 Fundamental Theorem of Arithmetic–Uniqueness Part. *Let n be a natural number. Let $\{p_1, p_2, \ldots, p_m\}$ and $\{q_1, q_2, \ldots, q_s\}$ be sets of primes with $p_i \neq p_j$ if $i \neq j$ and $q_i \neq q_j$ if $i \neq j$. Let $\{r_1, r_2, \ldots, r_m\}$ and $\{t_1, t_2, \ldots, t_s\}$ be sets of natural numbers such that*

$$\begin{aligned} n &= p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m} \\ &= q_1^{t_1} q_2^{t_2} \cdots q_s^{t_s}. \end{aligned}$$

Then $m = s$ and $\{p_1, p_2, \ldots, p_m\} = \{q_1, q_2, \ldots, q_s\}$. That is, the sets of primes are equal but their elements are not necessarily listed in the same order (i.e., p_i may or may not equal q_i). Moreover, if $p_i = q_j$, then $r_i = t_j$. In other words, if we express the same natural number as a product of distinct primes, then the expressions are identical except for the ordering of the factors. ***Hint.*** *Argue that the two sets are equal (how do we do that?). Then argue that the exponents must also be equal.*

Proof. Without loss of generality, assume $p_1 < p_2 < \cdots < p_m$ and $q_1 < q_2 < \cdots < q_s$. Given a p_i, we know that $p_i | q_1^{t_1} q_2^{t_2} \cdots q_s^{t_s}$, which implies that $p_i | q_j$ for some j by Claim 2 in the proof of Lemma 2.8. Further, Lemma 2.8 implies that $p_i = q_j$, and similarly, given a q_j, $q_j = p_i$ for some i. Thus, $m = s$, and by the ordering of the p_i's and q_j's, we have $p_i = q_j$, $i = 1, 2, \ldots, m$. Therefore,

$$n = p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m} = p_1^{t_1} p_2^{t_2} \cdots p_m^{t_m}.$$

Now, assume by way of contradiction, that $r_i \neq t_i$ for some i. Without loss of generality, we may assume $r_i < t_i$. Then $p_i^{t_i} | p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m}$, which implies that

$$p_i^{t_i - r_i} | p_1^{r_1} p_2^{r_2} \cdots \hat{p}_i^{r_i} \cdots p_m^{r_m}.$$

Since $t_i - r_i > 0$, we have that

$$p_i | p_1^{r_1} p_2^{r_2} \cdots \hat{p}_i^{r_i} \cdots p_m^{r_m},$$

from which $p_i | p_j$ for some $j \neq i$ by Lemma 2.8. This is a contradiction, so $r_i = t_i$ for $i = 1, 2, \ldots, m$. ■

Our first major result is in hand: we can factor natural numbers $n > 2$ uniquely as a product of primes. Much of the rest of this book seeks to deduce a generalization of this result that relies on structural arithmetic properties enjoyed by $\mathbb{Z}$ and similar objects.

References

D. Marshall, E. Odell, M. Starbird, *Number Theory Through Inquiry*, MAA Textbooks, Mathematical Association of America, 2007

1.4 The Integers modulo m

Guiding Questions.

In this section, we'll seek to answer the questions:

- What are equivalence relations?
- What is congruence modulo m?
- How does arithmetic in $\mathbb{Z}_m$ compare to arithmetic in $\mathbb{Z}$?

The foundation for our exploration of abstract algebra is nearly complete. We need the basics of one more "number system" in order to appreciate the abstract approach developed in subsequent chapters. To build that number system, a brief review of relations and equivalence relations is required. Recall that given sets S and T, the Cartesian product of S with T, denoted $S \times T$ ("S cross T"), is the set of all possible ordered pairs whose first element is from S and second element is from T. Symbolically,

$$S \times T = \{(s,t) : \ s \in S, \ t \in T\}.$$

Definition 1.4.1 Let S be a nonempty set. A **relation** R on S is a subset of $S \times S$. If $x, y \in S$ such that $(x,y) \in R$, we usually write xRy and say that x and y are **related under** R. $\Diamond$

The notion of a relation as presented above is extremely open-ended. *Any* subset of ordered pairs of $S \times S$ describes a relation on the set S. Of course, some relations are more meaningful than others; the branch of mathematics known as order theory studies *order* relations (such as the familiar $<$). Our focus will be on *equivalence relations*, which isolate the important features of $=$.

Definition 1.4.2 Let S be a nonempty set. We say a relation $\sim$ on S is an **equivalence relation** if the following properties hold:

- $\sim$ is *reflexive*: if $a \in S$, then $a \sim a$.
- $\sim$ is *symmetric*: if $a, b \in S$ with $a \sim b$, then $b \sim a$.

- $\sim$ is *transitive*: if $a, b, c \in S$ with $a \sim b$ and $b \sim c$, then $a \sim c$.

Given $x \in S$, the set

$$\overline{x} = \{y \in S : \ x \sim y\}$$

is called the **equivalence class of** x. Any element $z \in \overline{x}$ is called a **representative** of the equivalence class. $\Diamond$

Activity 1.4.1 Prove that "has the same birthday as" is an equivalence relation on the set P of all people.

Solution. Given any $x \in P$, clearly x has the same birthday as x.

Moreover, if $x, y \in P$ such that x has the same birthday as y, then it is clear that y has the same birthday as x.

Finally, if $x, y, z \in P$ such that x has the same birthday as y and y has the same birthday as z, then x must have the same birthday as z.

Exploration 1.4.2 What other relations can you think of? Write down one example and one non-example of an equivalence relation.

Activity 1.4.3 Prove that $\leqslant$ is *not* an equivalence relation on $\mathbb{Z}$.

Solution. The relation $\leqslant$ fails the symmetry condition. As an example, note that $2 \leqslant 3$, but $3 \not\leqslant 2$.

For our purposes, a particularly important equivalence relation is congruence modulo m on the set of integers.

Definition 1.4.3 Let $a, b \in \mathbb{Z}$ and $m \in \mathbb{N}$, $m > 1$. We say a **is congruent to** b **modulo** m if $m \mid a - b$. We write $a \equiv b \mod m$. $\Diamond$

Activity 1.4.4 Justify the following congruences.

1. $18 \equiv 6 \mod 12$
2. $47 \equiv 8 \mod 13$
3. $71 \equiv 1 \mod 5$
4. $21 \equiv -1 \mod 11$
5. $24 \equiv 0 \mod 6$

Theorem 1.4.4 *Given an integer $m > 1$, congruence modulo m is an equivalence relation on $\mathbb{Z}$.*

Proof. Let $a \in \mathbb{Z}$. Then $m|a - a$, so $a \equiv_m a$. Thus, $\equiv_m$ is reflexive.

Let $a, b \in \mathbb{Z}$ such that $a \equiv_m b$. This means that $m|a - b$, so there is some $k \in \mathbb{Z}$ such that $mk = a - b$. Then $m(-k) = b - a$, so $m|b - a$ and $b \equiv_m a$. Thus, $\equiv_m$ is symmetric.

Finally, let $a, b, c \in \mathbb{Z}$ such that $a \equiv_m b$ and $b \equiv_m c$. Then $m|a - b$ and

$m|b-c$, so there are integers k_1, k_2 such that $mk_1 = a-b$ and $mk_2 = b-c$. Summing these equations yields $m(k_1+k_2) = (a-b)+(b-c) = a-c$, so $m|a-c$ and $a \equiv_m c$. ■

Exploration 1.4.5 Find all of the equivalence classes of $\mathbb{Z}_5$ and $\mathbb{Z}_7$.

Answer. The equivalence classes of $\mathbb{Z}_5$ are

$$\overline{0} = \{7k : k \in \mathbb{Z}\} = \{\ldots, -5, 0, 5, 10, \ldots\}$$
$$\overline{1} = \{7k+1 : k \in \mathbb{Z}\} = \{\ldots, -4, 1, 6, 11, \ldots\}$$
$$\overline{2} = \{7k+2 : k \in \mathbb{Z}\} = \{\ldots, -3, 2, 7, 12, \ldots\}$$
$$\overline{3} = \{7k+3 : k \in \mathbb{Z}\} = \{\ldots, -2, 3, 8, 13, \ldots\}$$
$$\overline{4} = \{7k+4 : k \in \mathbb{Z}\} = \{\ldots, -1, 4, 9, 14, \ldots\}$$

Since every integer is in one of the above equivalence classes, we know we have found them all.

Theorem 1.4.5 *Let $a, b, c, d \in \mathbb{Z}$ and $m > 1$ such that $a \equiv c \mod m$ and $b \equiv d \mod m$. Then $a+b \equiv c+d \mod m$.*
Proof. Write $mk_1 = a-c$ and $mk_2 = b-d$ for some $k_1, k_2 \in \mathbb{Z}$. Then $m(k_1+k_2) = (a-c)+(b-d) = (a+b)-(c+d)$, so $a+b \equiv c+d \mod m$. ■

Theorem 1.4.6 *Let $a, b, c, d \in \mathbb{Z}$ and $m > 1$ such that $a \equiv c \mod m$ and $b \equiv d \mod m$. Then $ab \equiv cd \mod m$.*
Proof. Write $mk_1 = a-c$ and $mk_2 = b-d$ for some $k_1, k_2 \in \mathbb{Z}$. Observe

$$\begin{aligned} ab - cd &= ab - bc + bc - cd \\ &= b(a-c) + c(b-d) \\ &= bmk_1 + cmk_2 \\ &= m(bk_1 + ck_2). \end{aligned}$$

Thus, $m|ab-cd$ and $ab \equiv cd \mod m$. ■

Definition 1.4.7 Let S be a set and $\sim$ an equivalence relation on S. Then a statement P about the equivalence classes of S is **well-defined** if the representative of the equivalence class does not matter. That is, whenever $\overline{x} = \overline{y}$, $P(\overline{x}) = P(\overline{y})$. ◊

The previous exercises justify the following definitions.

Definition 1.4.8 Let $m > 1$ and $a, b \in \mathbb{Z}_m$. Then the following are well-defined operations on the equivalence classes:

1. *Addition modulo m*: $\overline{a} + \overline{b} := \overline{a+b}$.

2. *Multiplication modulo m*: $\overline{a} \cdot \overline{b} := \overline{a \cdot b}$.

$\Diamond$

Most elementary propositions about $\mathbb{Z}_m$ can be recast as statements about $\mathbb{Z}$. For instance, in proving Theorem 1.4.5 you likely proved that if $m|a-c$ and $m|b-d$ that $m|(a+b)-(c+d)$. However, as the statements become more complex, repeatedly reshaping statements about Z_m as statements about $\mathbb{Z}$ becomes cumbersome and unhelpful. Instead, you are encouraged to become comfortable doing arithmetic modulo m or, put another way, arithmetic with the equivalence classes of $\mathbb{Z}_m$ as defined in Definition 1.4.8.

Activity 1.4.6 Without passing back to $\mathbb{Z}$, find the smallest nonnegative integer representative of the resulting equivalence classes.

1. $\overline{5}+\overline{11}$ in $\mathbb{Z}_9$
2. $\overline{-3}+\overline{-3}$ in $\mathbb{Z}_6$
3. $\overline{8}\cdot\overline{3}$ in $\mathbb{Z}_{19}$
4. $\overline{-1}\cdot(\overline{3}+\overline{8})$ in $\mathbb{Z}_7$
5. $\overline{3}\cdot(\overline{5}^2+\overline{3}^3)$ in $\mathbb{Z}_{20}$

Solution. TBD

In the remainder of this section, we investigate fundamental properties of arithmetic in $\mathbb{Z}_m$.

Investigation 1.4.7 Let $\overline{a},\overline{b}\in\mathbb{Z}_m$ and $m>1$. If $\overline{a}\cdot\overline{c}=\overline{b}\cdot\overline{c}$, is it true that $\overline{a}=\overline{b}$? If so, prove it. If not, find an example of when the statement fails to hold.

Answer. It is not true. For example, in $\mathbb{Z}_{12}$, $\overline{6}\cdot\overline{3}=\overline{6}\cdot\overline{1}$, but $\overline{3}\neq\overline{1}$.

Theorem 1.4.9 *Let a,b,c, and m be integers with $m>1$ and $\gcd(c,m)=1$. Then there is some $x\in\mathbb{Z}$ such that $\overline{cx}=\overline{1}$.*

Conclude that if $\overline{a}\cdot\overline{c}=\overline{b}\cdot\overline{c}$ in $\mathbb{Z}_m$ that $\overline{a}=\overline{b}$.

Proof. We know that $m|ac-bc$, i.e., that $m|c(a-b)$. By Theorem 1.2.10, $m|a-b$. ■

Theorem 1.4.10 *Let $p\in\mathbb{N}$ be prime and $\overline{a},\overline{b},\overline{c}\in\mathbb{Z}_p$ such that $\overline{c}\neq\overline{0}$. Then*

1. *there is some $\overline{x}\in\mathbb{Z}_p$ such that $\overline{c}\cdot\overline{x}=\overline{1}$; and,*
2. *if $\overline{a}\cdot\overline{c}=\overline{b}\cdot\overline{c}$, $\overline{a}=\overline{b}$.*

Proof. In $\mathbb{Z}_p$, every nonzero equivalence class is represented by an x for which $\gcd(x,p)=1$. Apply Theorem 1.4.9. ■

Chapter 2

Fields and Rings

You have been exploring numbers and the patterns they hide within them since your earliest school days. In Chapter 1 we reminded ourselves about some of those patterns (with the goal of understanding factorization) and worked to express them in a more formal way. You may find yourself wondering why we are going out of our way to complicate ideas you have understood since elementary school. The reason for the abstraction (and the reason for this course!) is so that we can explore just how far we can push these patterns. How far does our understanding of factorization in the integers stretch to other types of numbers and other mathematical objects (like polynomials)? In this chapter we will set the ground work for answering that question by introducing ideas that will assist us in streamlining our investigation into factorization.

2.1 Fields

Guiding Questions.

In this section, we'll seek to answer the questions:

- What are binary operations?
- What is a field? What sorts of things can one do in a field?
- What are examples of fields?

We now begin the process of abstraction. We will do this in stages, beginning with the concept of a *field*. First, we need to formally define some familiar sets of numbers.

Definition 2.1.1 The rational numbers, denoted by $\mathbb{Q}$, is the set

$$\mathbb{Q} = \left\{ \frac{a}{b} : \ a, b \in \mathbb{Z}, \ b \neq 0 \right\}.$$

$\Diamond$

Recall that in elementary school, you learned that two fractions $\frac{a}{b}, \frac{c}{d} \in \mathbb{Q}$ are equivalent if and only if $ad = bc$.

Activity 2.1.1 Prove that our elementary school definition of equivalent fractions is an equivalence relation. Recall Definition 1.4.2.

We likely have an intuitive idea of what is meant by $\mathbb{R}$, the set of real numbers. Defining $\mathbb{R}$ rigorously is actually quite difficult, and occupies a significant amount of time in a first course in real analysis. Thus, we will make use of your intuition.

Out of $\mathbb{R}$ we may build the complex numbers.

Definition 2.1.2 The **complex numbers** consist of all expressions of the form $a + bi$, where $a, b \in \mathbb{R}$ and $i^2 = -1$. Given $z = a + bi$, we say a is the **real part** of z and b is the **imaginary part**. The set of complex numbers is denoted $\mathbb{C}$. $\Diamond$

As was mentioned in the Introduction, *algebra* comes from an Arabic word meaning "the reunion of broken parts". We therefore need a way of combining two elements of a set into one; we turn to a particular type of function, known as a binary operation, to accomplish this.

Definition 2.1.3 Let X be a nonempty set. A function $\star : X \times X \to X$ is called a **binary operation**. If $\star$ is a binary operation on X, we say that X is **closed under the operation** $\star$. [Given $a, b \in X$, we usually write $a \star b$ in place of the typical function notation, $\star(a, b)$.] $\Diamond$

Investigation 2.1.2 Which of $+, -, \cdot, \div$ are binary operations:

1. on $\mathbb{R}$?
2. on $\mathbb{Q}$?
3. on $\mathbb{Z}$?
4. on $\mathbb{N}$?
5. on $\mathbb{C}$? (Recall that for $a_1 + b_1 i, a_2 + b_2 i \in \mathbb{C}$, $(a_1 + b_1 i) + (a_2 + b_2 i) := (a_1 + a_2) + (b_1 + b_2)i$ and $(a_1 + b_1 i)(a_2 + b_2 i) := (a_1 a_2 - b_1 b_2) + (a_1 b_2 + b_1 a_2)i$.)

Answer. Division is never a binary operation. The others are binary operations on $\mathbb{C}, \mathbb{R}, \mathbb{Q}$, and $\mathbb{Z}$. The only binary operation on $\mathbb{N}$ is addition.

Activity 2.1.3 Choose your favorite nonempty set X and describe a binary operation different than those in Investigation 2.1.2.

Answer. Answers vary, but one option is to define $\min : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ by

$$\min(a, b) := \begin{cases} a & \text{if } a \leqslant b \\ b \end{cases}$$

The hallmark of modern pure mathematics is the use of *axioms*. An axiom is essentially an unproved assertion of truth. Our use of axioms serves several purposes.

From a logical perspective, axioms help us avoid the problem of infinite regression (e.g., asking *How do you know?* over and over again). That is, axioms give us very clear starting points from which to make our deductions.

To that end, our first abstract algebraic structure captures and axiomatizes familiar behavior about how numbers can be combined to produce other numbers of the same type.

Definition 2.1.4 A **field** is a nonempty set F with at least two elements and binary operations $+$ and $\cdot$, denoted $(F, +, \cdot)$, and satisfying the following **field axioms**:

1. Given any $a, b, c \in F$, $(a + b) + c = a + (b + c)$.
2. Given any $a, b \in F$, $a + b = b + a$.
3. There exists an element $0_F \in F$ such that for all $a \in F$, $a+0_F = 0_F+a = a$.
4. Given any $a \in F$ there exists a $b \in F$ such that $a + b = b + a = 0_F$.
5. Given any $a, b, c \in F$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
6. Given any $a, b \in F$, $a \cdot b = b \cdot a$.
7. There exists an element $1_F \in F$ such that for all $a \in F$, $1_F \cdot a = a \cdot 1_F = a$.
8. For all $a \in F$, $a \neq 0_F$, there exists a $b \in F$ such that $a \cdot b = b \cdot a = 1_F$.
9. For all $a, b, c \in F$, $a \cdot (b + c) = a \cdot b + a \cdot c$.
10. For all $a, b, c \in F$, $(a + b) \cdot c = a \cdot c + b \cdot c$.

$\Diamond$

We will usually write $a \cdot b$ as ab. Additionally, we will usually drop the

subscripts on $0, 1$ unless we need to distinguish between fundamentally different identities in different fields.

Investigation 2.1.4 Which of the following are fields under the specified operations? For most, a short justification or counterexample is sufficient.

1. $\mathbb{N}$ under the usual addition and multiplication operations
2. $\mathbb{Z}$ under the usual addition and multiplication operations
3. $2\mathbb{Z}$, the set of even integers, under the usual addition and multiplication operations
4. $\mathbb{Q}$ under the usual addition and multiplication operations
5. $\mathbb{Z}_6$ under addition and multiplication modulo 10
6. $\mathbb{Z}_5$ under addition and multiplication modulo 11
7. $\mathbb{R}$ under the usual addition and multiplication operations
8. $\mathbb{C}$ under the complex addition and multiplication defined in Investigation 2.1.2
9. $\mathcal{M}_2(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R} \right\}$[1], the set of 2×2 matrices with real coefficients using the usual definition of matrix multiplication[2] and matrix addition.

Solution.

1. $\mathbb{N}$ is not closed under taking additive inverses, so is not a field.
2. $\mathbb{Z}$ is not a field, as there is no integer a such that $2a = 1$.
3. $2\mathbb{Z}$ is not a field for the same reason.
4. $\mathbb{Q}$ is a field.
5. $\mathbb{Z}_6$ is not a field; there is no $\overline{a} \in \mathbb{Z}_{10}$ for which $\overline{2} \cdot \overline{a} = \overline{1}$ (you can check them all; there are only 4 viable options).
6. $\mathbb{Z}_5$ *is* a field. The axioms pertaining to addition and multiplication were established earlier or rely on the same axioms holding for $\mathbb{Z}$.
7. $\mathbb{R}$ is a field

8. $\mathbb{C}$ is a field. Given $a+bi \neq 0$, $(a+bi)^{-1} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i$.

9. $\mathcal{M}_2(\mathbb{R})$ is not a field. Matrix multiplication is not commutative.

In the Investigation 2.1.4, you determined which of sets of familiar mathematical objects are and are not fields. Notice that you have been working with fields for years and that our abstraction of language to that of fields is simply to allow us to explore the common features at the same time - it is inefficient to prove the same statement about every single field when we can prove it once and for all about fields in general.

Theorem 2.1.5 Properties of Fields. *Let F be a field.*

1. *The additive identity 0 is unique.*

2. *For all $a \in F$, $a \cdot 0 = 0 \cdot a = 0$.*

3. *Additive inverses are unique.*

4. *The multiplicative identity 1 is unique.*

5. *Multiplicative inverses are unique.*

6. $(-1)\cdot(-1) = 1$ ***Hint.*** *Note that we are saying that the additive inverse of the multiplicative identity times itself equals the multiplicative identity. You should use only the field axioms and the properties previously established in this theorem.*

Proof.

1. Assume $0, 0'$ both satisfy the additive identity axiom. Observe that $0 = 0 + 0' = 0'$.

2. As $0 = 0 + 0$, we may write $a \cdot 0 = a \cdot (0+0) = a \cdot 0 + a \cdot 0$. Now add the additive inverse of $a \cdot 0$ to both sides to obtain $0 = a \cdot 0 = 0 \cdot a$.

3. Let $a \in F$ and suppose b and c are such that $a + c = 0$ and $a + b = 0$. Then $a + c = a + b$, and we may add b (or c) to both sides to obtain $b+(a+b) = (b+a)+b = 0+b = b$ and $c+(a+c) = (c+a)+c = 0+c = c$.

[1] For students who have taken a linear algebra course.

[2] Recall that, if $\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}, \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} \in \mathcal{M}_2(\mathbb{R})$, then $\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \cdot \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = \begin{pmatrix} a_1a_2 + b_1c_2 & a_1b_2 + b_1d_2 \\ c_1a_2 + d_1c_2 & c_1b_2 + d_1d_2 \end{pmatrix}$.

Thus $b = c$.

4. Suppose 1 and $1'$ are multiplicative identities. Then $1 = 1 \cdot 1' = 1'$.

5. Let $a \in F$ be nonzero and suppose b, c are multiplicative inverses for a. Then $ab = 1 = ac$, and we may multiply by b (or c) to obtain $b(ab) = (ba)b = 1 \cdot b = b = b(ac) = (ba)c = 1 \cdot c = c$.

6. Observe that $(-1)+1 = 0$ by the definition of additive inverses. Multiply both sides by -1 and distribute to obtain $(-1)(-1) + (-1) \cdot 1 = (-1) \cdot 0$. By part 2 of this theorem and the definition of a multiplicative identity, we obtain $(-1)(-1) + (-1) = 0$. We now add the multiplicative identity to obtain $(-1)(-1) + (-1) + 1 = 0 + 1$, which simplifies to $(-1)(-1) = 1$.

■

One consequence of Theorem 2.1.5 is that, given $a \in F$, $b \in F \setminus \{0\}$, we may refer to $-a$ as *the* additive inverse of a, and b^{-1} as *the* multiplicative inverse of b. We will thus employ this familiar terminology henceforth.

Investigation 2.1.5 For which $n > 1$ is $\mathbb{Z}_n$ a field? Compute some examples, form a conjecture, and prove your conjecture.

2.2 Rings

Guiding Questions.

In this section, we'll seek to answer the questions:

- What are rings and integral domains, and how do they relate to fields?
- What are subrings, and how can we tell if a given subset of a ring is a subring?
- What special types of elements do rings have?

In the previous section, we observed that many familiar number systems are fields but that some are not. As we will see, these non-fields are often more structurally interesting, at least from the perspective of factorization; thus, in this section, we explore them in more detail. Before we proceed with that endeavor we will give a formal definition of polynomial so that we can include it in our work.

Definition 2.2.1 Let A be a set with a well-defined addition operation $+$ and additive identity 0, and x a variable. We define a **polynomial in x with coefficients in A** to be an expression of the form

$$p = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n,$$

where $a_n \neq 0$. We call $n \in \mathbb{N}_0$ the **degree** of the polynomial p, denoted $\deg(p) = n$, and $a_0, a_1, \ldots, a_n$ the **coefficients** of the polynomial. The coefficient a_n is known as the **leading coefficient** of p, and a_nx^n is the **leading term** of p. By

$$A[x] := \{a_0 + a_1x + a_2x + \cdots + a_nx^n : \ n \in \mathbb{N}_0, \ a_i \in A\}$$

we denote the set of all polynomials with coefficients in A. The additive identity of $A[x]$ is 0, called the **zero polynomial**, and is the polynomial whose coefficients are all 0. The degree of the zero polynomial is $-\infty$. $\Diamond$

Exploration 2.2.1 Give some examples of polynomials in $A[x]$ for various choices of number systems A. Identify their coefficients, leading terms, and degrees.

Exploration 2.2.2 In the following table, fill in a Y if the set has the property; fill in a N if it does not.

Table 2.2.2 A list of properties and sets.

	$\mathbb{N}$	$\mathbb{Z}$	$2\mathbb{Z}$	$\mathbb{Q}$	$\mathbb{Q}[x]$	$\mathbb{Z}_8$	$\mathbb{Z}_2$	$\mathbb{R}$	$\mathbb{C}$	$\mathcal{M}_2(\mathbb{R})$
Closure under $+$										
Closure under $\cdot$										
$+$ is associative										
$\cdot$ is associative										
$+$ is commutative										
$\cdot$ is commutative										
$\cdot$ distributes over $+$										
There is an additive identity										
All elements have additive inverses										
There is an multiplicative identity										
All nonzero elements have mult. inverses										

Exploration 2.2.3 Which of the field axioms in Definition 2.1.4 hold for $F[x]$, where F is a field, and which fail to hold in general?

Answer. All the axioms hold, except F8. For instance, there is no polynomial $f(x) \in \mathbb{Q}[x]$ for which $xf(x) = 1$.

As a result of the answer to Exploration 2.2.3 and the completed Table 2.2.2, we make the following definition.

Definition 2.2.3 A **ring** R is a nonempty set, together with binary operations $+$ and $\cdot$, denoted $(R, +, \cdot)$, and satisfying the following axioms.

1. Given any $a, b, c \in R$, $(a + b) + c = a + (b + c)$.
2. Given any $a, b \in R$, $a + b = b + a$.
3. There exists an element $0_R \in R$ such that for all $a \in R$, $a+0_R = 0_R+a = a$.
4. Given any $a \in R$ there exists a $b \in R$ such that $a + b = b + a = 0_R$.
5. Given any $a, b, c \in R$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
6. For all $a, b, c \in R$, $a \cdot (b + c) = a \cdot b + a \cdot c$.
7. For all $a, b, c \in R$, $(a + b) \cdot c = a \cdot c + b \cdot c$.

As with fields, when the ring R is clear from context, we will often write 0 in place of 0_R. ◊

Investigation 2.2.4 Compare and contrast Definitions 2.1.4 and Definition 2.2.3. What are the similarities? What are the differences?

While rings do not enjoy all the properties of fields, they are incredibly useful even in applied mathematics (see, e.g., [2.2.1] for one recent example).

Definition 2.2.4 A ring R is said to be **commutative** if, for all $a, b \in R$, $ab = ba$. Additionally, R is said to have a **unity** or **multiplicative identity** if there is an element $1_R \in R$ such that for all $a \in R$, $a \cdot 1_R = 1_R \cdot a = a$. ◊

If R is noncommutative, it may have a left (respectively, right) identity, i.e., an element $e \in R$ such that for all $r \in R$, $er = r$ (respectively, $re = r$). If R has an element e for which $er = re = r$ for all $r \in R$, e is often called a two-sided identity. In short, noncommutative rings may have left, right, or two-sided identities (or none at all).

Exploration 2.2.5 Consider the sets given in Table 2.2.2. Which are rings? Which are commutative rings with identity?

Exploration 2.2.6 Which properties of fields in Theorem 2.1.5 hold for (commutative) rings?

Investigation 2.2.7 Are all rings fields? Are all fields rings? Justify.

Investigation 2.2.8 Most familiar rings are commutative, though not all. Most familiar (commutative) rings have identities, but not all. Find:

1. A ring that does not have an identity[1].

2. A noncommutative ring that *does* have an (two-sided) identity.

Solution.

1. $2\mathbb{Z}$

2. $\mathcal{M}_2(\mathbb{R})$; the 2×2 identity matrix is a two-sided identity.

In the 1920s, Emmy Noether was the first to explicitly describe the ring axioms as we know them today, and her definition of a (not-necessarily-commutative) ring has led to a great deal of interesting work in algebra, number theory, and geometry, including the (see Section 3.3 for more on the historical development of the proof of Fermat's Last Theorem). Most modern definitions of *ring* agree with our Definition 2.2.3 and allow for rings with noncommutative multiplication and no multiplicative identity.

The following theorem states that the set of polynomials with coefficients in a ring R is itself a ring under the usual operations of polynomial addition of like terms, and multiplication via distribution. The proof is not tricky, but a rigorous justification (especially of, e.g., the associativity of polynomial multiplication) is tedious, and thus is omitted.

Theorem.

If R is a (commutative) ring (with identity 1_R), then $R[x]$ is a (commutative) ring (with identity $1_{R[x]} = 1_R$).

One of the ways to better understand mathematical structures is to understand their similar substructures (e.g., given a vector space $V \subseteq \mathbb{R}^n$ and a subspace $W \subseteq V$, we may write $V = W + W^{\perp}$).

Definition 2.2.5 Let $(R, +, \cdot)$ be a ring and let $S \subseteq R$. If S is itself a ring under $+$ and $\cdot$, we say S is a **subring** of R. In this case, R is often called an **overring** of S. ◊

The following theorem provides a easy-to-apply test to check if a given subset S of a ring R is in fact a subring of R.

[1] Sometimes called a *rng*. ☺

Theorem 2.2.6 *Let R be a ring and S a subset of R. Then S is a subring if and only if:*

1. *$S \neq \emptyset$;*
2. *S is closed under multiplication; and*
3. *S is closed under subtraction.*

Proof. TBD. ■

Activity 2.2.9 Determine whether the following rings S are subrings of the given rings R.

1. $S = \mathbb{Z}$, $R = \mathbb{Q}$
2. $S = \mathbb{Z}_5$, $R = \mathbb{Z}_7$
3. S is any ring, $R = S[x]$
4. $S = \mathbb{R}$, $R = \mathbb{C}$

Answer.

In our study of rings, we are primarily interested in special types of subrings known as *ideals*, to be studied in more depth in Chapter 4.

Definition 2.2.7 Let R be a ring and let $u \in R$ be nonzero. If there is a $v \in R$ such that $uv = vu = 1$, we say u is **unit** of R. We denote the set of units of R by $R^\times$. We say $x, y \in R$ are **associates** if there exists some $u \in R^\times$ such that $x = uy$. ◊

Exploration 2.2.10 Explicitly describe the set $\mathbb{Z}^\times$. What are the associates of 7 in $\mathbb{Z}$?

In other words, a unit in a ring is a nonzero element with a multiplicative inverse. The existence of units is the primary difference between fields and commutative rings with identity: in a field, all nonzero elements are units, while in a commutative ring with identity, no nonzero elements need be units, as Theorem 2.2.8 demonstrates.

Theorem 2.2.8 *A commutative ring with identity R in which every nonzero element is a unit is a field.*

Proof. Compare the axioms for a commutative ring with identity and a field. The only thing missing from the ring axioms is the existence of multiplicative inverses for nonzero elements. ■

A useful tool for analyzing the structure of rings with finitely many elements are addition and multiplication tables. As an example, consider the addition and multiplication tables for $R = \mathbb{Z}_3$ shown in [title] **??** and Table 2.2.10.

Table 2.2.9 Addition table for $R = \mathbb{Z}_3$.

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{0}$	$\overline{1}$

Table 2.2.10 Multiplication table for $R = \mathbb{Z}_3$.

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{1}$

Investigation 2.2.11 Calculate addition and multiplication tables for the following rings.

1. $R = \mathbb{Z}_5$
2. $R = \mathbb{Z}_6$

List 2-3 observations about your tables.

Solution. TBD.

One of the interesting side effects of our definition of *ring* is that it allows for behavior that may at first appear unintuitive or downright weird.

Definition 2.2.11 A **zero divisor** in a ring R is a nonzero element $z \in R$ such that there is a nonzero $x \in R$ with $xz = zx = 0$. $\Diamond$

Notice that the reason the idea of zero divisors at first appears weird is that they are not something we encounter when working with our familiar sets of numbers, such as $\mathbb{Z}$ or $\mathbb{R}$. In fact, we specifically use the fact that there are no zero divisors in our familiar numbers systems to solve equations in high school algebra (e.g., if $(x-2)(x+5) = 0$, then $(x-2) = 0$ or $x+5 = 0$). The lack of zero divisors is one of the properties that does not persist in our abstraction from the integers to rings in general.

Exploration 2.2.12 Find, with justification, all of the zero divisors in $\mathbb{Z}_{10}$ and $\mathbb{Z}_{11}$. Make and prove a conjecture about the existence of zero divisors in

$\mathbb{Z}_m$, where $m > 1$.

Solution. The zero divisors in $\mathbb{Z}_{10}$ are $\overline{2}, \overline{4}, \overline{5}, \overline{6}, \overline{8}$. There are no zero divisors in $\mathbb{Z}_{11}$.

conjecture. $\overline{x} \in \mathbb{Z}_m$ is a zero divisor if and only if $\gcd(x, m) \neq 1$.

Investigation 2.2.13 Are there any other rings in which you've seen zero divisors? Recall your answers to Exploration 2.2.5.

Answer. Matrix rings, if students have had linear algebra (and/or completed Table 2.2.2). Otherwise, this may be a new concept.

Theorem 2.2.12 *Let R be a ring and suppose $a, b \in R$ such that ab is a zero divisor. Then either a or b is a zero divisor.*

Proof. Let $a, b \in R$ such that ab is a zero divisor. Then $a, b \neq 0$ (else $ab = 0$). Since ab is a zero divisor, there is some $c \neq 0$ such that $(ab)c = 0$. If $bc \neq 0$, then a is a zero divisor, as $a(bc) = 0$. On the other hand, if $bc = 0$, then b is a zero divisor, as $b, c \neq 0$. ■

Theorem 2.2.13 *Let R be a ring and $u \in R^\times$. Then u is not a zero divisor.*

Proof. Let $u \in R^\times$ and suppose u is a zero divisor. Then there is some $v \neq 0$ such that $uv = 0$. But then $0_R = u^{-1}0_R = u^{-1}(uv) = (u^{-1}u)v = 1_R v = v$. \Lightning ■

Investigation 2.2.14 How can we reinterpret Investigation 1.4.7 in light of our new language of units and zero divisors? State a theorem that uses this new language.

Answer. Answers may vary, but how about this: Let $\overline{a}, \overline{b}, \overline{c} \in \mathbb{Z}_m$ such that $\overline{a} \cdot \overline{c} = \overline{b} \cdot \overline{c}$. Then $\overline{a} = \overline{b}$ if $\overline{c}$ is not a zero divisor.

While there is a well-developed body of literature on (noncommutative) rings (possibly without identity), from this point on, and unless stated otherwise, when we use the word *ring* we mean *commutative ring with identity*.

Moreover, while even commutative rings with identity and zero divisors are of interest to mathematicians, we will focus our study on rings with no zero divisors. As these rings share many properties of the integers, they are known as **integral domains**.

Definition 2.2.14 A commutative ring with identity R is an **integral domain**, or just **domain**, if R has no zero divisors. ◊

The next activities and theorems help us identify examples of domains, as well as situate the notion of a domain in its proper place relative to fields and rings in general.

Activity 2.2.15 Which of the following rings are domains? Justify your answers.

1. $\mathbb{Z}$
2. $\mathbb{Z}_8$
3. $\mathbb{Z}_{19}$
4. $\mathbb{R}$
5. $\mathbb{Q}[x]$

Theorem 2.2.15 *Every field is a domain.*
Proof. If F is a field, the nonzero elements of F are units, which cannot be zero divisors. Thus, F has no zero divisors. ■

Theorem 2.2.16 *Let $m > 1$ and $R = \mathbb{Z}_m$. Then R is a field if and only if R is a domain.*
Proof. The forward direction holds by Theorem 2.2.15.

For the reverse, assume R is a domain. Then R has no zero divisors. If m is composite, there exist integers a, b satisfying $1 < a, b < m$ such that $m = ab$. Then $\overline{a}, \overline{b} \neq \overline{0}$ in $\mathbb{Z}_m$, but $\overline{a} \cdot \overline{b} = \overline{0}$. Thus, m may not be composite, and is therefore prime. By an earlier theorem, $\mathbb{Z}_p$ is a field. ■

Theorem 2.2.17 *If R is a domain and S is a subring of R, then S is a domain.*
Proof. Any zero divisors in S are also zero divisors in R. Since R has no zero divisors, neither does S. ■

Theorem 2.2.18 *If R is a domain, then so is $R[x]$.*
Proof. Let $f(x) = a_i x^i + a_{i+1}x^{i+1} + \cdots a_n x^n$ and $g(x) = b_j x^j + b_{j+1}x^{j+1} + \cdots b_k x^k$ be nonzero polynomials in $R[x]$, where $a_i, b_j \neq 0$. Then the lowest-degree term in $f(x)g(x)$ is $a_i b_j x^{i+j}$. Since R is a domain, $a_i b_j \neq 0$, and thus $f(x)g(x)$ is not the zero polynomial. ■

Investigation 2.2.16 Is the converse of Theorem 2.2.18 true? If so, give a short proof. If not, find a counterexample.

Answer. Yes. Apply Theorem 2.2.17.

Corollary 2.2.19 *Given a field F, the set of polynomials $F[x]$ is a domain.*

When considering sets of polynomials, as we do in Chapter 3 (particularly in Section 3.1), the following results will be quite useful.

Theorem 2.2.20 *Let R be a domain, and let $p(x), q(x) \in R[x]$ be nonzero polynomials. Then* $\deg(p(x)q(x)) = \deg(p(x)) + \deg(q(x))$.

Proof. Let the leading term of $p(x)$ be $a_n x^n$ and the leading term of $q(x)$ be $b_m x^m$. Then the leading term of $p(x)q(x)$ is $a_n b_m x^{n+m}$. (Observe that since R is a domain, $a_n b_m \neq 0$.) Thus, $\deg(p(x)q(x)) = n+m = \deg(p(x)) + \deg(q(x))$. ■

Exploration 2.2.17 Can the hypotheses of Theorem 2.2.20 be relaxed? If so, provide more general hypotheses and adapt the proof. If not, give an illustrative example.

Solution. No; let $p = \overline{1} + \overline{2}x, q = \overline{2} + \overline{3}x \in \mathbb{Z}_6[x]$. Observe that $\deg(p) + \deg(q) = 1 + 1$, but $pq = \overline{2} + \overline{5}x$ has degree 1.

Investigation 2.2.18 Let R be a domain. What are the units of $R[x]$? Prove your answer.

Answer. The units are $R^\times$. Clearly, $R^\times \subseteq R[x]^\times$.

Suppose $p(x), q(x) \in R[x]^\times$ such that $p(x)q(x) = 1$. Then $\deg p(x) + \deg q(x) = \deg(p(x)q(x)) = \deg(1) = 0$. Thus $\deg p(x) = \deg q(x) = 0$, and consequently, $p(x), q(x) \in R^\times$.

References

C. Curto, V. Itskov, A. Veliz-Cuba, N. Youngs, *The Neural Ring: An Algebraic Tool for Analyzing the Intrinsic Structure of Neural Codes*, Bull. Math. Bio. 75 (2013), 1571-1611, DOI 10.1007/s11538-013-9860-3

2.3 Divisibility in Integral Domains

Guiding Questions.

In this section, we'll seek to answer the questions:

- What multiplicative properties can we generalize from $\mathbb{Z}$ to any integral domain?
- What are the differences between a prime and irreducible element in a commutative ring?

When we introduced the notion of integral domain, we said that part of the reason for the definition was to capture some of the most essential properties of the integers. This is the heart of abstraction and generalization in mathematics: to distill the important properties of our objects of interest and explore the consequences of *those* properties. One such important property of

$\mathbb{Z}$ is *cancellation*.

Theorem 2.3.1 *Let R be a ring. Then R is a domain if and only if for all $a, b, c \in R$ with $c \neq 0$ and $ac = bc$, we have $a = b$.*

Proof. Assume R is a domain and $ac = bc$. Then $ac - bc = 0$, so $c(a - b) = 0$. Since R is a domain, it has no zero divisors, and therefore either $c = 0$ or $a - b = 0$. The first possibility is ruled out by our assumptions on a, b, c, so we must have $a - b = 0$, or $a = b$.

Conversely, assume that R is not a domain. Let $z \in R$ be a zero divisor; then there is a nonzero $x \in R$ such that $xz = 0 = 0z$. This implies that $x = 0$, a contradiction. ■

We may read Theorem 2.3.1 as saying that the defining property of an integral domain is the ability to cancel common nonzero factors. Note that we have not *divided*; division is not a binary operation, and nonzero elements of rings need not be units. However, as was the case in $\mathbb{Z}$, there are notions of *divisibility* and *factorization* in rings.

Definition 2.3.2 Let R be a commutative ring with identity, and let $a, b \in R$. We say a **divides** b and write $a \mid b$ if there is a $c \in R$ such that $ac = b$. We then say that a is a **factor** of b. ◊

Investigation 2.3.1 Find all factors of $\overline{2}$ in the following rings:

1. $\mathbb{Z}_5$
2. $\mathbb{Z}_6$
3. $\mathbb{Z}_{10}$

Solution. TBD.

Our definition of prime also extends nicely to domains. Indeed, the desire to extend the familiar notion of prime from $\mathbb{Z}$ to any ring is the reason for our less-familiar definition given in Definition 1.3.1.

Definition 2.3.3 Let R be a domain. We say a nonzero nonunit element $a \in R$ is **prime** if whenever $a \mid bc$ for some $b, c \in R$, either $a|b$ or $a|c$. ◊

A notion related to primality is irreducibility. In fact, one might reasonably say that irreducibility is the natural generalization of the typical definition of prime one encounters in school mathematics.

Definition 2.3.4 Let R be a domain. We say a nonzero nonunit element $a \in R$ is **irreducible** if whenever $a = bc$ for some $b, c \in R$, one of b or c is a unit. (Note that in some areas of the literature, the word **atom** is used interchangeably with irreducible.) ◊

Exploration 2.3.2 Find the units, primes, and irreducibles in the following rings.

1. $\mathbb{R}$
2. $\mathbb{Z}$
3. $\mathbb{Z}_5$
4. $\mathbb{Z}_6$

Solution. TBD.

In domains, all primes are irreducible.

Theorem 2.3.5 *Let R be a domain. If $a \in R$ is prime, then a is irreducible.*
Proof. Compare to the proof of Theorem 1.3.2.

Let $a \in R$ be prime, and suppose that $a = bc$ for some $b, c \in R$. Then $a|bc$, so by definition either $a|b$ or $a|c$. Without loss of generality, assume $a|b$. Then there is a $k \in R$ such that $ak = b$, so $a = (ak)c$, and we may cancel a to obtain $1_R = kc$. Thus, c is a unit, making a irreducible. ■

In familiar settings, the notion of prime and irreducible exactly coincide.

Theorem 2.3.6 *Every irreducible in $\mathbb{Z}$ is prime.*
Proof. Let $p \in \mathbb{Z}$ be irreducible, and suppose that $d|p$. Then $p = de$. Since p is irreducible, either d or e is a unit. However, the only units are ± 1, so either one of d or e is p or $-p$. In either case, p is prime. ■

Despite their overlap in familiar settings, primes and irreducibles are distinct types of elements. As the next exploration demonstrates, not all primes are irreducible. What is more, Exploration 2.3.4 will show that not all irreducibles are primes, even in domains!

Exploration 2.3.3 Find an example of a ring R and prime $p \in R$ such that p is not irreducible.

Solution. In $R = \mathbb{Z}_6$, $p = \overline{3}$ is prime by Exploration 2.3.2 but not irreducible as $\overline{3}^2 = \overline{3}$ and $\overline{3}$ is not a unit.

Exploration 2.3.4 Consider the set R of all polynomials in $\mathbb{Z}[x]$ for which the coefficient on the linear term is zero. That is,

$$R = \left\{a_0 + a_2x^2 + \cdots + a_{n-1}x^{n-1} + a_nx^n : \ a_i \in \mathbb{Z}, \ n \in \mathbb{N}\right\}.$$

(You should convince yourself that R is an integral domain, but do not need to prove it.) Then, find a polynomial of the form x^n in R that is irreducible, but not prime.

Solution. Consider $f(x) = x^2$. Then $f(x)$ is irreducible, as it cannot be

factored into a product of linear polynomials (there aren't any in R), so any factorization of f is degree 2 times degree 0. Then the leading coefficients must be units, i.e., both 1 or both -1.

However, f is not prime, as $f|x^3 \cdot x^3$ but $f \nmid x^3$.

Our last straightforward generalization from the multiplicative structure of $\mathbb{Z}$ is the notion of greatest common divisor. As our next definition again demonstrates, our careful work in the context of $\mathbb{Z}$ generalizes nicely to all domains. Indeed, we intentionally did not appeal to $\leqslant$ to define the greatest common divisor in Definition 1.2.7, as not all rings have a natural order relation like $\mathbb{Z}$ does.

Definition 2.3.7 Let R be a domain, and let $a, b \in R$. A nonzero element $d \in R$ is a **greatest common divisor** of a and b if

1. $d \mid a$ and $d \mid b$ and,
2. if $e \in R$ with $e \mid a$ and $e \mid b$, then $e \mid d$.

$\Diamond$

Theorem 2.3.8 *Let R be a domain and $a, b \in R$ and suppose d is a greatest common divisor of a and b. Then any associate of d is also a greatest common divisor of a and b. (Recall Definition 2.2.7.)*

Proof. Let d be a gcd of a and b, let $u \in R^\times$, and $e = ud$. We claim e is also a gcd of a and b.

Since $d|a$ and $d|b$ there are $k_1, k_2 \in R$ such that $dk_1 = a$ and $dk_2 = b$. Then $e(u^{-1}k_1) = a$ and $e(u^{-1}k_2) = b$, so $e|a$ and $e|b$.

Let f be a common divisor of a and b. Since d is a gcd, $f|d$, i.e., $fk_3 = d$. Then $f(uk_3) = e$, so $f|e$.

Thus, e is a gcd of a and b. ■

Exploration 2.3.5 In most familiar domains, GCDs exist. However, they don't always! Find an example of elements in the ring from Exercise 2.3.4 which do not have a GCD. Justify your assertion.

Solution. Consider x^5 and x^6. First note that x^4 is not a common divisor in R.

Both x^5 and x^6 are divisible by x^3 and x^2 in R. However, neither can be the gcd, as $x^2 \nmid x^3$ and $x^3 \nmid x^2$.

Exploration 2.3.6 Fill in the following blanks in order of increasing generality with the words *ring*, *integral domain*, *field*, and *commutative ring*.

_____ ⇒ _____ ⇒ _____ ⇒ _____

2.4 Principal Ideals and Euclidean Domains

Guiding Questions.

In this section, we'll seek to answer the questions:

- What are principal ideals, and what are principal ideal domains?
- What are Euclidean domains, and how are they related to PIDs?

One of the ways in which mathematicians study the *structure* of an abstract object is by considering how it interacts with other (related) objects. This is especially true of its *sub*objects. Thus, in linear algebra, we are often concerned with *subspaces* of a vector space as a means of understanding the vector space, or even submatrices as a way of understanding a matrix (see, e.g., the cofactor expansion formula for the determinant). In real analysis and topology, the important subobjects are usually open sets, or subsequences, and the study of a graph's subgraphs is an important approach to many questions in graph theory.

In this section, we begin a set-theoretic structural exploration of the notion of ring by considering a particularly important class of subring which will be integral to our understanding of factorization.

These subrings are called *ideals*. They arose in the work of Kummer and Dedekind as a way of trying to recover some notion of unique factorization in rings that do not have properties like the fundamental theorem of arithmetic in $\mathbb{Z}$.

Definition 2.4.1 A subset I of a (not necessarily commutative) ring R is called an **ideal** if:

1. $0 \in I$
2. for all $x, y \in I$, $x + y \in I$; and,
3. for all $x \in I$ and for all $r \in R$, $xr \in I$ and $rx \in I$.

$\Diamond$

Observe that the third requirement for a set I to be an ideal of R is simplified slightly if R is commutative (which, we recall, all of our rings are).

There are many important examples and types of ideals, but there are also some trivial ideals contained in every ring.

Theorem 2.4.2 *Let R be a ring. Then R and $\{0\}$ are ideals of R.*

Theorem 2.4.3 *All ideals are subrings.*
Proof. It is straightforward to check that all conditions of Theorem 2.2.6 are satisfied. ■

The following theorem provides a useful characterization of when an ideal I is in fact the whole ring.

Theorem 2.4.4 *Let R be a ring and I an ideal of R. Then $I = R$ if and only I contains a unit of R.*
Proof. If $I = R$, $1 \in R$ will do.

If there is a unit $u \in I$, then given any $r \in R$, $r = ru^{-1}u = (ru^{-1})u \in I$, so $I = R$. ■

The most important type of ideals (for our work, at least), are those which are the sets of all multiples of a single element in the ring. Such ideals are called *principal ideals.*

Theorem 2.4.5 *Let R be commutative with identity and let $a \in R$. The set*

$$\langle a \rangle = \{ra : r \in R\}$$

is an ideal (called the principal ideal generated by *a).*
Proof. Observe that $0 = 0 \cdot a \in \langle a \rangle$.

Moreover, if $r_1a, r_2a \in \langle a \rangle$, $r_1a - r_2a = (r_1 - r_2)a \in \langle a \rangle$. Finally, if $x \in R$ and $ra \in \langle a \rangle$, $x(ra) = (xr)a \in \langle a \rangle$.

Thus, $\langle a \rangle$ is an ideal. ■

The element a in the theorem is known as a *generator* of $\langle a \rangle$.

Investigation 2.4.1 Let R be commutative with identity, and let $x, y, z \in R$. Give necessary and sufficient conditions for $z \in \langle x \rangle$ and, separately, $\langle x \rangle \subseteq \langle y \rangle$.

That is, fill in the blanks: "$z \in \langle x \rangle \Leftrightarrow$ _____" and "$\langle x \rangle \subseteq \langle y \rangle \Leftrightarrow$ _____."

Justify your answers.

Answer. We have "$z \in \langle x \rangle \Leftrightarrow x|z$" and "$\langle x \rangle \subseteq \langle y \rangle \Leftrightarrow y|x$"[1].

Note that $z \in \langle x \rangle \Leftrightarrow \exists r \in R, z = xr \Leftrightarrow x|z$.

Similarly, suppose $\langle x \rangle \subseteq \langle y \rangle$. Then $x \in \langle y \rangle$, so $y|x$. Conversely, if $y|x$, then there is some $r \in R$ such that $x = yr$, and thus for all $ax \in \langle x \rangle$, $ax = (ar)y \in \langle y \rangle$.

Note that this means that if we want to know if $\langle x \rangle \subseteq \langle y \rangle$, it's enough to check that $x \in \langle y \rangle$.

Principal ideals may have more than one generator.

[1] An acceptable alternative would be: $x \in \langle y \rangle$. Make sure students are aware of this!

Theorem 2.4.6 *Let R be a ring and $a \in R$. Then $\langle a \rangle = \langle ua \rangle$, where u is any unit of R.*
Proof. Apply the answer to the question. ■

Activity 2.4.2 In $R = \mathbb{Z}$, describe the principal ideals generated by

1. 2
2. -9
3. 9
4. 0
5. 27
6. 3

Determine the subset relations among the above ideals.

Solution.

1. All multiples of 2
2. All multiples of -9
3. All multiples of 9; same as the previous part.
4. $\{0\}$
5. All multiples of 27
6. All multiples of 3

We have $\langle 0 \rangle \subsetneq \langle 27 \rangle \subsetneq \langle -9 \rangle = \langle 9 \rangle \subsetneq \langle 3 \rangle$. The ideal $\langle 2 \rangle$ only contains $\langle 0 \rangle$, which is a subset of all ideals.

It is the case in many familiar settings that all ideals are principal. Such domains are given a special name.

Definition 2.4.7 An integral domain R in which every ideal is principal is known as a **principal ideal domain (PID)**. ◊

Theorem 2.4.8 *The ring $\mathbb{Z}$ is a principal ideal domain.* ***Hint****. Use properties specific to $\mathbb{Z}$, perhaps from Section 1.*
Proof. Let $I \subseteq Z$ be an ideal. If $I = \{0\}$, then $I = \langle 0 \rangle$, so suppose there is some nonzero $x \in I$. Define $S = \{m \in \mathbb{Z} : m > 0\}$. Note that $S \neq \emptyset$, as if $m \in I$, $(-1)m \in I$ also.

By WOP, S has a least element, call it d.

Claim: $I = \langle d \rangle$.

It is clear that $\langle d \rangle \subseteq I$. Now let $x \in I$ be nonzero, and write $x = dq + r$ using the division algorithm. Observe that $0 \leqslant r = x - dq < d$, but as $x \in I$ and $-dq \in I$, we must have $r \in I$. To avoid contradicting the WOP, we must have $r = 0$. Thus, $x = dq$ and $x \in \langle d \rangle$. ■

Activity 2.4.3 Find an integer d such that $I = \langle d \rangle \subseteq \mathbb{Z}$, if

1. $I = \{4x + 10y : x, y \in \mathbb{Z}\}$
2. $I = \{6s + 7t : s, t \in \mathbb{Z}\}$
3. $I = \{9w + 12z : w, z \in \mathbb{Z}\}$
4. $I = \{am + bn : m, n \in \mathbb{Z}\}$

You do not need to prove that each of the sets above are ideals (though you should make sure you can do it).

Solution. We see:

1. $I = \langle 2 \rangle$
2. $I = \langle 1 \rangle = \mathbb{Z}$
3. $I = \langle 3 \rangle$
4. $I = \langle \gcd(a, b) \rangle$

Theorem 2.4.9 *Let R be a principal ideal domain and $x, y \in R$ be not both zero. Let $I = \{xm + yn : m, n \in R\}$. Then:*

1. *I is an ideal, and*
2. *$I = \langle d \rangle$, where d is any greatest common divisor of x and y.*

We conclude that there exist $s, t \in R$ such that $d = xs + yt$.

Proof. Observe that $0 = x0 + y0 \in I$. Additionally, if $xm_1 + yn_1, xm_2 + yn_2 \in I$, then $(xm_1 + yn_1) + (xm_2 + yn_2) = x(m_1 + m_2) + y(n_1 + n_2) \in I$, and $r(xm_1 + yn_1) = x(rm_1) + y(rn_1) \in I$. Thus, I is an ideal.

Since R is a PID, there exists $d \in R$ such that $I = \langle d \rangle$. We claim that d is a GCD of x and y.

It is clear that $d|x$, as $x \cdot 1 + y \cdot 0 \in I = \langle d \rangle$. Similarly, $d|y$.

Now let $e \in R$ be a common divisor of x and y. We wish to show that $e|d$. Write $x = ek_1$ and $y = ek_2$. Since $d \in \langle d \rangle$, there exist $s, t \in R$ such that $d = xs + yt = (ek_1)s + (ek_2)t = e(k_1 s + k_2 t)$, and thus $e|d$.

In particular, there exist $s, t \in R$ such that a GCD d of x and y can be

written as $d = xs + yt$. ■

We have so far abstracted and axiomatized several important algebraic properties of $\mathbb{Z}$ that we discussed in § 1. In particular, we have our usual operations of addition and multiplication, and their interactions; we have notions of divisibility/factorization, irreducibility, and primality; we also have cancellation and greatest common divisors.

Our last major abstraction from $\mathbb{Z}$ is the division algorithm. The main obstacle to postulating domains with a division algorithm is a clear notion of comparison relations. That is, if R is an arbitrary domain with $r, s \in R$, is it possible to clearly and sensibly say which of r or s is "bigger" ? (Recall that this was a requirement for the division algorithm with nonzero remainders.) However, if there is a way to relate elements of a domain R to $\mathbb{N}_0$, we can sensibly define a division algorithm.

Definition 2.4.10 Let R be an integral domain. We call R a **Euclidean domain** if there is a function $\delta : R \setminus \{0\} \to \mathbb{N}_0$ such that:

1. If $a, b \in R \setminus \{0\}$, then $\delta(a) \leqslant \delta(ab)$.
2. If $a, b \in R$, $b \neq 0$, then there exist $q, r \in R$ such that $a = bq + r$, where either $r = 0$ or $\delta(r) < \delta(b)$.

We call the function δ a **norm** for R. ◊

Thus, a Euclidean domain is an integral domain with a division algorithm that behaves in a familiar way. In the remainder of this section, we will investigate the properties of Euclidean domains. First, we consider some examples.

Theorem 2.4.11 *The field $\mathbb{Q}$ is a Euclidean domain under ordinary addition and multiplication, with $\delta(x) = 0$ for all $x \in \mathbb{Q}$.*

Proof. TBD. ■

Investigation 2.4.4 Is $\mathbb{Z}$ a Euclidean domain? If so, what is the norm function δ, and why does this function have the required properties of a norm?

Answer. Yes. The norm is the absolute value function.

Lemma 2.4.12 *Let F be a field and $S \subseteq F[x]$ a set containing a nonzero polynomial. Prove that S contains a polynomial f such that $\deg(f) \leqslant \deg(g)$ for all nonzero $g \in S$.*

Proof. Let $T = \{\deg g : g \in S\}$. Since S contains a nonzero polynomial, $T \neq \emptyset$. By WOP, T contains a minimal element $d \geqslant 0$, which must be the degree of some polynomial in S. ■

Lemma 2.4.13 *Let F be a field and $f(x), g(x) \in F[x]$ with $g(x) \neq 0$. If $\deg f(x) \geqslant \deg g(x) > 0$, and $f(x) = a_0 + a_1x + \cdots + a_mx^m$ and $g(x) = b_0 + b_1x + \cdots + b_nx^n$, then $h(x) = f(x) - a_mb_n^{-1}x^{m-n}g(x)$ has degree strictly less than $\deg f(x)$.*

Proof. The leading term of $f(x)$ is a_mx^m, while the leading term of $a_mb_n^{-1}x^{m-n}g(x)$ is $a_mb_n^{-1}x^{m-n}(b_nx^n) = a_mx^m$. Thus, the leading term of h has degree less than m. ■

Theorem 2.4.14 (Polynomial Division Algorithm).

Let F be a field and $f(x), g(x) \in F[x]$ with $g(x) \neq 0$. Then there exist unique $q(x), r(x) \in F[x]$ such that

$$f(x) = g(x)q(x) + r(x),$$

where $\deg(r(x)) < \deg g(x)$. ***Hint.*** *For existence, consider three cases: $f(x) = 0$; $f(x) \neq 0$ and $\deg f < \deg g$; $f(x) \neq 0$ and $\deg f \geqslant \deg g$. In the last case, use induction on $m = \deg f(x)$. For uniqueness, mimic the uniqueness proof of Theorem 1.2.5.*

Proof. Existence: If $f(x) = 0$, then $q(x) = 0$ and $r(x) = 0$ will do. If $f(x) \neq 0$ and $\deg f < \deg g$, then $q(x) = 0$ and $r(x) = f(x)$ will suffice. Thus, we need only consider the case in which $f(x) \neq 0$ and $\deg f \geqslant \deg g$. We use induction on $\deg f = m$.

When $m = 0$, $\deg g \leqslant 0 = \deg f$, and as $g(x) \neq 0$, both f and g are nonzero constants. Then $r = 0$ and $q = fg^{-1}$ will work.

Now assume q and r exist whenever $\deg f < m$. Assume $\deg f = m$ and write $f(x) = a_0 + a_1x + \cdots + a_mx^m$ and $g(x) = b_0 + b_1x + \cdots + b_nx^n$. Use Lemma 2.4.13 and set $h(x) = f(x) - a_mb_n^{-1}x^{m-n}g(x)$, which must have degree less than f. Thus, by induction, there exist $q_1, r_1 \in F[x]$ such that $h = gq_1 + r_1$, with $r_1 = 0$ or $\deg r_1 < \deg g$.

We therefore have

$$\begin{aligned} f(x) &= a_mb_n^{-1}x^{m-n}g(x) + h(x) \\ &= a_mb_n^{-1}x^{m-n}g(x) + g(x)q_1(x) + r_1(x) \\ &= (a_mb_n^{-1}x^{m-n} + q_1(x))g(x) + r_1(x), \end{aligned}$$

where $q = a_mb_n^{-1}x^{m-n} + q_1(x)$ and $r = r_1(x)$ have the desired properties.

Uniqueness: Suppose $f = gq + r$ and $f = g\hat{q} + \hat{r}$, where $r, \hat{r}$ both have the desired properties. Then

$$0 = g[q - \hat{q}] + [r - \hat{r}],$$

or $\hat{r} - r = g[q - \hat{q}]$. Thus either $\hat{r} - r = 0$, or $\hat{r} - r$ has degree at least $\deg g$. The latter is clearly impossible, so $\hat{r} = r$ and $\hat{q} = q$. ■

Theorem 2.4.15 *Let F be a field. Then the ring $F[x]$ is a principal ideal domain.* ***Hint.*** *Mimic the proof of Theorem 2.4.8 and use Lemma 2.4.12!*

Proof. Let I be a nonzero ideal of $F[x]$ and let $f(x) \in I$ be a polynomial of smallest degree. We claim $I = \langle f(x) \rangle$.

Clearly $\langle f(x) \rangle \subseteq I$.

Let $g(x) \in I$ and use Theorem 2.4.14 to write $g(x) = f(x)q(x) + r(x)$, where $\deg r(x) < \deg f(x)$ or $r(x) = 0$. As in Theorem 2.4.8, write $r(x) = g(x) - f(x)q(x) \in I$, so we must have that $r(x) = 0$. Thus, $f(x)|g(x)$ and $g(x) \in \langle f(x) \rangle$. ■

Investigation 2.4.5 Is $F[x]$ a Euclidean domain for all fields F? If so, what is the norm function δ, and why does this function have the required properties of a norm? If not, why not? Prove your answer.

Answer. Yes. It's the degree function.

In fact, every Euclidean domain is a PID.

Theorem 2.4.16 *Every Euclidean domain is a principal ideal domain.* ***Hint.*** *Mimic the proof of Theorem 2.4.8.*

Proof. Let R be a euclidean domain, and I an ideal of R. If $I = \{0\}$, then I is principal, so assume that $I \neq \{0\}$.

Define $S = \{\delta(x) > 0 : x \in I\}$. Then the Well-Ordering Principle guarantees that S has a least element. Let $d \in I$ be such that $\delta(d) > 0$ is minimal. We claim that $I = \langle d \rangle$.

Clearly $\langle d \rangle \subseteq I$. Now assume that $a \in I$, and write $a = dq + r$, where either $r = 0$ or $r \neq 0$ and $\delta(r) < \delta(d)$. If $\delta(r) < \delta(d)$, then $r = a - dq \in I$, contradicting the minimality of $\delta(d)$. Thus, $r = 0$, and $a \in \langle d \rangle$.

Therefore, R is a PID. ■

Exploration 2.4.6 Where do Euclidean domains and PIDs fit in the hierarchy of abstraction found in Question 2.3.6?

Chapter 3

Factorization

In this chapter, we come to the heart of the text: a structural investigation of unique factorization in the familiar contexts of $\mathbb{Z}$ and $F[x]$. In Section 3.1, we explore theorems that formalize much of our understanding of that quintessential high school algebra problem: factoring polynomials. As we saw in Theorem 1.2.5 and Theorem 2.4.14, both $\mathbb{Z}$ and $F[x]$ have a division algorithm and, thus, are Euclidean domains. In Section 3.2, we explore the implications for multiplication in Euclidean domains. That is: given that we have a well-behaved division algorithm in an integral domain, what can we say about the factorization properties of the domain?

Finally, in the optional Section 3.3, we explore contexts in which unique factorization into products of irreducibles need not hold.

3.1 Factoring Polynomials

Guiding Questions.

In this section, we'll seek to answer the questions:

- What properties of divisibility in $\mathbb{Z}$ extend to $F[x]$?
- What is an irreducible polynomial? Are there any tools we can use to determine if a given polynomial is irreducible?

One of the most beautiful consequences of an abstract study of algebra is the fact that both $\mathbb{Z}$ and $F[x]$ are Euclidean domains. While they are not "the same", we can expect them to share many of the same properties. In

this section, our first goal will be to extend familiar properties from $\mathbb{Z}$ to $F[x]$. We will also see that particular features of a polynomial (e.g., its degree, or the existence of roots) allows for additional criteria for its irreducibility to be decided.

Since both $\mathbb{Z}$ and $F[x]$ have a division algorithm, it is reasonable to expect that, similar to the integers, we can also investigate the greatest common divisor of polynomials. In fact, our method for finding the greatest common divisor of two integers extends nicely to polynomials.

Investigation 3.1.1 Given $f(x), g(x) \in F[x]$, state a conjecture that gives a means for finding $\gcd(f(x), g(x))$. Prove your conjecture is correct.

Solution. The Euclidean algorithm! Apply the division algorithm for polynomials and mimic the proof of the Euclidean algorithm in $\mathbb{Z}$.

Investigation 3.1.2 Carefully state and prove a Bézout-like theorem (recall Theorem 1.2.9) for polynomials in $F[x]$.

Solution. Let $f(x), g(x) \in F[x]$ such that f and g are not both the zero polynomial. Then there exist polynomials $s(x), t(x) \in F[x]$ such that $f(x)s(x) + g(x)t(x) = \gcd(f(x), g(x))$.

One of the most useful things we can do with polynomials is *evaluate* them by "plugging in" elements from our coefficient set (or some superset that contains it) and performing the resulting arithmetic in an appropriate ring. We can make this completely rigorous using the language of functions: given a commutative ring R and all polynomials $p(x) = a_0 + a_1x + \cdots + a_nx^n \in R[x]$, we define the function $p_f : R \to R$ by $p_f(r) = a_0 + a_1r + \cdots + a_nr^n$. However, we will not belabor this point; instead, we will generally write $p(r)$ in place of $p_f(r)$ and appeal to our common notions of evaluating polynomials.

Given a polynomial $p(x) \in R[x]$, we have frequently been interested in finding all $r \in R$ for which $p(r) = 0$.

Definition 3.1.1 Let R be commutative with identity and suppose $p(x) \in R[x]$. We say $r \in R$ is a **zero** or **root** of $p(x)$ if $p(r) = 0$. ◊

When considering polynomials with integer coefficients, any rational roots are particularly well-behaved.

Theorem 3.1.2 *Let $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \in \mathbb{Z}[x]$ with $a_0, a_n \neq 0$. If $r, s \in \mathbb{Z}$ such that $s \neq 0, \gcd(r, s) = 1$, and $p(r/s) = 0$, then $r|a_0$ and $s|a_n$.*

Proof. TBD. ■

Activity 3.1.3 Use Theorem 3.1.2 to find the *possible* rational roots of $p(x) = 3 - x + x^2 + 5x^3 - 10x^4 - 6x^5$. Which of the possibilities you found are actually roots? Justify.

Solution. TBD.

Theorem 3.1.2 gave a condition to check to see if polynomials in $\mathbb{Z}[x]$ had roots in $\mathbb{Q}$. However, the lack of a rational root for a polynomial $q(x) \in \mathbb{Z}[x]$ is not sufficient to say that a polynomial is irreducible in $\mathbb{Z}[x]$ according to Definition 2.3.4.

Activity 3.1.4 Find a polynomial $q(x) \in \mathbb{Z}[x]$ that has no roots in $\mathbb{Q}$ but is nonetheless reducible *over* $\mathbb{Z}$.

Solution. Any polynomial with a nonunit integer factor will do, such as $q(x) = 4x^2 + 2$.

To simplify matters, we will focus henceforth on polynomials with coefficients in a field. The following theorem is a result that you learned in high school algebra (and have likely used countless times since then), but as with the other familiar topics we have explored so far, it is necessary to formalize prior to continuing.

Theorem 3.1.3 Factor Theorem. *Let F be a field, and $p(x) \in F[x]$. Then $\alpha \in F$ is a root of $p(x)$ if and only if $x - \alpha$ divides $p(x)$.*

Proof. If $x-\alpha$ divides $p(x)$, then $p(x) = q(x)(x-\alpha)$ and $p(\alpha) = q(\alpha)(\alpha-\alpha) = 0$, so α is a root.

Otherwise, use the division algorithm to divide $p(x)$ by $x - \alpha$. Then $p(x) = q(x)(x - \alpha) + r$, where $\deg r < \deg(x - \alpha) = 1$. Thus, r is a nonzero constant. If α is a root of p, then $0 = p(\alpha) = q(\alpha)(\alpha - \alpha) + r = 0 + r = r$, so $r = 0$ and $x - \alpha | p(x)$. ■

Note that while $F[x]$ is a ring, and we already have a definition of an irreducible element of a ring, we will find it useful to have a ready definition of irreducible in the context of polynomials with coefficients in a field. It is to that task that we now turn.

Exploration 3.1.5 Given a field F, define an irreducible element of $F[x]$, keeping in view Theorem 2.2.20 and Definition 2.3.4. **Hint**. What are the units in $F[x]$?

Solution. An irreducible polynomial is a nonzero nonconstant $p(x) \in F[x]$ such that whenever $p(x) = a(x)b(x)$, where $a(x), b(x) \in F[x]$, either $a(x)$ or $b(x)$ is a nonzero constant.

Definition 3.1.4 A polynomial $f(x) \in F[x]$ is **reducible** if it is not irreducible.

$\Diamond$

Exploration 3.1.6 State a positive definition for a reducible polynomial with coefficients in a field F. That is, state a definition which does not refer to the notion of irreducibility.

Solution. A polynomial $r(x)$ is reducible if it can be written as $r(x) = s(x)t(x)$, where $1 \leqslant \deg s(x) < \deg r(x)$ and $1 \leqslant \deg t(x) < \deg r(x)$.

Theorem 3.1.5 *Every polynomial of degree 1 in $F[x]$ is irreducible.*
Proof. Let $f(x)$ be degree 1 and write $f(x) = s(x)t(x)$. Then $\deg f = 1 = \deg s + \deg t$. Since $\deg s, \deg t \geqslant 0$, one of $\deg s$ or $\deg t$ is 0, hence s or t is constant. ■

Theorem 3.1.6 *A nonconstant polynomial $f(x) \in F[x]$ of degree 2 or 3 is irreducible over F if and only if it has no zeros in F.*
Proof. We prove the double contrapositive: $f(x) \in F[x]$ of degree 2 or 3 is reducible if and only if it has a zero in F.

If $f(x)$ is reducible there exist nonconstant $s(x), t(x) \in F[x]$ such that $f(x) = s(x)t(x)$. Since $\deg f = 2$ or 3, one of $s(x)$ or $t(x)$ has degree 1, and is thus of the form $x - \alpha$, where $\alpha \in F$. Thus, f has a zero.

Similarly, if f has a zero $\alpha \in F$, $f(x) = (x-\alpha)g(x)$, where $\deg g \geqslant 1$. Thus, f is reducible over F. ■

The preceding theorems allow us to explore the (ir)reducibility of polynomials of small degree with coefficients in *any field*.

Activity 3.1.7 Determine which of the following polynomials are irreducible over the given fields. Justify your answer.

1. Over $\mathbb{Z}_2$:
   (a) $x^2 + 1$,
   (b) $x^2 + x$,
   (c) $x^2 + x + 1$,
   (d) $x^3 + x^2 + 1$,
   (e) $x^4 + x^2 + 1$.
2. Over $\mathbb{Z}_3$:
   (a) $x^2 + 1$,
   (b) $x^2 + x$,
   (c) $x^2 + x + 1$,

(d) $x^2 + x + 2$,

(e) $x^3 + x + 1$,

(f) $x^3 + x^2 + 1$,

(g) $x^3 + x^2 + x + 1$.

Solution. TBD.

As the following theorem illustrates, in $F[x]$, all irreducibles are primes.

Theorem 3.1.7 *Let F be a field and $p(x), f(x), g(x) \in F[x]$ such that $p(x)$ is irreducible and $p(x)$ divides $f(x)g(x)$. Then $p(x)$ divides $f(x)$ or $p(x)$ divides $g(x)$.*

Proof. Assume $p(x)$ does not divede $f(x)$. Then $\gcd(p(x), f(x) = 1$ and $1 = s(x)p(x) + t(x)f(x)$. Multiplying by $g(x)$ yields $g(x) = g(x)s(x)p(x) + g(x)t(x)f(x)$ which implies that $p(x)$ divides $g(x)$ (since $p(x)$ divides $f(x)g(x)$) . ■

We next state the Fundamental Theorem of Algebra. Despite its name, its proof relies on analytic properties of the real numbers; there is no purely algebraic proof. Moreover, it is not essential for the work we do in following sections, but given its close relationship to the question of factorization, we include it here for completeness.

Fundamental Theorem of Algebra.

Every nonconstant polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$.

We conclude with one consequence of the Fundamental Theorem of Algebra.

Theorem 3.1.8 *Every nonconstant polynomial in $\mathbb{C}[x]$ can be written as a product of linear polynomials.* ***Hint****. What are the irreducibles in $\mathbb{C}[x]$?*

Proof. Induction on degree of polynomial using previous theorem. ■

Thus, the multiplicative structure of $\mathbb{C}[x]$ is straightforward: everything can be factored as a product of linear polynomials. Fields of coefficients like $\mathbb{C}$ for which this is true are said to be **algebraically closed**; not all fields satisfy this property. For instance, $x^2 + 1 \in \mathbb{R}[x]$ does not factor into a product of linear polynomials. Consequently, $\mathbb{R}$ is not algebraically closed.

However, regardless of whether our field is algebraically closed, we have not yet determined that any $p \in F[x]$ can be factored uniquely into a product of irreducibles, or even that such factorizations into irreducibles exist. In Section 3.2, we do just that.

3.2 Factorization in Euclidean Domains

Guiding Questions.

In this section, we'll seek to answer the questions:

- What is a unique factorization domain? What examples of UFDs do we possess?
- What is the ascending chain condition on ideals? What are Noetherian rings?
- What does the ascending chain condition have to do with unique factorization?

In this section, our explorations of the structural arithmetic properties that guarantee unique factorization culminate in (((Unresolved xref, reference "thm_everyedisufd"; check spelling or use "provisional" attribute))) . Specifically, we'll see that all Euclidean domains possess the unique factorization property. To prove this theorem, we will rely in part on an interesting property of *chains* of ideals in Euclidean domains.

3.2.1 Unique Factorization Domains

We begin by describing exactly what we mean by unique factorization. The reader may find it helpful to compare Definition 3.2.1 to The Fundamental Theorem of Arithmetic.

Definition 3.2.1 An integral domain R is called a **unique factorization domain** (or **UFD**) if the following conditions hold.

1. Every nonzero nonunit element of R is either irreducible or can be written as a finite product of irreducibles in R.
2. Factorization into irreducibles is unique up to associates. That is, if $s \in R$ can be written as

$$s = p_1 p_2 \cdots p_k \text{ and } s = q_1 q_2 \cdots q_m$$

for some irreducibles $p_i, q_j \in R$, then $k = m$ and, after reordering, p_i is an associate of q_i.

$\Diamond$

Activity 3.2.1 Using $\mathbb{Z}$ as an example, illustrate the definition of UFD by factoring 20 into two sets of *different* irreducibles which nonetheless can be paired up as associates.

We are already familiar with several examples.

Theorem 3.2.2 *The integers $\mathbb{Z}$ form a UFD.*
Proof. We have already seen that primes and irreducibles coincide in $\mathbb{Z}$ (see Theorem 2.3.5 and Theorem 2.3.6). The result follows from the Fundamental Theorem of Algebra. ■

Theorem 3.2.3 *Every field is a UFD.*
Proof. There are no nonzero nonunits in a field. The UFD conditions are therefore trivially satisfied. ■

3.2.2 The Ascending Chain Condition and Noetherian Rings

We now set our sights on a proof of Theorem 3.2.12. In order to prove it, we will make use of an important property of ideals in Euclidean domains. First, a definition.

Definition 3.2.4 A commutative ring R is called **Noetherian** if it satisfies the **ascending chain condition** on ideals.

That is, R is Noetherian if whenever

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

is an ascending chain of ideals in R, then there exists some n for which $I_n = I_{n+1} = I_{n+2} = \cdots$. ◊

Exploration 3.2.2 Consider the ideals $I_1 = \langle\, 30 \,\rangle$ in $\mathbb{Z}$ and $J_1 = \langle\, 32 \,\rangle$. Find the longest ascending chains of ideals starting first with I_1 and then with J_1 that you can. When does each chain stabilize?

Solution. TBD.

We next show that every PID is Noetherian.

Theorem 3.2.5 *Every principal ideal domain is Noetherian.* ***Hint.*** *Let $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ and set $I = \cup I_j$. Show that I is an ideal, and use your assumptions!*
Proof. It is a straightforward definition check to see that I as defined in the hint is an ideal. That we are in a PID means there exists a such that $I = \cup I_j = \langle\, a \,\rangle$. Therefore there exists j such that $a \in I_j$. It follows that $I = I_j$ and in particular that $I = I_k$ for all $k \geqslant j$. ■

Corollary 3.2.6 *Every Euclidean domain is Noetherian.*
Proof. Every ED is a PID. ■

3.2.3 Euclidean Domains are UFDs

We now begin collecting results to prove that every Euclidean domain is a UFD. The first condition in the UFD definition is that every nonzero nonunit factors as a product of irreducibles. We first show that every nonzero nonunit is divisible by at least one irreducible (Lemma 3.2.7), which we apply to show that every nonzero nonunit can be written as a finite product of irreducibles (Theorem 3.2.8).

Lemma 3.2.7 *Let R be a principal ideal domain, and $r \in R$ a nonzero nonunit. Then r is divisible by an irreducible.* ***Hint.*** *Let $r \in R$ be reducible and write $r = r_1 r_2$. Continue to factor reducibles and build an ascending chain of ideals.*

Proof. Suppose that r is not irreducible. Write $r = r_1 r_2$, where the r_i are nonzero nonunits. Then $\langle r \rangle \subsetneq \langle r_1 \rangle$. If r_1 is not irreducible, we may write $r_1 = r_{1,1} r_{1,2}$, where $r_{1,1}, r_{1,2}$ are nonzero nonunits, and observe that

$$\langle r \rangle \subsetneq \langle r_1 \rangle \subsetneq \langle r_{1,1} \rangle .$$

(If r_1 *is* irreducible and we nonetheless write $r_1 = r_{1,1} r_{1,2}$, then either $\langle r_{1,1} \rangle = R$ if $r_{1,1}$ is a unit, or $\langle r_{1,1} \rangle = \langle r_1 \rangle$ if $r_{1,1}$ is associate to r_1.)

Continuing in this way, we may continue to factor the reducible factors of r_1; since R is a PID and thus has the ascending chain condition, we must eventually reach a point where the chain stabilizes, i.e., that we have found an irreducible factor of r_1. ■

Theorem 3.2.8 *Let R be a PID. Then every nonzero nonunit element of R is either irreducible or can be written as a finite product of irreducibles in R.*

Proof. We may perform the analysis from Lemma 3.2.7 for all factors of r, and thus r can be factored into a product of irreducibles. ■

The second condition that must be satisfied for a domain to be a UFD is that the product of irreducibles must be unique (up to associates). In order to prove that, we will make use of Theorem 3.2.10, which states that in PIDs, primes and irreducibles are identical concepts.

Lemma 3.2.9 *Let R be a PID and let $p \in R$ be irreducible. Let $a \in R$ be such that $p \nmid a$. Then $1 \in I = \{ax + py : x, y \in R\}$ and thus there exist $s, t \in R$ such that $1 = as + pt$.*

Proof. Assume that p is irreducible. Suppose that p divides ab for some $a, b \in R$ and that p does not divide a. Since R is a PID, $I = \langle \gcd(a, p) \rangle = \langle 1 \rangle$. Thus there exists $s, t \in R$ such that $1 = as + pt$. ■

Theorem 3.2.10 *Let R be a PID and let $p \in R$. Then p is prime if and only if p is irreducible.*

Proof. Assume that p is prime. Suppose that $p = ab$ for some $a, b \in R$. Then p divides ab which implies that p divides a or p divides b. WLOG, assume that p divides a. Then there exists $c \in R$ such that $a = pc$ which implies that $p = pcb$. Therefore $cb = 1$ and b is a unit which implies that p is irreducible.

Assume that p is irreducible. Suppose that p divides ab for some $a, b \in R$ and that p does not divide a. Then $\langle a, p \rangle = R$ and there exists $x, y \in R$ such that $1 = ax + py$. Multiplying boths sides by b yiels $b = abx + pby = p(cx + by)$ which implies that p divides b and therefore p is prime. ■

Observe that Theorem 3.2.10 implies that if R is a PID and $p \in R$ is irreducible with $p|ab$, then $p|a$ or $p|b$.

Our crucial final step on the road to Theorem 3.2.12 is the following.

Theorem 3.2.11 *Every PID is a UFD.* ***Hint****. For part 2 of the definition, use induction on the number of irreducible divisors of an arbitrary nonzero nonunit. Mimic the proof of Theorem 1.3.8.*

Proof. Let R be a PID, and observe that by Theorem 3.2.8 every nonzero nonunit can be written as a product of irreducibles. We thus need only show that this product is unique (up to associates).

To that end, we adopt the notation of Definition 3.2.1. Let $s \in R$. We perform induction on the number of irreducible factors k of s. If $k = 0$, then s is a unit. If we had $s = pa$ for some irreducible p, then p divides a unit and would thus be a unit itself, which is a contradiction.

Suppose now that $k \geqslant 1$ and we write

$$s = p_1 p_2 \cdots p_k \text{ and } s = q_1 q_2 \cdots q_m, \ m \geqslant k,$$

where the p_i and q_j are not necessarily distinct irreducibles. Since $p_1 | q_1 q_2 \cdots q_m$, we know by Theorem 3.2.10 that p_1 must divide one of the factors.

Renumbering, we may assume that $p_1 | q_1$, whence $q_1 = p_1 v_1$, where v_1 must be a unit as q_1 is irreducible. Canceling p_1 yields

$$p_2 p_3 \cdots p_k = v_1 q_2 q_3 \cdots q_n, m \geqslant k.$$

By induction on k we conclude that each of the irreducible factors on the left matches with precisely one factor on the right (up to associates). This completes the inductive step, and thus the proof. ■

Theorem 3.2.12 *Every Euclidean domain is a unique factorization domain.*

Proof. Apply Theorems 2.4.16 and Theorem 3.2.11. ■

Theorem 3.2.13 (Unique Factorization of Polynomials). *Let F be a field. Then $F[x]$ is a UFD.*

That is, if $f(x) \in F[x]$ with $\deg(f(x)) \geqslant 1$, then $f(x)$ is either irreducible or a product of irreducibles in $F[x]$. What is more, if

$$f(x) = p_1(x)p_2(x)\cdots p_k(x) \text{ and } f(x) = q_1(x)q_2(x)\cdots q_m(x)$$

*are two factorizations of f into irreducibles p_i, q_j, then $m = k$ and after reordering, p_j and q_j are associates.**Hint**. Handle existence and uniqueness separately. For each, (strong) induction on $\deg(f(x))$ will work. Or do something entirely different.*

Thus, we see that the existence of a well-behaved division algorithm and (a lack of zero divisors) is sufficient to guarantee unique factorization. However, it is not necessary. The following theorem is included for reference, but is not intended to be proved.

Theorem.

If R is a UFD, then $R[x]$ is a UFD.

Thus, $\mathbb{Z}[x]$ is a UFD. That is, every nonconstant polynomial in $\mathbb{Z}[x]$ is either irreducible or can be factored uniquely into a product of irreducibles. However, as we will see later, $\mathbb{Z}[x]$ is not a PID.

3.3 Nonunique Factorization

Guiding Questions.

In this section, we'll seek to answer the questions:

- How can unique factorization fail, and why does it matter?
- What is an example of a nonatomic domain?
- What is an example of an element that does not factor uniquely into a product of irreducibles?

Despite the evidence to the contrary, not every ring has the unique factorization property. That is, there are commutative rings with identity which are not UFDs. In fact, the failure of certain rings in algebraic number theory to have the unique factorization property played a role in several failed attempts to prove Fermat's Last Theorem, which says that there are no nontrivial inte-

ger solutions (x, y, z) to the equation $x^n + y^n = z^n$ if $n \geqslant 3$. Pierre de Fermat famously claimed that he had a "marvelous proof" of this fact, but the margin of the book in which he was writing was "too narrow to contain it." Fermat's supposed proof was never found, and many now doubt that he had one. The search for a valid proof would not be complete until the work of Andrew Wiles and Richard Taylor in the mid-1990s.

In 1847, Gabriel Lamé claimed he had completely solved the problem. His solution relied on the factorization of $x^p + y^p$, where p is an odd prime, as

$$x^p + y^p = (x + y)(x + \zeta y) \cdots (x + \zeta^{p-1} y),$$

where $\zeta = e^{2\pi i/p}$ is a primitive p-th root of unity in $\mathbb{C}$. However, the ring $\mathbb{Z}[\zeta] = \left\{a_0 + a_1\zeta + a_2\zeta^2 + \cdots + a_{p-1}\zeta^{p-1} : a_i \in \mathbb{Z}\right\}$ is not a unique factorization domain.

There are two ways that unique factorization in an integral domain can fail: there can be a failure of a nonzero nonunit to factor into irreducibles, or there can be nonassociate factorizations of the same element. We investigate each in turn.

Exploration 3.3.1 A non-atomic domain. We say an integral domain R is **atomic** if every nonzero nonunit can be written as a finite product of irreducibles in R.

In this exploration, we encounter a non-atomic domain.

Let

$$\begin{aligned} R &= \mathbb{Z} + x\mathbb{Q}[x] \\ &= \left\{a + b_1 x + b_2 x^2 + \cdots + b_n x^n : a \in \mathbb{Z}, b_1, \ldots, b_n \in \mathbb{Q}, n \geqslant 0\right\}, \end{aligned}$$

the set of polynomials with integer constant terms and rational coefficients.

1. Convince yourself that R is an integral domain. You do not need to prove it in detail, but you should at least argue that R is closed under the usual polynomial addition and multiplication, and that R is a domain.

2. Describe the irreducibles in R.

3. Use the notion of degree to argue that any factorization of x in R has the form
$$x = m\left(\frac{x}{m}\right).$$

4. Explain why the factorization in the previous part cannot lead to a factorization of x into irreducibles in R.

Solution. TBD.

We now explore the atomic domain $R = \mathbb{Z}[\sqrt{-7}] = \left\{a + b\sqrt{-7} : \ a, b \in \mathbb{Z}\right\}$. As we will see, even when a nonzero nonunit can be written as a product of irreducibles, it may be the case that this factorization is not unique.

Activity 3.3.2 Verify that $8 = (1 + \sqrt{-7})(1 - \sqrt{-7})$.

Next, we develop a multiplicative function δ which enables us to explore the multiplicative properties of $\mathbb{Z}[\sqrt{-7}]$.

Theorem 3.3.1 *Define $\delta : R \to \mathbb{N}_0$ by $\delta(a + b\sqrt{-7}) = a^2 + 7b^2$. Then for all $x, y \in R$, $\delta(xy) = \delta(x)\delta(y)$.*

Theorem 3.3.2 *An element $u \in R$ is a unit if and only if $\delta(u) = 1$.*
Proof. Observe that u is a unit if and only if $uv = 1$ for some v, which menas that $1 = \delta(1) = \delta(u)\delta(v)$, so $\delta(u) = \delta(v) = 1$. ■

Lemma 3.3.3 *There do not exist $x, y \in \mathbb{N}_0$ such that $2 = x^2 + 7y^2$.*
Proof. Suppose there exist $x, y \in \mathbb{N}_0$ such that $2 = x^2 + y^2$. Then we must have $y = 0$, which means that $x^2 = 2$, a contradiction. ■

Theorem 3.3.4 *The elements 2, $1 + \sqrt{-7}$, and $1 - \sqrt{-7}$ are irreducible in R. We conclude that R is not a UFD.*
Proof. Suppose $2 = ab$. Then $4 = \delta(2) = \delta(a)\delta(b)$. By the lemma, we may not have $\delta(a) = 2$, which means without loss of generality that $\delta(a) = 1$, and thus a is a unit. Therefore, 2 is irreducible.

Now suppose that $1 + \sqrt{-7} = ab$. Then $8 = \delta(1 + \sqrt{-7}) = \delta(a)\delta(b)$. The possible values for $\delta(a)$ are 1, 2, 4, and 8. If $\delta(a) = 1$ or 8, then $1 + \sqrt{-7}$ is irreducible, as either a or b is necessarily a unit. By the lemma, we may not have $\delta(a) = 2$ or $\delta(b) = 2$, so in fact either $\delta(a) = 1$ or 8. Therefore, $1 \pm \sqrt{-7}$ is irreducible.

Since we have factored 8 into two different products of irreducibles, R is not a UFD. ■

Chapter 4

Ideals and Homomorphisms

The first three chapters of this text tell the story of unique factorization. The culmination is the result that any Euclidean domain is a unique factorization domain; that is, in an integral domain with a well-behaved division algorithm, a nonzero nonunit necessarily factors uniquely into irreducibles. In order to expediently develop that result, we ignored many concepts that are otherwise interesting and useful in a first course in abstract algebra. This chapter is a coda that seeks to fill in some of those gaps.

In Section 4.1, we expand on the definition of ideal introduced in Section 2.4 and explore non-principal ideals. No math course is complete without a discussion of functions of some sort; we explore homomorphisms in Section 4.2, concluding with an exploration and proof of the First Isomorphism Theorem. Finally, in Section 4.3, we introduce prime and maximal ideals, as well as the notion of congruence modulo I and use ideals to build new rings from old.

4.1 Ideals in general

Guiding Questions.

In this section, we'll seek to answer the questions:

- What operations can we perform on existing ideals to create new ideals?
- How can we describe (non-principal) ideals in general?

Recall that one of the ways in which we understand a mathematical object

is to study its relationship to other mathematical objects. In algebra, we learn about a ring by studying its relationship to other rings via functions (introduced in Section 4.2) and to its ideals, introduced in Definition 2.4.1.

The notion of an ideal number was first introduced by Ernst Kummer in the middle of the nineteenth century. Kummer was studying the cyclotomic integers in connection to work on Fermat's Last Theorem and reciprocity laws in number theory, and discovered, to use our modern terminology, that these rings of cyclotomic integers were not UFDs. In particular, he found irreducible cyclotomic integers that were not prime. His work, which was finished by Richard Dedekind by 1871, was to define a new class of complex number, an *ideal number* for which unique factorization into prime ideal numbers held. This notion of ideal number was later elaborated on by David Hilbert and Emmy Noether into the more general version which we stated in Definition 2.4.1.

In this section, we explore ways of describing non-principal ideals. We also explore properties of ideals, as well as their connections to other fields of mathematics.

We first explore the behavior of ideals under the usual set-theoretic operations of intersection and union.

Theorem 4.1.1 *Let R be a ring and let $\{I_\alpha\}_{\alpha\in\Gamma}$ be a family of ideals. Then $I = \bigcap_{\alpha\in\Gamma} I_\alpha$ is an ideal.*

Proof. It is clear that $0 \in I$. Moreover, if $x, y \in I$, then $x, y \in I_\alpha$ for all α, so $x - y \in I_\alpha$ and thus $x - y \in I$. Finally, if $x \in I$ and $r \in R$, $rx \in I_\alpha$ for all α, and thus $rx \in I$. ■

Investigation 4.1.1 Let R be a ring and $I, J \subseteq R$ be ideals. Must $I \cup J$ be an ideal of R? Give a proof or counterexample of your assertion.

Solution. Given $R = \mathbb{Z}$, $I = \langle 2 \rangle$, and $J = \langle 3 \rangle$, note that since $5 = 2 + 3 \notin I \cup J$, the union $I \cup J$ is not an ideal.

In addition to the set-theoretic properties described above, we can do arithmetic with ideals.

Theorem 4.1.2 *Let R be a ring and $I, J \subseteq R$ ideals of R. Then the **sum** of I and J,*

$$I + J := \{x + y : x \in I, y \in J\},$$

*is an ideal of R. Furthermore, the **product** of I and J,*

$$IJ := \{x_1y_1 + x_2y_2 + \cdots + x_ny_n : n \geqslant 1, x_i \in I, y_j \in J\}$$

is an ideal of R.

Proof. TBD. ■

When we studied principal ideals, we were able to describe the principal ideal in terms of a single generating element. However, not every ideal is principal (see the Challenge 4.1.8). Still, we would like a way to more precisely describe the elements of such ideals; we begin with Definition 4.1.3.

Definition 4.1.3 Let R be a commutative ring with identity, and let $S \subseteq R$ be a subset. Then

$$\langle S \rangle := \bigcap_{\substack{J \supseteq S \\ J \text{ is an ideal}}} J \tag{4.1.1}$$

is called the **ideal generated by** S, and we call S the **generating set for the ideal**. ◊

A consequence of Definition 4.1.3 is the following theorem.

Theorem 4.1.4 *Let R be a ring. Then $\langle \emptyset \rangle = \{0\}$.*

Proof. Observe that for all ideals J in R, $\emptyset \subseteq J$. In particular, one of the ideals in the right-hand side of (4.1.1) is the zero ideal. Since the zero ideal is a subset of all ideals, the theorem follows. ■

One way to interpret Definition 4.1.3 is that $\langle S \rangle$ is the smallest ideal (with respect to subset inclusion) that contains S.

Theorem 4.1.5 *Given a commutative ring R and a subset S of R, $\langle S \rangle$ is the smallest ideal containing S in the sense that, if J is any ideal of R containing S, $\langle S \rangle \subseteq J$.*

Proof. Let I be any ideal containing S. Thus, I is one of the ideals on the right-hand side of (4.1.1). Since $\langle S \rangle$ is formed by the intersection of I with other ideals, $\langle S \rangle \subseteq I$. ■

The concept elucidated by Theorem 4.1.5 is helpful, but does not give us a handle on the structure of the elements of $\langle S \rangle$. Such a description is provided by Theorem 4.1.6.

Theorem 4.1.6 *Given a commutative ring with identity R and a nonempty subset S of R:*

1. *The set $I = \{r_1 s_1 + r_2 s_2 + \cdots + r_n s_n : r_i \in R,\ s_j \in S,\ n \geqslant 1\}$ is an ideal of R;*
2. *$S \subseteq I$; and*
3. *$I = \langle S \rangle$.*

Proof. It is clear that $0 \in I$, and that if $r \in R$ and $x \in I$, $rx \in I$. Moreover, the sum of two R-linear combinations of elements of S is yet another R-linear combination of elements of S. Thus, I is an ideal. Further, if $s \in S$, $1 \cdot s \in I$,

so $S \subseteq I$. Therefore, $\langle S \rangle \subseteq I$.

Now assume that $x \in I$. Then x has the form $x = r_1 s_1 + \cdots + r_n s_n$. Each $s_i \in S$, so each $r_i s_i \in J$ if J is any ideal containing S. In particular, $r_i s_i \in \langle S \rangle$, and thus the sum $x = \sum r_i s_i \in \langle S \rangle$. ■

In other words, the ideal $\langle S \rangle$ contains all possible finite sums of products of ring elements with elements from S.

Definition 4.1.7 If R is a ring and $S = \{s_1, s_2, \ldots, s_n\}$ is a finite subset of R, the ideal I generated by R is denoted by $I = \langle s_1, s_2, \ldots, s_n \rangle$, and we say I is **finitely generated**. ◊

Challenge 4.1.8 *The ring $\mathbb{Z}[x]$ is not a PID.* ***Hint.*** *Consider the ideal $I = \langle 2, x \rangle$.*

Note that the set S in Theorem 4.1.6 need not be finite. However, in many familiar rings, every ideal will have a finite generating set, as the next theorem demonstrates.

Theorem 4.1.9 *Let R be a ring. If R is Noetherian*[1]*, then every ideal I of R is finitely generated.* ***Hint.*** *Consider an arbitrary ideal I and inductively build an ascending chain of finitely generated ideals contained in I.*

Proof. Assume that R is Noetherian. Inductively build an ascending chain of ideals as follows:

- Let $x_1 \in I$ be arbitrary.
- For all $i \geqslant 2$, let $x_i \in I \setminus \langle x_1, \ldots, x_{i-1} \rangle$ be arbitrary.

We thus obtain an ascending chain of ideals

$$\langle x_1 \rangle \subseteq \langle x_1, x_2 \rangle \subseteq \langle x_1, x_2, x_3 \rangle \subseteq \cdots$$

which must stabilize at $\langle x_1, x_2, \ldots, x_n \rangle$. That is, at some point, $I \setminus \langle x_1, x_2, \ldots, x_n \rangle = \emptyset$, and since $\langle x_1, x_2, \ldots, x_n \rangle \subseteq I$, we have $I = \langle x_1, x_2, \ldots, x_n \rangle$. ■

In fact, we could have used the finite generation of ideals as the definition of Noetherian rings, as the two notions are equivalent. First, a lemma.

Lemma 4.1.10 *Let R be a ring and $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ an ascending chain of ideals. Then*

$$I = \bigcup_{j=1}^{\infty} I_j$$

is an ideal.

Proof. Straightforward definition check. Nearly identical to Theorem 3.2.5. ■

[1] Recall Definition 3.2.4.

Theorem 4.1.11 *Let R be a ring such that every ideal of R is finitely generated. Then R is Noetherian.* ***Hint.*** *Argue that the ideal I defined in Lemma 4.1.10 is a finitely generated ideal of R, and use this to conclude that the ascending chain stabilizes.*

Proof. Following the hint and Lemma 4.1.10, we have $I = \langle x_1, x_2, \ldots, x_n \rangle$ while $I_j \subseteq I$ for all j. Since $I = \bigcup_{j=1}^{\infty} I_j$, for each $k, 1 \leqslant k \leqslant n$, there is a j_k satisfying $x_k \in I_{j_k}$. What is more, we may rename $x_1, x_2, \ldots, x_n$ so that $j_1 \leqslant j_2 \leqslant \ldots \leqslant j_n$ and thus

$$I_{j_1} \subseteq I_{j_2} \subseteq \cdots \subseteq I_{j_n}.$$

Observe that this means that $x_{j_k} \in I_{j_n}$ for all k, so that $I = I_{j_n}$. Thus, the chain stabilizes, and R is Noetherian. ■

As one might expect, not every ring is Noetherian. However, most familiar rings are.

Exploration 4.1.2 Show that the ring $R = \mathbb{Q}[x_1, x_2, x_3, \ldots]$ of polynomials in infinitely many variables over $\mathbb{Q}$ is not Noetherian either by exhibiting an ascending chain of ideals that never stabilizes, or an ideal without a finite generating set.

Solution. TBD.

We close with a discussion of a class of ideals which are the object of active mathematical research. Recall that a (simple) graph G consists of a set $V = \{x_1, x_2, \ldots, x_n\}$ of *vertices* together with a collection E of *edges*, which are just pairs of vertices and can be written $x_i x_j$. This notation suggests the following definition.

Definition 4.1.12 Let K be a field, G a graph on the vertex set $V = \{x_1, x_2, \ldots, x_n\}$ with edge set E, and let $R = K[x_1, x_2, \ldots, x_n]$ be the ring of polynomials whose variables are the vertices of G with coefficients in K. Define the **edge ideal** of G to be

$$I(G) := \langle x_i x_j \mid x_i x_j \in E \rangle.$$

That is, $I(G)$ is generated by the products of the variables corresponding to the edges of the graph. ◊

Activity 4.1.3 Consider the graph G in Figure 4.1.13. List the generators of $I(G)$ and an appropriate ring in which $I(G)$ may live.

Figure 4.1.13 A graph G.

As one might hope, we do not make Definition 4.1.12 merely for fun; given a graph G, it is possible to relate the graph-theoretic properties of G (e.g., the chromatic number) with the ideal-theoretic properties of $I(G)$. See [4.1.1] and [4.1.2], among others, for more.

References

A. Van Tuyl, *A Beginner's Guide to Edge and Cover Ideals*, in *Monomial Ideals, Computations, and Applications*, Lecture Notes in Mathematics Volume 2083, 2013, pp 63-94C. Bocci, S. Cooper, E. Guardo, et al., *The Waldschmidt constant for squarefree monomial ideals*, *J Algebr Comb* (2016) 44:875

4.2 Homomorphisms

Guiding Questions.

In this section, we'll seek to answer the questions:

- What is a ring homomorphism?
- What are some examples of ring homomorphisms?

Central to modern mathematics is the notion of *function*[1]. Functions arise in all areas of mathematics, each subdiscipline concerned with certain types of functions. In algebra, our concern is with *operation-preserving* functions, such as the linear transformations $L : V \to W$ of vector spaces you have seen in a course in linear algebra. Those linear transformations had the properties that $L(\mathbf{v} + \mathbf{u}) = L(\mathbf{v}) + L(\mathbf{u})$ (addition is preserved) and $L(c\mathbf{u}) = cL(\mathbf{u})$ (scalar multiplication is preserved).

We find something similar at work in the study of homomorphisms of rings, which we define to be functions that preserve both addition and multiplication.

Definition 4.2.1 Let R and S be commutative rings with identity. A function $\varphi : R \to S$ is a called **ring homomorphism** if it preserves addition, multiplication, and sends the identity of R to the identity of S. That is, for all $x, y \in R$:

- $\varphi(x + y) = \varphi(x) + \varphi(y)$,

[1] This section assumes a familiarity with the idea of function from a set-theoretic point of view, as well as the concepts of injective (one-to-one), surjective (onto), and bijective functions (one-to-one correspondences).

- $\varphi(xy) = \varphi(x)\varphi(y)$, and
- $\varphi(1_R) = 1_S$.

If φ is a bijection, we say that φ is an **isomorphism** and write $R \cong S$. If $\varphi : R \to R$ is an isomorphism, we say φ is an **automorphism** of R. $\Diamond$

Our first job when glimpsing a new concept is to collect a stock of examples.

Exploration 4.2.1 Determine whether the following functions are homomorphisms, isomorphisms, automorphisms, or none of these. Note that R denotes an arbitrary commutative ring with identity.

1. $\varphi : R \to R$ defined by $\varphi(x) = x$
2. $\psi : R \to R$ defined by $\psi(x) = -x$
3. $\alpha : \mathbb{Z} \to \mathbb{Z}$ defined by $\alpha(x) = 5x$
4. $F : \mathbb{Z}_2[x] \to \mathbb{Z}_2[x]$ defined by $F(p) = p^2$
5. $\iota : \mathbb{C} \to \mathbb{C}$ defined by $\iota(a + bi) = a - bi$, where $a, b \in \mathbb{R}, i^2 = -1$
6. $beta : \mathbb{Z} \to \mathbb{Z}_5$ defined by $\beta(x) = \overline{x}$
7. $\epsilon_r : R[x] \to R$ defined by $\epsilon_r(p(x)) = p(r)$ (this is known as the r-evaluation map)
8. $\xi : \mathbb{Z}_5 \to \mathbb{Z}_{10}$ defined by $\xi(\overline{x}) = \overline{5x}$

Solution. TBD Note that $\varphi(x + y) = \overline{x + y} = \overline{x} + \overline{y} = \varphi(x) + \varphi(y)$, and similarly for multiplication. We observe that φ is not an isomorphism, as $\varphi(2) = \varphi(7) = \overline{2}$.

Homomorphisms give rise to a particularly important class of subsets: kernels.

Definition 4.2.2 Let $\varphi : R \to S$ be a ring homomorphism. Then $\ker \varphi = \{r \in R : \varphi(r) = 0_S\}$ is the **kernel** of φ. $\Diamond$

Activity 4.2.2 For each homomorphism in Exploration 4.2.1, find (with justification), the kernel.

Solution. TBD

In fact, kernels are not just important subsets of rings; they are ideals.

Theorem 4.2.3 *Given a ring homomorphism $\varphi : R \to S$, $\ker \varphi$ is an ideal.*

Proof. If $x, y \in \ker\varphi$, $\varphi(x+y) = \varphi(x) + \varphi(y) = 0_S + 0_S = 0_S$. Similarly, if $x \in \ker\varphi$ and $r \in R$, $\varphi(rx) = \varphi(r)\varphi(x) = \varphi(r)0_S = 0_S$. ■

Kernels also give a useful way of determining whether their defining homomorphisms are one-to-one.

Theorem 4.2.4 *Let $\varphi : R \to S$ be a homomorphism. Then φ is one-to-one if and only if $\ker\varphi = \{0\}$.*

Proof. TBD ■

4.3 Quotient Rings: New Rings from Old

Guiding Questions.

In this section, we'll seek to answer the questions:

- How can we use ideals to build new rings out of old?
- What sorts of ideals allow us to build domains? Fields?
- What is the First Isomorphism Theorem?

If the only rings that existed were polynomial rings, familiar systems of numbers like $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, and matrix rings, there would still be enough to justify the defining the concept of a ring and exploring its properties. However, these are not the only rings that exist. In this section, we explore a way of building new rings from old by means of ideals. To better understand these new rings, we will also define two new classes of ideals: prime ideals, and maximal ideals. We end by briefly connecting these rings to a familiar problem from high school algebra.

4.3.1 Congruence modulo I

The major concept of this section is the notion of congruence modulo I. One can reasonably think of this idea as a generalization of congruence modulo m in $\mathbb{Z}$.

Definition 4.3.1 Let R be a ring and I an ideal of R. Then elements $a, b \in R$ are said to be **congruent modulo** I if $b - a \in I$. If this is the case, we write $a + I = b + I$. ◊

Activity 4.3.1 Determine (with brief justification) whether $a + I = b + I$ in the following rings R.

1. $a = 9$, $b = 3$, $I = \langle 6 \rangle$, $R = \mathbb{Z}$

2. $a = 10$, $b = 4$, $I = \langle 7 \rangle$, $R = \mathbb{Z}$

3. $a = 9$, $b = 3$, $I = \langle 6 \rangle$, $R = \mathbb{Z}[x]$

4. $a = x^2 + x - 2$, $b = x - 1$, $I = \langle x + 1 \rangle$, $R = \mathbb{Q}[x]$

5. (*Challenge.*) $a = x^3$, $b = x^2 + 2x$, $I = \langle y - x^2, y - x - 2 \rangle$, $R = \mathbb{Q}[x, y]$

Solution. TBD last one: $x^3 - x^2 - 2x = x(x-2)(x+1) = -x(y - x^2) + x(y - x - 2)$.

Exploration 4.3.2 Given a ring R, ideal I, and $a \in R$, when is it the case that $a + I = 0 + I = I$?

Answer. When $a \in I$.

Observe that if $b - a \in I$, then there is some $x \in I$ such that $b - a = x$, and so $b = a + x$.

As was the case in $\mathbb{Z}_m$, congruence modulo I is an equivalence relation.

Theorem 4.3.2 *Let R be a ring and I an ideal of R. Then congruence modulo I is an equivalence relation on R.*

Proof. Since $0 \in I$, $a - a \in I$ for all a, so $a + I = a + I$ and the relation is reflexive.

Moreover, if $b - a \in I$, then $-1(b - a) = a - b \in I$, so the relation is symmetric.

If $b - a \in I$ and $c - b \in I$, then $(b - a) + (c - b) = c - a \in I$, so the relation is transitive. ■

The set of equivalence classes under this relation is denoted R/I. What is more, this is not merely a set of equivalence classes. As the next two theorems demonstrate, this set possesses two algebraic operations that extend naturally from those of R.

Theorem 4.3.3 *Let R be a ring and I an ideal of R. If $a, b, c, d \in R$ such that $a + I = b + I$ and $c + I = d + I$, then $(a + c) + I = (b + d) + I$.*

Proof. Suppose $a + I = b + I$ and $c + I = d + I$. Then $b - a = x \in I$ and $d - c = y \in I$. Adding, we have $(b - a) + (d - c) = (b + d) - (a + c) = x + y \in I$. Thus, $(b + d) + I = (a + c) + I$. ■

Theorem 4.3.4 *Let R be a ring and I an ideal of R. If $a, b, c, d \in R$ such that $a + I = b + I$ and $c + I = d + I$, then $ac + I = bd + I$.*

Proof. As before, we observe that $b - a = x \in I$ and $d - c = y \in I$. Write $b = x + a$ and $d = c + y$. Then $bd = (x + a)(c + y) = xc + xy + ac + ay = ac + \underbrace{(xc + xy + ay)}_{\in I}$, so $bd - ac \in I$, and thus $bd + I = ac + I$. ■

The previous two theorems together show that addition and multiplication on the set R/I is well-defined. As these operations are built on the operations of R, it will likely not surprise you to learn that the usual axioms defining a ring also hold.

Theorem 4.3.5 *Let R be a commutative ring with identity 1_R and I an ideal of R. The set of equivalence classes modulo I, denoted R/I, is a commutative ring with identity $1_R + I$ under the operations of addition modulo I and multiplication modulo I defined in Theorem 4.3.3 and Theorem 4.3.4.*

Proof. TBD ■

Thus, given a ring R and ideal I of R, we may build a new ring R/I. In Subsection 4.3.2, we will explore the question of when R/I possesses some of the properties we've previously explored, e.g., when is R/I a domain? A field? First, we conclude with two explorations. The first gives us a sense of what these rings can look like. The second connects quotient rings to *solution sets* of polynomial equations.

Exploration 4.3.3 Consider the ring $R = \mathbb{Z}_2[x]$ and the ideals $I = \langle\, x^2 - 1 \,\rangle$ and $J = \langle\, x^3 - x - 1 \,\rangle$.

1. List the elements of R/I and R/J.
2. What happens to x^2 in R when you pass to the quotient ring R/I? How about x^3 as you pass from R to R/J?
3. In view of your answer to the previous question, how does x behave as you "mod out" by I and J?
4. Build addition and multiplication tables for each of R/I and R/J.

Solution. TBD

Exploration 4.3.4 One of the most useful connections made in high school algebra is the connection between a function f (in particular, a polynomial function) and its *graph*. We may extend this notion to ideals via the concept of a *zero set* as follows.

Let F be a field and $R = F[x, y]$ with $I \subseteq R$ a nonzero ideal. We define the **zero set** of I, denoted $Z(I)$, as the set of all points $(a, b) \in F^2$ for which $f(a, b) = 0$ for all $f \in I$.

1. Suppose $I = \langle\, f_1, f_2, \ldots, f_n \,\rangle$. Prove that $(a, b) \in Z(I)$ if and only if $f_j(a, b) = 0$ for each $j \in \{1, \ldots, n\}$. Thus, $Z(I)$ can be determined entirely by examining the generators of I.

2. Describe $Z(I)$ given $I = \langle\, y - x^2 \,\rangle$.

3. (Challenge) Given $I = \langle\, y - x^2 \,\rangle$ and $J = \langle\, y - x - 2 \,\rangle$, describe $Z(I+J)$ and $Z(I \cap J)$.

4. Given $I = \langle\, y - x^2 \,\rangle$, describe the relationship between the variables x and y in the quotient R/I. In what way have we restricted our polynomial "inputs" to the parabola $y = x^2$?

Solution. TBD.

4.3.2 Prime and Maximal Ideals

In this section, we continue our exploration of quotient rings by looking more closely at properties of ideals. We focus on particular properties of ideals that ensure that the quotient R/I is either a domain or a field.

Definition 4.3.6 Let R be commutative with identity and $P \subsetneq R$ a nonzero ideal. We say P is **prime** if whenever $a, b \in R$ such that $ab \in P$, we have $a \in P$ or $b \in P$. ◊

Theorem 4.3.7 *Let R be a domain and $p \in R$ be prime. Then $\langle\, p \,\rangle$ is a prime ideal.*
Proof. TBD ■

Activity 4.3.5 Which of the following ideals are prime?

1. $\langle\, 9 \,\rangle$ in $\mathbb{Z}$

2. $\langle\, 11 \,\rangle$ in $\mathbb{Z}$

3. $\langle\, x^2 + 1 \,\rangle$ in $\mathbb{R}[x]$

4. $\langle\, x^2 - 1 \,\rangle$ in $\mathbb{R}[x]$

5. $\langle\, x^2 - 5x + 6, x^4 + 2x^3 - 10x^2 + 5x - 2 \,\rangle$ in $\mathbb{R}[x]$

Solution.

1. Not prime. $3 \cdot 3 = 9 \in \langle\, 9 \,\rangle$, but $3 \notin \langle\, 9 \,\rangle$.

2. Prime. If $x \in \langle\, 11 \,\rangle$, then $11|x$ and 11 is prime.

3. Prime. Easy explanation is that $R[x]/\langle\, x^2 + 1 \,\rangle$ is isomorphic to $\mathbb{C}$. ⌣ For now, though, we know that $x^2 + 1$ is irreducible, and $R[x]$ is a PID, so irreducibles are prime.

4. Not prime. $x^2 - 1 = (x - 1)(x + 1)$, but $x \pm 1 \notin \langle\, x^2 - 1 \,\rangle$ for degree reasons.

5. One may use the EA to show that $\gcd(x^2 - 5x + 6, x^4 + 2x^3 - 10x^2 + 5x - 2) = x - 2$, so $\langle\, x^2 - 5x + 6, x^4 + 2x^3 - 10x^2 + 5x - 2 \,\rangle = \langle\, x - 2 \,\rangle$. This is prime.

It is this precise condition that guarantees that the resulting quotient is a domain.

Theorem 4.3.8 *Let R be commutative with identity and I an ideal of R. Then I is prime if and only if R/I is an integral domain.*

Proof. Begin by assuming that I is prime, and suppose $(a+I)(b+I) = 0+I = I$. Then $ab \in I$, and since I is prime, either $a \in I$ or $b \in I$. Thus either $a+I = 0+I$ or $b + I = 0 + I$.

Now assume that R/I is a domain. Further, let $a, b \in R$ be such that $(a + I)(b + I) = ab + I = 0 + I$. Then $ab \in I$, and since R/I is a domain, $a + I = 0 + I$ or $b + I = 0 + I$, i.e., $a \in I$ or $b \in I$. Thus, I is prime. ■

We now consider another important class of ideals: the maximal ideals.

Definition 4.3.9 Let R be commutative with identity and let $M \subsetneq R$ be a nonzero ideal. We say that M is a **maximal ideal** if no proper ideal of R properly contains M. That is, if J is an ideal satisfying $M \subseteq J \subseteq R$, either $J = M$ or $J = R$. ◊

In other words, an ideal M is maximal if no "larger" ideal (with respect to inclusion) properly contains it. As we will see later, rings can have many maximal ideals.

It is a fact that any ring R with $0_R \neq 1_R$ has a maximal ideal. This follows from *Zorn's Lemma*; a rigorous exploration of Zorn's Lemma lies outside of the scope of this text, but suffice it to say that Zorn's Lemma is incredibly useful in all areas of algebra for proving existence theorems. For example, a proof that every vector space has a basis relies on Zorn's Lemma.

Rings with only one maximal ideal are said to be *local rings*, and are actively studied in modern research in commutative algebra (the study of commutative rings and their properties).

The next two results demonstrate that the maximality of I is precisely the condition that guarantees that R/I is a field.

Lemma 4.3.10 *Let R be commutative with identity and M a maximal ideal of R. Let $x \in R \setminus M$, and set $J = \{xr + y : r \in R,\ y \in M\}$. Then $M \subsetneq J$, and thus there exist $r' \in R$, $y' \in M$ such that $1 = xr' + y'$.*

Proof. TBD ■

Theorem 4.3.11 *Let R be commutative with identity and I an ideal of R. Then I is maximal if and only if R/I is a field.* ***Hint****. For the forward direction, apply the previous lemma to construct an inverse for $x + I$ given any $x \in R \setminus I$.*

Proof. If R/I is a field, assume J is an ideal of R that properly contains I. Let $x \in J \setminus I$; then $x + I$ is a nonzero element of R/I, and since R/I is a field, there is some $y + I$ such that $(xy) + I = 1 + I$. Since $x \in J$, $xy \in J$. As $1 + I = (x + I)(y + I) = xy + I$, we have $1 - xy \in I \subsetneq J$, and thus $1 = (1 - xy) + xy \in J$, which means $J = R$. Thus, I is maximal.

Now, suppose that I is maximal and let $x \in R \setminus I$. Apply the previous lemma to obtain $1 = xr' + y'$, where $y' \in I$. Then

$$1 + I = xs + y + I = xs + I = (x + I)(y + I).$$

■

Theorem 4.3.12 *Every maximal ideal is prime.*

Proof. All fields are integral domains. Thus, if I is maximal, R/I is a field, thus a domain, and thus I is prime. ■

In general, the converse is not true (see the Challenge below). However, it holds in sufficiently nice rings.

Theorem 4.3.13 *In a principal ideal domain, every prime ideal is maximal.*

Proof. Let R be a PID and $\langle p \rangle$ a prime ideal. By previous work, p is prime. Suppose that $\langle p \rangle \subseteq \langle m \rangle$. Thus, $p \in \langle m \rangle$, so $m|p$. That is, $p = mk$. Since p is prime and R is a domain, it is irreducible. Thus, either m or k is a unit. If m is a unit, then $\langle m \rangle = R$. If k is a unit, then $m = k^{-1}p$, and thus $m \in \langle p \rangle$, which means that $\langle m \rangle = \langle p \rangle$.

Thus, $\langle p \rangle$ is maximal. ■

Exploration 4.3.6 Describe the prime and maximal ideals of $\mathbb{Z}$ and $\mathbb{Q}[x]$.

Hint. For which ideals I is $\mathbb{Z}/I$ a domain? A field? Similarly for $\mathbb{Q}[x]$. Or, use Theorem 4.3.13.

Solution. TBD

Challenge.

Find a commutative ring with identity, R, and a nonmaximal prime ideal P of R.

4.3.3 Homomorphisms and Quotient Rings

As quotient rings provide fertile soil for building new examples of rings, it should not surprise us to find that homomorphisms interact with quotient rings in interesting and useful ways. Chief among them are the *isomorphism theorems*. In this subsection, we focus primarily on the First Isomorphism Theorem.

We have seen that any homomorphism $\varphi : R \to S$ gives rise to an ideal of R, namely $\ker \varphi$. Our next theorem demonstrates that, given a commutative ring with identity R, every ideal is the kernel of some homomorphism defined on R.

Theorem 4.3.14 *Let R be commutative with identity and I an ideal of R. Define $\varphi : R \to R/I$ by $\varphi(r) = r + I$. Then φ is a homomorphism with $\ker \varphi = I$.*

In what follows, we work toward a proof of the First Isomorphism Theorem for Rings.

Throughout, let R and S be commutative rings with identity, and let $\varphi : R \to S$ be a homomorphism. Recall that $\text{im}\varphi = \{s \in S : \ \varphi(r) = s \text{ for some } r \in R\}$.

Define $f : R/\ker \varphi \to \text{im}\varphi$ by $f(r + \ker \varphi) = \varphi(r)$.

Lemma 4.3.15 *Using the notation from above, f is a well-defined function.*
Proof. Suppose that $r_1 + \ker \varphi = r_2 + \ker \varphi$. Then $r_2 - r_1 \in \ker \varphi$, so $\varphi(r_2 - r_1) = 0_S$, and thus $\varphi(r_1) = \varphi(r_2)$. Therefore, $f(r_1 + \ker \varphi) = f(r_2 + \ker \varphi)$, and f is well-defined. ■

Lemma 4.3.16 *Using the notation above, f is a homomorphism.*
Proof. We show that f preserves addition. That it preserves multiplication will follow similarly. Observe that $f((x + \ker \varphi) + (y + \ker \varphi)) = f((x + y) + \ker \varphi) = \varphi(x + y) = \varphi(x) + \varphi(y) = f(x + \ker \varphi) + f(y + \ker \varphi)$. ■

Lemma 4.3.17 *Using the notation above, f is one-to-one.*
Proof. Suppose that $f(r_1 + \ker \varphi) = f(r_2 + \ker \varphi)$. That is, $\varphi(r_1) = \varphi(r_2)$. Then $\varphi(r_1 - r_2) = 0_S$, so $r_1 - r_2 \in \ker \varphi$, and therefore $r_1 + \ker \varphi = r_2 + \ker \varphi$. Thus, f is one-to-one. ■

Lemma 4.3.18 *Using the notation above, f is onto.*
Proof. Since φ is onto $\text{im}\varphi$ by definition, given any $s \in \text{im}\varphi$ there is some $r \in R$ such that $\varphi(r) = s$. Then $f(r + \ker \varphi) = \varphi(r)$. ■

We thus obtain:

Theorem 4.3.19 (First Isomorphism Theorem). *Let $\varphi : R \to S$ be a homomorphism of commutative rings. Then $R/\ker \varphi \cong im\varphi$.*

In particular, if $\varphi : R \to S$ is onto, $R/\ker \varphi \cong S$.

The First Isomorphism Theorem gives a useful way of establishing an isomorphism between a quotient ring R/I and another ring S: find an onto homomorphism $R \to S$ with kernel I.

Theorem 4.3.20 *We have the following isomorphisms of rings.*

1. $\mathbb{Z}/\langle m \rangle \cong \mathbb{Z}_m$
2. $\mathbb{Q}[x]/\langle x-5 \rangle \cong \mathbb{Q}$
3. $\mathbb{R}[x]/\langle x^2+1 \rangle \cong \mathbb{C}$

Proof. TBD Define $\varphi : \mathbb{R}[x] \to \mathbb{C}$ by $\varphi(p(x)) = p(i)$. We saw earlier that this evaluation map is a homomorphism. It is easy to see that φ is onto as $\varphi(a+bx) = a+bi$ for any $a, b \in \mathbb{R}$. Thus, $\mathbb{R}[x]/\ker\varphi \cong \mathbb{C}$ by the First Isomorphism Theorem.

We claim that $\ker\varphi = \langle x^2+1 \rangle$. Clearly, $\langle x^2+1 \rangle \subseteq \ker\varphi$. Moreover, $\ker\varphi \subsetneq \mathbb{R}[x]$. Finally, $\langle x^2+1 \rangle$ is prime and thus maximal, as $\mathbb{R}[x]$ is a PID. Thus, $\ker\varphi = \langle x^2+1 \rangle$. ■

Activity 4.3.7 Let $R = \mathbb{Z}_6$ and define $\varphi : \mathbb{Z}_6 \to \mathbb{Z}_2$ by $\varphi(\overline{x}) = \overline{x}$. That is, φ sends an equivalence class $\overline{x} \in \mathbb{Z}_6$ represented by $x \in \mathbb{Z}$ to the equivalence class represented by x in $\mathbb{Z}_2$.

1. Show that φ is a well-defined function.
2. Prove that φ is a homomorphism.
3. Is φ onto? Justify.
4. Compute $\ker\varphi$ (that is, list the elements in the set). Is φ one-to-one?
5. Without appealing to the definition, is $\ker\varphi$ prime? Explain.

Solution. TBD

Index