Modern Algebra

Modern Algebra

Michael Janssen
Dordt University

Melissa Lindsey
Dordt University

Edition: 2019 Beta 1

Website: https://book.mkjanssen.org

Introduction

While it is difficult to intuitively define the way in which mathematics is structural, its concern with *number* is clear, especially for students of elementary mathematics such as arithmetic. However, the study of *number* rapidly leads one to structural questions regarding the properties of numbers themselves, which has historically been the domain of *algebra*.

In (((Unresolved xref, reference "franklin2014aristotelian"; check spelling or use "provisional" attribute))) , 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 these notes is to explore the idea of factorization from an abstract perspective. We will first give attention to factorization in the integers, and then factorization properties of polynomials. Our larger goal is to precisely describe deep structural properties common to both the integers and polynomials that guarantee that factorization into fundamental pieces, like primes and irreducibles, can be done *uniquely*. In order to accomplish this goal, we will walk in the realms of abstraction, and catch glimpses of the beauty and incredible power of this perspective on mathematics. ReferencesReferencesg:references:idm416673863728

A Note to Students

The note will go here.

A Note to Instructors

Strong opinions informed the writing of these notes. The strongest are (a) that students learn math best by doing it, and (b) that students–especially preservice 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. 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.

Groups are not present in these notes. 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*.

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 familiar 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

Note that in this section we will assume the basic algebraic/arithmetic properties of the integers, most of which we will formalize via axioms in subsequent sections.

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 number. (A model of $\mathbb{N}$ via set theory and the Peano axioms describes this order using *successors*.) A consequence of the Peano postulates is the *well-ordering principle*, which we take 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,* $\leqslant$.

Definition 1.1.3 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[1]. 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).

It can be proved that the principle of mathematical induction is logically equivalent to the well-ordering principle. That is, if we assume either one to be true, we can deduce the other. We will generally take the well-ordering principle on $\mathbb{N}$ to be an axiom, and treat induction as a theorem.

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 (**??**), 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 (**??**) and (**??**) 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 (**??**). ■

We conclude with opportunities to practice induction.

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

Exploration 1.1.1 Induction on the sum of cubes. Prove that the sum of the first n cubes is $\dfrac{n^2(n+1)^2}{4}$.

Checkpoint 1.1.4 de Moivre's formula. Prove that, for all $n \geqslant 1$, $(\cos(x) + i\sin(x))^n = \cos(nx) + i\sin(nx)$.

Hint. You'll need some trig identities!

Write a conclusion!

1.2 Divisibility in the integers

In this section, we begin to explore some of the arithmetic and algebraic properties of $\mathbb{Z}$. One of the primary goals of this sections is to formalize definitions that you've been using and have an intuitive understanding of. At first glance, this might seem to add an unnecessary layer of complication. However, it will be abundantly clear as we move forward that a formalized mathematical language and notation are necessary to extend these properties to a more abstract setting. We begin with a twist on 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 $ac = b$. If no such $c \in Z$ exists, we write $a \nmid b$. ◊

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).

Checkpoint 1.2.2 Determine whether $a \mid b$ if:

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

Theorem 1.2.3 *Let a, b, and c be integers. 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 1.2.4 *Let a, b, and c be integers. 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 $a \mid bc$. ■

Definition 1.2.5 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 \marginpar{\setstretch{0.7}**This formalizes the idea of greatest common factors that is introduced around sixth grade .**}

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 $d \geqslant e$. 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 $\mathbb{Z}$.

Checkpoint 1.2.6 Compute $\gcd(a, b)$ if:

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

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 1.2.7 The Division algorithm. *Let $a, b \in \mathbb{N}$. Then there exist unique integers q, r such that $a = bq + r$, where $0 \leqslant r < b$.*

Note that this theorem has two parts: existence and uniqueness. Don't try to do them both at once!

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. Further, note that S has a largest element: every element $x \in S$ satisfies $x \leqslant a$, since $a, b \in \mathbb{N}$. Let q be the largest element of S, and set $r = a - bq$.

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., $q = q'$. ∎

Theorem 1.2.8 *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 $d|a$. ∎

Theorem 1.2.9 *Let a, b, n, and r be integers with a and b not both 0. If $a = nb + r$, then $\gcd(a, b) = \gcd(b, r)$.*

Proof. By the previous theorem, if d is a common divisor of a and b, then d also divides r, and $d \leqslant \gcd(b, r)$. Suppose $d = \gcd(a, b)$, and $\gcd(b, r) = e > d$ is an integer that divides both b and r, say $eq_1 = b$ and $eq_2 = r$. Then $a = nb + r = n(eq_1) + (eq_2) = e(nq_1 + q_2)$, so $e|a$, contradicting the assumption that $d = \gcd(a, b)$. Thus $e = \gcd(b, r) \leqslant d$, hence $e = d$. ∎

Checkpoint 1.2.10 As an illustration of the above theorem, note that

$$\begin{aligned} 56 &= 3 \cdot 12 + 8, \\ 12 &= 1 \cdot 8 + 4, \\ 8 &= 2 \cdot 4 + 0. \end{aligned}$$

Use the preceding theorem to show that if $a = 56$ and $b = 12$, then $\gcd(56, 12) = \gcd(8, 4) = 3$.

Solution. Applying the previous theorem to line 1 tells us that $\gcd(51, 15) = \gcd(15, 6)$. The second line tells us that $\gcd(15, 6) = \gcd(6, 3)$. The third line tells us that $\gcd(6, 3) = \gcd(3, 0) = 3$. The result follows by transitivity.

Checkpoint 1.2.11 [1] (Euclidean algorithm.) Using the previous theorem and the Division algorithm successively, devise a procedure for finding the greatest common divisor of two integers.

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

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

Checkpoint 1.2.12 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.

Theorem 1.2.13 Bézout's Identity. *For any integers a and b not both 0, there are integers x and y such that*

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

Proof. If, e.g., $a = 0$ and $b \neq 0$, then $\gcd(a, b) = b$ and $x = 0, y = 1$ suffices. If $a = b$, then $\gcd(a, b) = a = b$, so $x = 0, y = 1$ suffices.

So, suppose $a, b \neq 0$, and let $d = \gcd(a, b)$. Without loss of generality assume $|a| > |b|$. The EA provides a sequence of integers $r_1, r_2, \ldots, r_k, q_1, q_2, \ldots, q_{k+1}$ and equations

$$\begin{aligned} |a| &= |b|q_1 + r_1 \\ |b| &= r_1q_2 + r_2 \\ r_1 &= r_2q_3 + r_3 \end{aligned}$$

[1] The lead-up to the Euclidean algorithm has been adapted from (((Unresolved xref, reference "NTI"; check spelling or use "provisional" attribute))) .

$$\vdots$$
$$r_{k-2} = r_{k-1}q_k + r_k$$
$$r_{k-1} = r_k q_{k+1},$$

where $0 \leqslant r_i < |b|$ and $r_k \neq 0$. Then $r_k = \gcd(a, b)$ and we may back-substitute to find integers x and y such that $ax + by = r_k = \gcd(a, b)$. ■

Theorem 1.2.14 *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 **??**, 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 $a|c$. ■

1.3 Primes and Factorization

As described in the Introduction, 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 numbers. 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. ◊

Theorem 1.3.2 *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 $p|b$. ■

Question 1.3.3 Using the previous theorem as a guide, give a biconditional characterization for composite numbers. That is, fill in the blank: "A number $m \in \mathbb{N}$ is composite if and only if \makebox[0.75in]{\hrulefill}."

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

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 **??**, there are integers $a, b \neq 1$ such that $k+1 = ab$. By induction, there is a prime p such that $p|a$, so $p|k+1$. ■

Our first major theorem of the semester 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.*

Theorem 1.3.5 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.*

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. *Use the Well-Ordering Principle (or something equivalent to it.)*

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 $p = q_2$. ■

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*

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

[1] This approach to the Fundamental Theorem of Arithmetic is adapted from (((Unresolved xref, reference "NTI"; check spelling or use "provisional" attribute))) .

$$= q_1^{t_1} q_2^{t_2} \cdots q_s^{t_s}.$$

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; that is, 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. 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 $i = 1, 2, \ldots, m$. ■

1.4 The Integers modulo m

If you are familiar with 'military time' then you understand that eighteen hundred hours and 6:00pm are equivalent ways of representing the same time. Similarly, it is common knowledge that if it is 10:00am now, it will again be 10:00am in 24 hours (though on a different day). Whether you realize it or not, when you are working with time (in hours) you are working in the integers modulo 12, a concept that is rooted in the mathematical idea of equivalence classes.

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. ◊

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. ◊

Checkpoint 1.4.3 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.

Question 1.4.4 What other relations can you think of? Write down one example and one non-example of an equivalence relation. □

Checkpoint 1.4.5 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 $3 \not\leqslant 2$.

Definition 1.4.6 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 $a \equiv_m b$. ◊

Checkpoint 1.4.7 Justify the following congruences.

1. $18 \equiv_{12} 6$
2. $47 \equiv_{13} 8$
3. $71 \equiv_5 1$
4. $21 \equiv_{11} -1$

Theorem 1.4.8 *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 $a \equiv_m c$. ■

Question 1.4.9 What are the equivalence classes of $\mathbb{Z}$ under the relation $\equiv_5$?

Answer. The equivalence classes are

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

□

Checkpoint 1.4.10 List the elements of $\mathbb{Z}_7$.

Theorem 1.4.11 *Let $a, b, c, d \in \mathbb{Z}$ and $m > 1$ such that $a \equiv_m c$ and $b \equiv_m d$. Then $a + b \equiv_m c + d$.*

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 $a + b \equiv_m c + d$. ■

Theorem 1.4.12 *Let $a, b, c, d \in \mathbb{Z}$ and $m > 1$ such that $a \equiv_m c$ and $b \equiv_m d$. Then $ab \equiv_m cd$.*

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 $ab \equiv_m cd$. ■

Definition 1.4.13 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 1.4.14 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:* $\overline{a} + \overline{b} := \overline{a + b}$.\marginpar{\setstretch{0.7}*The symbol $:=$ is often used to indicate that we are* defining *the expression on the left to equal the expression on the right.*}
2. *Multiplication modulo m:* $\overline{a} \cdot \overline{b} := \overline{a \cdot b}$.

◊

Question 1.4.15 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 1.4.16 *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 **??**, $m|a - b$. ■

Theorem 1.4.17 *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 the previous theorem. ■

Chapter 2

Fields and Rings

You have been exploring numbers and the patterns they hide within them since you began elementary school (if not before). In Chapter 1 we reminded ourselves about what some of those patterns are in the integers (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 more formal notation that will assist us in streamlining our investigation into factorization.

2.1 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 $ad = bc$.

Checkpoint 2.1.2 Prove that our elementary school definition of equivalent fractions is an equivalence relation.

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 will occupy a significant amount of time in a first course in real analysis. Thus, we will make use of your intuition. It is certainly the case that $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$, but $\mathbb{R}$ also contains many irrational numbers, such as π, e, $\sqrt{2}$, and so on.

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

Definition 2.1.3 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 $\mathbb{C}$. $\Diamond$

Definition 2.1.4 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 $\star(a, b)$.] $\Diamond$

Question 2.1.5 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 $\mathbb{C}$?

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. $\square$

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.

More generally, the use of any hypotheses when asking mathematical or scientific questions is a form of creativity, and even *play*. God has created us in His image, and he is a creative being. The definitions we make and the theorems we prove are reflections of the *imago Dei* in us. Of course, we do not have the power to create *ex nihilo*; instead, our tools are the mathematical *aspects* God has imbued in his creation. We can postulate structures with whichever axioms we like, but our creativity is most valuable if it reflects the world *as it is*.

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 2.1.6 A **field** is a nonempty set F with at least two elements and binary operations $+$ and $\cdot$, 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 $a, b, c \in F$, $(a + b) \cdot c = a \cdot c + b \cdot c$.

$\Diamond$

For brevity's sake, we will usually write $a \cdot b$ as ab.

Checkpoint 2.1.7 Which of the following are fields? If no operations are specified, assume the most obvious operations (e.g., the addition and multiplication you have known for years). For most, a short justification or counterexample is sufficient.

1. $\mathbb{N}$
2. $\mathbb{Z}$
3. $2\mathbb{Z}$, the set of even integers
4. $\mathbb{Q}$
5. $\mathbb{Z}_{10}$
6. $\mathbb{Z}_{11}$
7. $\mathbb{R}$
8. $\mathbb{C}$, with addition defined as $(a_1+b_1i)+(a_2+b_2i) := (a_1+a_2)+(b_1+b_2)i$, and multiplication defined as $(a_1+b_1i)(a_2+b_2i) := (a_1a_2-b_1b_2)+(a_1b_2+b_1a_2)i$.
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.

In the previous exercise, 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 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.**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. $\mathbb{Z}_{10}$ is not a field; there is no $\overline{a} \in \mathbb{Z}_{10}$ for which $\overline{5} \cdot \overline{a} = \overline{1}$ (you can check them all; there are only 9 viable options).
6. $\mathbb{Z}_{11}$ *is* a field. The axioms pertaining to addition and multiplication were established earlier or rely on the same axioms holding for $\mathbb{Z}$.

[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 $\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \cdot \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = \begin{pmatrix} a_1a_2 + b_1c_2 & a_1b_2 + b_1d_2 \\ c_1a_2 + d_1c_2 & c_1b_2 + d_1d_2 \end{pmatrix}$.

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.

Theorem 2.1.8 properties of fields. *Let F be a field.*

1. *The additive identity 0_F is unique.*
2. *For all $a \in F$, $a \cdot 0_F = 0_F \cdot a = 0_F$.*
3. *Additive inverses are unique.*
4. *The multiplicative identity 1_F is unique.*
5. *Multiplicative inverses are unique.*

Proof.

1. Assume $0_F, 0'_F$ both satisfy the additive identity axiom. Observe that $0_F = 0_F + 0'_F = 0'_F$.
2. As $0_F = 0_F + 0_F$, we may write $a \cdot 0_F = a \cdot (0_F + 0_F) = a \cdot 0_F + a \cdot 0_F$. Now add the additive inverse of $a \cdot 0_F$ to both sides to obtain $0_F = a \cdot 0_F = 0_F \cdot a$.
3. Let $a \in F$ and suppose b and c are such that $a + c = 0_F$ and $a + b = 0_F$. Then $a + c = a + b$, and we may add b (or c) to both sides to obtain $b+(a+b) = (b+a)+b = 0_F+b = b$ and $c+(a+c) = (c+a)+c = 0_F+c = c$. Thus $b = c$.
4. Suppose 1_F and $1'_F$ are multiplicative identities. Then $1_F = 1_F 1'_F = 1'_F$.
5. Let $a \in F$ be nonzero and suppose b, c are multiplicative inverses for a. Then $ab = 1_F = ac$, and we may multiply by b (or c) to obtain $b(ab) = (ba)b = 1_F b = b = b(ac) = (ba)c = 1_F c = c$.

■

One consequence of Theorem **??** 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.

Definition 2.1.9 Let K be a field. If F is a nonempty subset satisfying

1. $F \subseteq K$ and
2. F is a field under the same operations as K,

then we call F a **subfield** of K, and K an **extension field** of F, and say "K/F is a **field extension**." ◊

Checkpoint 2.1.10 The set of complex numbers $\mathbb{C}$ is an extension field of $\mathbb{R}$.

Theorem 2.1.11 *For all primes p, $\mathbb{Z}_p$ is a field.*

2.2 Rings

In the previous section, we observed that many familiar number systems are fields but that some are not. In this section we will explore which field properties usually hold for familiar sets of numbers and which ones are more specialized. Before we proceed with that endeavor we will give a formal definition of polynomial so that we can include it in our exploration.

Definition 2.2.1 Let A be a set, 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 $n \in \mathbb{N}_0$ is the **degree** of the polynomial p, denoted $\deg(p) = n$, and $a_0, a_1, \ldots, a_n$ are 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. If A has an additive identity 0, then $A[x]$ has a **zero polynomial**, the polynomial whose coefficients are all 0. The zero polynomial is also often denoted by 0. The degree of the zero polynomial is $-\infty$. $\Diamond$

Checkpoint 2.2.2 Give some examples of polynomials in $A[x]$ for various choices of number systems A. Identify their coefficients and degrees.

Checkpoint 2.2.3 In the following table, fill in a Y if the set has the property; fill in a N if it does not.

	$\mathbb{N}$	$\mathbb{Z}$	$2\mathbb{Z}$	$\mathbb{Q}$	$\mathbb{Q}[x]$	$\mathbb{Z}_8$	$\mathbb{Z}_2$	$\mathbb{R}$	$\mathbb{C}$	$\mathcal{M}_2(\mathbb{R})$
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										

Table 2.2.4: A list of properties and sets.

Question 2.2.5 Which of the field axioms 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 Question **??** and the completed Table **??**, we make the following definition.

Definition 2.2.6 A ring R is a nonempty set, together with binary operations $+$ and $\cdot$ 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 $a \in R$, $a+0_R = 0_R+a = a$[1].

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 $a, b, c \in R$, $(a + b) \cdot c = a \cdot c + b \cdot c$.

$\Diamond$

Checkpoint 2.2.7 Compare and contrast Definitions **??** and Definition **??**. What are the similarities? What are the differences?

While rings seem even more complicated than fields, they are incredibly useful even in applied mathematics (see, e.g., (((Unresolved xref, reference "Curto2013"; check spelling or use "provisional" attribute)))).

Definition 2.2.8 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 $a \in R$, $a \cdot 1_R = 1_R \cdot a = a$. $\Diamond$

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).

Checkpoint 2.2.9 Consider the sets given in Table **??**. Which are rings? Which are commutative rings with identity?

Question 2.2.10 Which properties of fields in Theorem **??** hold for (commutative) rings? $\square$

Question 2.2.11 Are all rings fields? Are all fields rings? Justify. $\square$

Checkpoint 2.2.12 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[2].

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 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 § **??** for more on the historical development of the proof of Fermat's Last Theorem). Most modern definitions of *ring* agree with our Definition **??** and allow for rings with noncommutative multiplication and no multiplicative identity.

[1] When the ring R is clear from context, we will often write 0 in place of 0_R.
[2] Sometimes called a *rng*. ☺

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 multiplication) is tedious, and thus is omitted.

Theorem 2.2.13 *If R is a (commutative) ring (with identity 1_R), then $R[x]$ is a (commutative) ring (with identity 1_R).*

One of the ways to better understand mathematical structures is to understand their similar substructures (e.g., subspaces of vector spaces in linear algebra). Again, this is a formalization of an idea that you have been using since you were a child. There are certain properties of the rational numbers that you understand because you understand the real numbers - this is because the rational numbers are a subring of the real numbers.

In our study of rings, we are primarily interested in special types of subrings, to be studied in more depth in Section **??**.

Definition 2.2.14 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. $\Diamond$

The following theorem provides a nice encapsulation of what needs to be checked to see if a subset S of a ring R is in fact a subring of R.

Theorem 2.2.15 *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.*

Checkpoint 2.2.16 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. $S = \mathbb{R}$, $R = \mathbb{C}$

Definition 2.2.17 Let R be a ring and let $u \in R$ be nonzero. If there is a $v \in R$ such that $uv = vu = 1_R$, 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 $x = uy$. $\Diamond$

That is, a unit in a ring is a nonzero element with a multiplicative inverse.

Checkpoint 2.2.18 Explicitly describe the set $\mathbb{Z}^\times$. What are the associates of 7 in $\mathbb{Z}$?

Theorem 2.2.19 *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. ■

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

Definition 2.2.20 A **zero divisor** in a ring R is a nonzero element $z \in R$ such that there is a nonzero $x \in R$ with $xz = zx = 0$. ◊

Notice that the reason the idea of zero divisors at first appears weird is that they're not something we encounter when working with numbers. 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 (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 doesn't persist in our abstraction from the integers to rings in general.

Checkpoint 2.2.21 Find, with justification, the zero divisors in $\mathbb{Z}_{10}$ and $\mathbb{Z}_{11}$. Make and prove a conjecture about 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 $\gcd(x, m) \neq 1$.

Question 2.2.22 Are there any other rings in which you've seen zero divisors?

Answer. Matrix rings? □

Theorem 2.2.23 *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 2.2.24 *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 ■

Question 2.2.25 How can we reinterpret Question **??** in light of our new language of units and zero divisors? State a theorem that uses this new language.

Answer. 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 and only 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 (for the sake of brevity), 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 useful in many realms of mathematics, we will focus our study on rings with no zero divisors. As these rings share many (though not all) properties of the integers, they are known as integral domains.

Definition 2.2.26 A commutative ring with identity R is an **integral domain**, or just **domain**, if R has no zero divisors. ◊

You have already seen several domains.

Checkpoint 2.2.27 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 2.2.28 *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 2.2.29 *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 **??**.

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 2.2.30 *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 2.2.31 *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. ■

Question 2.2.32 Is the converse of Theorem **??** true? If so, give a short proof. If not, find a counterexample.

Answer. Yes. Apply Theorem **??**. □

When considering sets of polynomials, as we do in § **??**, the following results will be quite useful.

Theorem 2.2.33 *Let F be a field, and let $p(x), q(x) \in F[x]$ 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 $a_n b_m x^{n+m}$. Thus, $\deg(p(x)q(x)) = n + m = \deg(p(x)) + \deg(q(x))$. ■

Question 2.2.34 What are the units of $F[x]$? Prove your answer.

Answer. The units are $F^\times$.

Suppose $p(x), q(x) \in F\times$ such that $p(x)q(x) = 1$. Then $\deg p(x) + \deg q(x) = \deg(p(x)q(x)) = \deg(1) = 0$. Thus $\deg p(x) = \deg q(x) = 0$. □

2.3 Divisibility in Integral Domains

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. One such idea is that of 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 $c \neq 0_R$ 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. ■

Thus, 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 in rings! 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. ◊

Our notion of primality also extends nicely to domains.

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.

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. ◊

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

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 **??**.

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. ■

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. ■

It is not the case in all domains that irreducibles are primes, as the next exercise illustrates.

Checkpoint 2.3.7 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.) 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 $f \nmid x^3$.

Definition 2.3.8 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 $e \mid d$.

◊

Theorem 2.3.9 *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.*

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. ■

Checkpoint 2.3.10 In most familiar domains, GCDs exist. However, they don't always! Find an example of elements in the ring from Exercise **??** 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 $x^3 \nmid x^2$.

Checkpoint 2.3.11 Fill in the following blanks in order of increasing generality with the words *ring*, *integral domain*, *field*, and *commutative ring*.

\makebox[0.75in]{\hrulefill} ⇒ \makebox[0.75in]{\hrulefill} ⇒ \makebox[0.75in]{\hrulefill} ⇒ \makebox[0.75in]{\hrulefill}

2.4 Principal Ideals

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 $rx \in I$.

◊

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 **??** 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 $I = R$, $1_R \in R$ 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 $0_R = 0_R a \in \langle a \rangle$.

Moreover, if $r_1 a, r_2 a \in \langle a \rangle$, $r_1 a - r_2 a = (r_1 - r_2)a \in \langle a \rangle$. 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 the *generator* of $\langle a \rangle$.

Question 2.4.6 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: "$z \in \langle x \rangle \Leftrightarrow$ \makebox[0.75in]{\hrulefill}" and "$\langle x \rangle \subseteq \langle y \rangle \Leftrightarrow$ \makebox[0.75in]{\hrulefill}."

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.

Theorem 2.4.7 *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. ■

Checkpoint 2.4.8 In $R = \mathbb{Z}$, describe the principal ideals generated by

1. 2
2. -9
3. 9

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

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

Theorem 2.4.10 *The ring $\mathbb{Z}$ is a principal ideal domain.* ***Hint****. Use properties specific to $\mathbb{Z}$, perhaps from Section* **??***.*

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 $x \in \langle\, d\,\rangle$. ■

Checkpoint 2.4.11 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 2.4.12 *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 § **??**. 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 2.4.13 Let R be an integral domain. We call R a **Euclidean domain** if there is a function $\delta : R\backslash\{0_R\} \to \mathbb{N}_0$ such that: \marginpar{\setstretch{0.7}**This is the lowercase Greek letter delta.**}

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 $r = 0_R$ 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.

Question 2.4.14 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 2.4.15 *Let $S \subseteq F[x]$ be a set containing a nonzero polynomial. Prove that S contains a polynomial f such that $\deg(f) \leqslant \deg(g)$ for all $g \in S$.*

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 2.4.16 *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 $f(x) = a_0 + a_1 x + \cdots + a_m x^m$ and $g(x) = b_0 + b_1 x + \cdots + b_n x^n$, then $h(x) = f(x) - a_m b_n^{-1} x^{m-n} g(x)$ has degree strictly*

less than $\deg f(x)$.

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

Theorem 2.4.17 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* **??**

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 $f(x) = a_0 + a_1 x + \cdots + a_m x^m$ and $g(x) = b_0 + b_1 x + \cdots + b_n x^n$. Use Lemma **??** and set $h(x) = f(x) - a_m b_n^{-1} x^{m-n} g(x)$, 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

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

where $q = a_m b_n^{-1} x^{m-n} + q_1(x)$ 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 2.4.18 *The ring $F[x]$ is a principal ideal domain.* ***Hint.*** *Mimic the proof of Theorem* **??** *and use Lemma* **??***!*

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 **??** to write $g(x) = f(x)q(x) + r(x)$, where $\deg r(x) < \deg f(x)$ or $r(x) = 0$. As in Theorem **??**, 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 $g(x) \in \langle f(x) \rangle$. ■

Question 2.4.19 Is $F[x]$ 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. It's the degree function. □

Question 2.4.20 Where do Euclidean domains and PIDs fit in the hierarchy of abstraction found in Question **??**? □

Chapter 3

Factorization

3.1 Factoring Polynomials

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 § **??** to $F[x]$. We will also see that the particular nature of a polynomial allows for additional criteria for, e.g., its irreducibility to be discovered.

We saw in Theorem **??** and Theorem **??** that both $\mathbb{Z}$ and $F[x]$ have a division algorithm. It is reasonable (and correct) to expect that similar to the integers we can also investigate the greatest common divisor of polynomials.

Checkpoint 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 $\mathbb{Z}$.

Checkpoint 3.1.2 Carefully state and prove a Bézout-like theorem (recall Theorem **??**) 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 $f(x)s(x) + g(x)t(x) = \gcd(f(x), g(x))$.

We next make the somewhat pedantic — yet valuable — distinction between the polynomial $p(x) \in R[x]$ and the function $p_f : R \to R$ defined by the polynomial.

Definition 3.1.3 Let R be a commutative ring with identity and $p(x) = a_0 + a_1x + \cdots + a_nx^n \in R[x]$. A **polynomial function** $p_f : R \to R$ is defined, for all $r \in R$, by $p_f(r) = a_0 + a_1r + \cdots + a_nr^n \in R$. We usually denote $p_f(r)$ by $p(r)$, and say that $p(r)$ is the polynomial p evaluated at r. ◊

Definition 3.1.4 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 $p(r) = 0_R$. ◊

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 3.1.5 *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 $x-\alpha | p(x)$. ■

Checkpoint 3.1.6 Given a field F, define an irreducible element of $F[x]$, keeping in view Theorem **??** and Definition **??**.

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

Checkpoint 3.1.8 State a positive definition for a reducible polynomial with coefficients in a field F.

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 3.1.9 *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 3.1.10 *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. ■

Theorem 3.1.11 *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 $f(x)g(x)$) . ■

Theorem 3.1.12 *Every nonconstant polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$.*

Theorem 3.1.13 *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. ■

3.2 Factorization in Euclidean Domains

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.

$\Diamond$

Checkpoint 3.2.2 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.

Theorem 3.2.3 *Every field is a UFD.*

Our next main goal is to prove Theorem **??**. In order to do that, we need some information about the ideal-theoretic structure of Euclidean domains.

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

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

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

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 "clear" that I 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 $k \geqslant j$. ■

Lemma 3.2.6 *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 3.2.7 *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 **??** for all factors of r, and thus r can be factored into a product of irreducibles. ■

Lemma 3.2.8 *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 $1 = as + bt$.*

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 3.2.9 *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. ■

Theorem 3.2.10 *Let R be a PID and $p \in R$ be irreducible. If $a, b \in R$ with $p \mid ab$, then $p \mid a$ or $p \mid b$.*

Proof. Irreducible implies prime (use the previous theorem). ■

Theorem 3.2.11 *Every PID is a UFD.*

Proof. Let R be a PID and suppose that a non-zero non-unit element $a \in R$ can be factored in two different ways as a product of irreducibles.

$$a = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s$$

Assume that $s \geqslant r$. Since p_1 is irreducible (and therefore prime), there exists j such that $p_1 \mid q_j$. Since q_j is irreducible as well, there is a unit u_1 such that $p_1 = u_1 q_j$. Up to reordering we have

$$p_1 p_2 \cdots p_r = u_1 p_1 q_2 \cdots q_s$$

Therefore

$$p_2 \cdots p_r = u_1 q_2 \cdots q_s$$

We can continue in a similar fashion until we end up with

$$1 = u_1 \cdots u_r q_{r+1} \cdots q_s$$

Since the q_j's are non-units, this means that $r = s$ and that the p_i's are associates of the q_j's. Therefore R is a UFD. ■

Theorem 3.2.12 *Every Euclidean domain is a principal ideal domain.**Hint.** Mimic the proof of Theorem* **??**.

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. ■

Theorem 3.2.13 *Every Euclidean domain is a unique factorization domain.*

Proof. Apply Theorems **??** and Theorem **??**. ■

Theorem 3.2.14 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.

Theorem 3.2.15 *If R is a UFD, then $R[x]$ is a UFD.*

Thus, $\mathbb{Z}[x]$ is a UFD. However, as we will see later, $\mathbb{Z}[x]$ is not a PID.

3.3 Nonunique Factorization

If every ring had the unique factorization property, life would be very boring indeed. And 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 integer solutions to the equation $x^n + y^n = z^n$ if $n \geqslant 3$.

In 1847, Gabriel Lamé claimed he had complete 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.

In the following exercises, we explore factorization in a similar integral domain, $R = \mathbb{Z}[\sqrt{-7}] = \left\{a + b\sqrt{-7} : \ a, b \in \mathbb{Z}\right\}$.

Checkpoint 3.3.1 Verify that $8 = (1 + \sqrt{-7})(1 - \sqrt{-7})$.

Checkpoint 3.3.2 Define a function $\delta : R \to \mathbb{N}_0$ by $\delta(a + b\sqrt{-7}) = a^2 + 7b^2$. Prove that if $x, y \in R$, $\delta(xy) = \delta(x)\delta(y)$.

Theorem 3.3.3 *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 3.3.4 *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 3.3.5 *The elements 2,* $1 + \sqrt{-7}$*, and* $1 - \sqrt{-7}$ *are irreducible in R. Thus, 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

4.1 Ideals in general

Theorem 4.1.1 *Let R be commutative with identity 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 $rx \in I$. ■

We can extend the notion of a principal ideal to ideals generated by any subset.

Definition 4.1.2 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. ◊

One way to think of $\langle\, S\,\rangle$ is given in the following theorem.

Theorem 4.1.3 *Given a commutative ring R and a subset S of R, $\langle\, S\,\rangle$ is the smallest ideal containing S.*

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

Theorem 4.1.4 *Given a commutative ring with identity R and a subset S of R,*

$$\langle S\rangle = \{r_1s_1 + r_2s_2 + \cdots + r_ns_n :\ r_i \in R,\ s_j \in S,\ n \geqslant 0\}.$$

Hint. *Begin by showing that $I = \{r_1s_1 + r_2s_2 + \cdots + r_ns_n :\ r_i \in R,\ s_j \in S,\ n \geqslant 0\}$ is an ideal containing S.*

Proof. For the ease of notation, let

$$I = \{r_1s_1 + r_2s_2 + \cdots + r_ns_n :\ r_i \in R,\ s_j \in S,\ n \geqslant 0\}.$$

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 $x = r_1s_1 + \cdots + r_ns_n$. Each $s_i \in S$, so each $r_is_i \in J$ if J is any ideal containing S. In articular, $r_is_i \in \langle\, S\,\rangle$,

and thus the sum $x = \sum r_i s_i \in \langle S \rangle$. ■

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

Proof. Suppose that $I = \langle a(x) \rangle$. Since $2 \in I$, $2 = c(x)a(x)$ for some $c(x) \in \mathbb{Z}[x]$, so $\deg c(x) = \deg a(x) = 0$, and thus $a(x) = a \in \mathbb{Z}$. We also have $x \in I$, so $x = b(x)a$, which implies that $\deg b(x) = 1$, so $b(x) = b_1 x + b_0$ for some $b_1, b_0 \in \mathbb{Z}$. Then $x = (b_1 x + b_0)a$ implies $b_1 a = 1$. Thus, $a = 1$ or $a = -1$. In either case, a is a unit in $\mathbb{Z}[x]$ so $1 \in I$. By Theorem **??**, there are polynomials $p(x), q(x) \in \mathbb{Z}[x]$ such that $1 = 2p(x) + xq(x)$. If p_0 is the constant term of $p(x)$, it follows that the constant term of $2p(x) + xq(x)$ is $2p_0 = 1$, which is a contradiction.

Thus, I is not principal. ■

4.2 Quotient Rings

Definition 4.2.1 Let R be a ring and I an ideal of R. Then elements $r, s \in R$ are said to be **congruent modulo** I if $b - a \in I$. If this is the case, we write $a + I = b + I$. ◊

Question 4.2.2 Given a ring R, ideal I, and $r \in R$, when is it the case that $r + I = 0 + I = I$?

Answer. When $r \in I$. □

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

Theorem 4.2.3 *Let R be a ring and I an ideal of R. Then congruence modulo I is an equivalence relation on R.*

The set of equivalence classes under this relation is denoted R/I.

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. ■

Theorem 4.2.4 *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 $(a + c) + I = (b + d) + I$.*

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 4.2.5 *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.

Theorem 4.2.6 *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 $1_R + I$.*

Checkpoint 4.2.7 Let $R = \mathbb{Z}$ and $I = \langle 7 \rangle$. Find at least three distinct integers x such that $9 + I = x + I$.

Solution. Such x include $-5, 2, 16$, and anything in $\{2 + 7m : m \in \mathbb{Z}\}$.

Definition 4.2.8 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 $b \in P$. $\Diamond$

Theorem 4.2.9 *Let R be a domain and $p \in R$ be prime. Then $\langle p \rangle$ is a prime ideal.*

Checkpoint 4.2.10 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 $\mathbb{C}$. $\ddot\smile$ 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.

Definition 4.2.11 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 $J = R$. $\Diamond$

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 these notes, 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).

Theorem 4.2.12 *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. ■

Lemma 4.2.13 *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 $1 = xr' + y'$.*

Theorem 4.2.14 *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.*

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 4.2.15 *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. ■

Theorem 4.2.16 *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. ■

Find a non-maximal prime ideal.**Solution**. In $\mathbb{Z}[x]$, it is straightforward to see that $\langle x \rangle$ is prime. Moreover, $\langle x \rangle \subsetneq \langle 2, x \rangle$, and thus is not maximal.

4.3 Homomorphisms

Definition 4.3.1 Let R and S be commutative rings with identity. A function $\varphi : R \to S$ is a called **ring homomorphism** if it preserves addition and multiplication. That is, for all $x, y \in R$, $\varphi(x+y) = \varphi(x) + \varphi(y)$ and $\varphi(xy) = \varphi(x)\varphi(y)$. 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. ♢

Checkpoint 4.3.2 Show that the function $\varphi : Z \to \mathbb{Z}_5$ defined by $\varphi(x) = \overline{x}$ is a homomorphism. Is it an isomorphism? Why or why not?

Solution. 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 $\varphi(2) = \varphi(7) = \overline{2}$.

Theorem 4.3.3 *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.*

Checkpoint 4.3.4 Let $R = \mathbb{Z}_2[x]$ and consider the map $F : R \to R$ given by $F(p) = p^2$. Is F a homomorphism? An isomorphism?

Checkpoint 4.3.5 Give an example of other rings R and S and a homomorphism $\varphi : R \to S$. Show that your function is a homomorphism and determine whether or not it is an isomorphism. Multiple presentations (of different homomorphisms) are possible.

Solution. One is complex conjugation, which is an isomorphism $\mathbb{C} \to \mathbb{C}$. Another is $\varphi : \mathbb{Z}_5 \to \mathbb{Z}_{10}$ given by $\varphi(x) = 5x$ (this isn't easy).

Definition 4.3.6 Let $\varphi : R \to S$ be a ring homomorphism. Then $\ker\varphi = \{r \in R : \varphi(r) = 0_S\}$ is the **kernel** of φ. $\Diamond$

Checkpoint 4.3.7 Let R be commutative with identity and $r \in R$. The *r-evaluation homomorphism* is the map $\varphi_r : R[x] \to R$ defined by $\varphi_r(p(x)) = p(r)$. Prove that φ_r is a homomorphism. What is $\ker\varphi_r$?

Solution. This follows from the way sums and products of polynomials are evaluated. By definition, $\ker\varphi_r$ is the set of all $p(x) \in R[x]$ such that $p(r) = 0$, that is, the set of all polynomials which have r as a zero.

Checkpoint 4.3.8 Find the kernels of other homomorphisms we've looked at so far.

Theorem 4.3.9 *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 $r \in R$, $\varphi(rx) = \varphi(r)\varphi(x) = \varphi(r)0_S = 0_S$. ■

Theorem 4.3.10 *Let $\varphi : R \to S$ be a homomorphism. Then φ is one-to-one if and only if $\ker\varphi = \{0\}$.*

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 4.3.11 *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 4.3.12 *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 4.3.13 *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 4.3.14 *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 4.3.15 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, $R/\ker\varphi \cong S$.

Theorem 4.3.16 *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. $\mathbb{R}[x]/\left\langle x^2+1 \right\rangle \cong \mathbb{C}$

Proof. 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, $\mathbb{R}[x]/\ker\varphi \cong \mathbb{C}$.

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

Index