algebra party time
rings edition

nathan bloomfield

Contents

Preface

These are the lecture notes of a first abstract algebra course the author teaches at the Northeastern State University of Oklahoma. Each section is intended to be one-half to one class meeting's worth of material.

As these are functionally lecture notes, used in a specific class, the main expository sections develop only the basic theory of rings as dictated by the needs of my students. Plenty of supplementary material, including some important definitions and theorems, is relegated to the "exercises". The exercises come in a few flavors: some are straight computation, some ask for examples or counterexamples, and some require a proof. Some exercises – my favorites – will guide you through the proofs of reasonably high-powered theorems. Motivated readers are encouraged to work as many of these as they can stand.

Some additional supplementary material, appropriate to the main text but cut from a typical class due to time constraints, is given in the appendices. These include constructions of the real, complex, and p-adic numbers, Zorn's lemma, and a brief discussion of categories.

Introduction

These notes develop the basic theory of **rings**, which grew over time out of several historical threads. Mathematicians of "antiquity" (defined here as the period from prehistory to the Enlightenment) were broadly interested in solving equations of several different kinds. For example, Diophantine equations like $a^2+b^2=c^2$, of which we are only interested in integer solutions. Or polynomial equations such as $ax^2+bx+c=0$, where a, b, and c are known constants. Or systems of simultaneous linear equations. Over time techniques were discovered for understanding various types of equations, and by the 19th Century these techniques had developed into fairly sophisticated branches of mathematics. Diophantine equations became the motivation for a nascent Algebraic Number Theory, whose great white whale was Fermat's Last Theorem. The rich history of polynomial equations reached an apex with the Abel-Ruffini Theorem on the insolubility of the general quintic equation. And the study of simultaneous linear equations evolved into what we now call Linear and Multilinear Algebra at the hands of Hermann Grassmann and was used to simplify many ideas in geometry and physics. Underlying these more-or-less parallel developments was

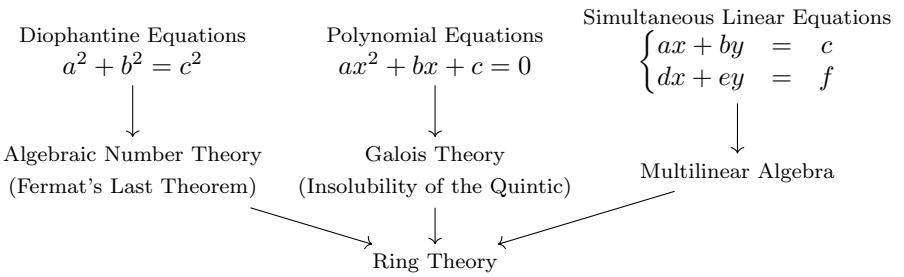


Figure 1: A grossly oversimplified genealogy of ring theory.

a common abstraction, which we call a *ring*, and which was only made explicit in the first couple of decades of the 20th Century. **The essence of 'ringhood' is the structure – the arithmetic – that the integers, polynomials, and square matrices have in common.**

The content of these notes is usually covered in a class called "Abstract

Algebra", a name which I dislike for two reasons. First, **all** mathematics is abstract. Yes, even the so-called "applied" branches. (Sorry if anyone is disappointed by this fact!) Show me where the number two exists in nature, or a continuous function, or a vector space, and I will reconsider this assertion. To single out only some ideas as being abstract is not a useful distinction, and can even hurt. How many of our students reflexively recoil at abstract subjects as being inherently pointless and less useful or interesting? The second reason for my dislike of the name "abstract algebra" is that, reason one notwithstanding, algebra is jam-packed with concrete objects which we can fiddle with and compute and draw pictures of. In fact the whole purpose of the "abstract" part is so that we can better understand the "concrete" part.

Yet the name persists, and is not likely to go away soon (if for no other reason than that it takes an act of nature to change a university catalog). This name reflects a central tension between two broad points of view in mathematics: the abstract and the concrete. The abstract point of view prefers theory-building; it likes for theorems to be as general as possible; and it doesn't mind nonconstructive proofs, especially if they are slick or insightful. The concrete point of view really likes constructive results; it tries to keep an eye on specific examples; and is never happier than when a proof can be interpreted as a clear and efficient algorithm. Neither point of view has a monopoly on enlightenment. The abstract point of view needs interesting concrete examples to abstract away from, and the concrete point of view is quickly bogged down in irrelevant details without abstraction. In these notes I have tried to balance these two perspectives.

— 0 —

Review of Arithmetic

A large amount of the material in this text is motivated by some basic ideas in number theory. In the interest of remaining self-contained, this chapter very briefly covers a few of these ideas. Some will be familiar, some not. The proofs here can be safely skipped over on the first reading.

1 The Integers

We begin with a formal definition of the set of numbers we call *integers*.

Axiom 1.1 (The Integers)**.** There is a set $\mathbb{Z}$, whose elements are called *integers*, which is equipped with two special distinct elements 0 (called *zero*) and 1 (called *one*), two binary operations $+$ (called *plus*) and $\cdot$ (called *times*, and typically not written explicitly) and a binary relation $\leq$ (pronounced "is less than or equal to") which together satisfy the following properties.

A1. Plus is *associative*: $(a+b)+c = a+(b+c)$ for all $a, b, c \in \mathbb{Z}$.

A2. Zero is *neutral* with respect to plus: $a+0 = 0+a = a$ for all $a \in \mathbb{Z}$.

A3. Every integer has an *additive inverse*: for every $a \in \mathbb{Z}$ there is an element $-a \in \mathbb{Z}$ (called a *negative* of a) such that $a+(-a) = (-a)+a = 0$.

A4. Plus is *commutative*: $a+b = b+a$ for all $a, b \in \mathbb{Z}$.

M. Times is *associative*: $(ab)c = a(bc)$ for all $a, b, c \in \mathbb{Z}$.

D. Times *distributes over* plus from either side: $a(b+c) = ab+ac$ and $(b+c)a = ba+ca$ for all $a, b, c \in \mathbb{Z}$.

C. Times is *commutative*: $ab = ba$ for all $a, b \in \mathbb{Z}$.

U. One is *neutral* with respect to times: $1 \cdot a = a \cdot 1 = a$ for all $a \in \mathbb{Z}$.

P1. $a \leq a$ for all $a \in \mathbb{Z}$.

P2. If $a \leq b$ and $b \leq c$, then $a \leq c$, for all $a, b, c \in \mathbb{Z}$.

P3. If $a \leq b$ and $b \leq a$, then $a = b$ for all $a, b \in \mathbb{Z}$.

T. If $a, b \in \mathbb{Z}$, then either $a \leq b$ or $b \leq a$.

O1. If $a \leq b$ then $a+c \leq b+c$ for all $a, b, c \in \mathbb{Z}$.

O2. If $0 \leq a$ and $0 \leq b$ then $0 \leq ab$.

O3. $0 \leq 1$.

WOP. Suppose S is a nonempty subset of $\mathbb{Z}$ which is bounded below; say $b \in \mathbb{Z}$ has the property that $b \leq s$ for all $s \in S$. Then there is an integer $m \in S$ such that $m \leq s$ for all $s \in S$. That is, every nonempty subset of $\mathbb{Z}$ which is bounded below has a least element. This is called the *well-ordering property* of $\mathbb{Z}$.

As usual, $a < b$ is short for $a \leq b$ and $a \neq b$; $a \geq b$ is equivalent to $b \leq a$, and $a > b$ is equivalent to $b < a$.

It may seem silly to introduce such a large "axiom". Indeed, it is possible to build up the integers out of simpler objects. However, doing so is not necessary

for our purposes. In developing any axiomatic system there is a balance to be made between *simplicity of axioms* and *distance to interesting results*. Taking the existence of $\mathbb{Z}$ as an axiom does not sacrifice too much simplicity – it is "only" one axiom – but allows us to say useful things very quickly.

Most of the properties of $\mathbb{Z}$ are familiar. The least intuitive is probably the last one, the well-ordering property. While strange at first, this property is extremely important. For example, we can use it to establish this "obvious" result.

Prop'n 1.2. There is no integer t such that $0 < t$ and $t < 1$.

Proof. Let $S = \{t \in \mathbb{Z} \mid 0 < t \text{ and } t < 1\}$, and suppose that S is not empty. Note that S is bounded below by 0 by definition; if $t \in S$ then $0 < t$. By WOP, then, S must have a least element; say m. Since $0 < m$, we have $0 < mm$ by 1.1. On the other hand, since $0 < m$ and $m < 1$, we have $mm < m$ by 1.1. That is, $mm \in S$, and $mm < m$ – a contradiction. □

It is not too hard to show that every integer a satisfies exactly one of $a > 0$, $a = 0$, and $a < 0$, a result we call the Trichotomy property. As usual if $a > 0$ we say a is *positive* and if $a < 0$ we say a is *negative*, and zero is neither positive nor negative. The nonnegative integers are special enough that we give them a name.

Def'n 1.3 (The Natural Numbers)**.** We denote by $\mathbb{N}$ the set of all integers n such that $n \geq 0$. Elements of $\mathbb{N}$ are called *natural numbers*.

Now the natural numbers satisfy a nice structural property.

Prop'n 1.4. If $n \in \mathbb{N}$, then either $n = 0$ or $n = m + 1$ for some $m \in \mathbb{N}$.

Proof. Suppose $n \in \mathbb{N}$ and $n \neq 0$; then $n > 0$. Let $m = n - 1$. Now if $m < 0$, we have $n - 1 < 0 < n$, and thus $0 < 1 - n < 1$, a contradiction. By trichotomy, either $m = 0$ or $m > 0$; that is, $m \in \mathbb{N}$. □

The well-ordering property of $\mathbb{Z}$ gives us a very powerful tool for working with natural numbers, called the Principle of Mathematical Induction.

Prop'n 1.5 (Principle of Mathematical Induction)**.** Let $B \subseteq \mathbb{N}$. If B satisfies the following two properties:

(i) $0 \in B$, and

(ii) If $n \in B$, then $n + 1 \in B$;

then $B = \mathbb{N}$.

Proof. We will prove this result by contradiction. Let $S = \{n \in \mathbb{N} \mid n \notin B\}$, and suppose S is not empty. Then by the Well-Ordering Property, S has a least element; say t. Since $t \in \mathbb{N}$, either $t = 0$ or $t = u + 1$ for some $u \in \mathbb{N}$. Since $0 \in B$, it must be the case that $t = u + 1$. Note that since $u < t$, and t is minimal among the natural numbers which are not in B, we have $u \in B$. But then $t = u + 1 \in B$, a contradiction. So in fact S is empty and we have $B = \mathbb{N}$. □

The Principle of Mathematical Induction (also called just "induction" or PMI) gives us a straightforward way to show that a given statement is true for all natural numbers. Proofs using PMI require two steps: the Base Case ($0 \in B$) and the Inductive Step (if $n \in B$ then $n + 1 \in B$). Most importantly, constructive proofs by induction can be turned into *recursive algorithms* which actually compute things. This is a powerful idea with implications far beyond numbers.

* * EXERCISES * *

1.1. **Trichotomy.** Show that if $a \in \mathbb{Z}$, then exactly one of $a < 0$, $a = 0$, or $a > 0$ is true.

1.2. Show that the following hold for all $a, b \in \mathbb{Z}$.

(i) If $a \geq 0$ and $b \leq 0$ then $ab \leq 0$.

(ii) If $a \leq 0$ and $b \leq 0$ then $ab \geq 0$.

1.3. Show that if a and b are integers and $ab = 1$, then either $a = b = 1$ or $a = b = -1$. (Hint: Apply Exercise 1.1 to a.)

1.4. Show that if $n \in \mathbb{N}$ then

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

1.5. Show that if $n \in \mathbb{N}$ then

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

1.6. Show that if $n \in \mathbb{N}$ then

$$\sum_{k=1}^{n} (2k-1) = n^2.$$

1.7. Show that if $n \in \mathbb{N}$ then

$$\sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}.$$

Def'n 1.6 (Absolute Value)**.** Given $n \in \mathbb{Z}$, the *absolute value* of n, denoted $|n|$, is n if $n \geq 0$ and is $-n$ if $n < 0$.

1.8. Show that for any integer a we have

$$-|a| \leq a \leq |a|.$$

1.9. Show that the following hold for all integers a and b.

(i) $|ab| = |a| \cdot |b|$

(ii) $|a + b| \leq |a| + |b|$

1.10. **Strong Induction.** Let $B \subseteq \mathbb{N}$ and suppose B satisfies the following two properties.

(i) $0 \in B$, and

(ii) If $n \in \mathbb{N}$ such that $a \in B$ for all $0 \leq a \leq n$, then $n + 1 \in B$;

Show that $B = \mathbb{N}$.

Def'n 1.7 (Integer Interval)**.** Let a and b be integers. The set

$$[\![a, b]\!] = \{x \in \mathbb{Z} \mid a \leq x \leq b\}$$

is called an *interval*.

Def'n 1.8 (Finite)**.** We say that a set S is *finite* if there is a natural number n and a bijection $\varphi : [\![1, n]\!] \to S$.

1.11. Let a and b be integers. Show that $[\![a, b]\!]$ is finite.

2 The Division Algorithm

Theorem 2.1 (Division Algorithm)**.** If a and b are integers with $b > 0$, then there exist integers q and r such that $a = qb + r$ and $0 \leq r < b$.

The q and r given by the Division Algorithm are called the *quotient* and *remainder*, respectively. We will give two proofs of this result, one constructive and one nonconstructive, and compare them.

Proof. **Nonconstructive.** Our strategy is to use the Well-Ordering Property. To this end, define a set S of integers as follows.

$$S = \{a - qb \mid q \in \mathbb{Z}\} \cap \mathbb{N}.$$

First, note that $S \subseteq \mathbb{N}$ by definition. Next, we claim that S is not empty. To see this, note that $-a \leq |a| \leq |a|b$, since $1 \leq b$. Rearranging, we have

$$0 \leq a + |a|b = a - (-|a|)b,$$

so that in particular $a - (-|a|)b \in S$. So S is a nonempty subset of $\mathbb{N}$. By WOP, S has a $\leq$-smallest element; say r. Now $0 \leq r$ and $r = a - qb$ for some integer q, so that $a = qb + r$.

Finally, we claim that $r < b$. To see this, assume by way of contradiction that $r \geq b$. Now $r - b \geq 0$, and moreover

$$r - b = a - qb - b = a - (q + 1)b.$$

Thus we have $r - b \in S$. However we also have $r > r - b$ (strict) since $b \geq 1$; this contradicts the minimalness of r in S. So our assumption that $r \geq b$ was false, and in fact $r < b$. Thus we have "found" integers q and r such that $a = qb + r$ and $0 \leq r < b$. □

Proof. **Constructive.** We will consider two cases separately; first with $a \geq 0$ and second with $a < 0$.

(i) Let B be the set

$$B = \{a \in \mathbb{N} \mid a = qb + r \text{ for some } q, r \in \mathbb{N} \text{ with } 0 \leq r < b\}.$$

We will show that $B = \mathbb{N}$ using induction.

- Base case: $a = 0$. Note that $0 = 0 \cdot b + 0$. Letting $q = r = 0$ we have $0 \in B$.
- Suppose that $a - 1 \in B$. That is, for any integer b, there exist integers q' and r' such that $a - 1 = q'b + r'$ and $0 \leq r' < b$. If $a = b$, then letting $q = 1$ and $r = 0$ we have $a \in B$. If $a < b$, then letting $q = 0$ and $r = a$ we have $a \in B$. Otherwise, there are two possibilities: either $r' + 1 < b$ or $r' + 1 \geq b$.

Case 1: Suppose $r' + 1 < b$. In this case we have $a = q'b + r' + 1$ and $0 \leq r' + 1 < b$. Letting $q = q'$ and $r = r' + 1$ we have $a \in B$.

Case 2: Suppose $r' + 1 \geq b$. Then in fact we must have $r' + 1 = b$. Now $a = q'b + r' + 1 = q'b + b = (q' + 1)b$. Letting $q = q' + 1$ and $r = 0$ we have $a \in B$.

By PMI, $B = \mathbb{N}$. That is, if $a \geq 0$ then there exist integers q and r such that $a = qb + r$ and $0 \leq r < b$.

(ii) Now suppose $a < 0$, so that $-a > 0$. By the previous discussion, there exist integers q' and r' such that $-a = q'b + r'$ and $0 \leq r' < b$. There are two possibilities: either $r' = 0$ or $r' > 0$.

Case 1: Suppose $r' = 0$. Now $a = -q'b$. Letting $q = -q'$ and $r = 0$ we have $a = qb + r$ with $0 \leq r < b$.

Case 2: Suppose $r' > 0$. Now

$$a = -q'b - r' = -q'b - b + b - r' = (-q' - 1)b + b - r'.$$

Moreover, note that $0 \leq b - r' < b$; the left inequality because $r' < b$, and the right because $r' > 0$. Letting $q = -q' - 1$ and $r = b - r'$ we have q and r so that $a = qb + r$ and $0 \leq r < b$.

□

Different proofs of the same result using different strategies, as we have here, can frequently offer different insights into why the result is true.

- The constructive proof is tedious, requiring several different case analyses. (This is when a proof splits into cases.) Generally a brute-force case analysis is seen as the least enlightening kind of proof. On the other hand, this is called a constructive proof because it gives us a strategy – an *algorithm* – to actually *find* the integers q and r. Given particular a and b, we can find q and r by tracing through the cases of the proof.

- The nonconstructive proof is short and sweet (the technical term for this is *elegant*), especially compared to the constructive proof. Logically it is much less complicated. It also does not depend as much on the details of $\mathbb{Z}$, and so is easier to generalize to other situations. However, it tells us nothing about how to *find* the q and r; it guarantees their existence, but offers no computational guidance.

These two proofs are a great example of the difference between the two major points of view in algebra – abstract and concrete. Abstraction can bring elegance and concision by throwing away unnecessary detail at the expense of computability, while keeping it concrete allows us to compute in exchange for getting our hands a little dirty, so to speak. Two important corollaries of the Division Algorithm will be useful later.

Cor. 2.2. If a and b are integers with $b \neq 0$, then there exist integers q and r such that $a = qb + r$ and $0 \leq r < |b|$.

Proof. If $b < 0$, then $-b > 0$. Using the Division Algorithm find q' and r' such that $a = q'(-b) + r'$ and $0 \leq r' < -b$. Let $q = -q'$ and $r = r'$. □

Cor. 2.3. The q and r given by the Division Algorithm are unique in the following sense; if $a = qb + r$ and $a = q'b + r'$ and $0 \leq r, r' < |b|$, then $q = q'$ and $r = r'$.

Proof. Note that $-|b| < -r' \leq 0$, so that $-|b| < r - r' < |b|$. In particular we have $|r - r'| < |b|$. On the other hand we have $qb + r = q'b + r'$, so that $r - r' = b(q' - q)$ and thus $|r - r'| = |b||q' - q|$. Thus $|b||q' - q| < |b|$, and so $|q' - q| < 1$. Thus $q' = q$. Now $r = a - qb = a - q'b = r'$. □

* * Exercises * *

2.1. Write a computer program that takes two integers a and b such that $b \neq 0$ and returns integers q and r such that $0 \leq r < |b|$ and $a = qb + r$. Test your program on the following inputs.

(i) $a = 5$ and $b = 2$

(ii) $a = 0$ and $b = 11$

(iii) $a = -10$ and $b = 3$

(iv) $a = -10$ and $b = -3$

(v) $a = 3^{700}$ and $b = 2^{100}$

3 Divisors

The integers have the nice property that, for any constants a and b, the equation

$$a + x = b$$

always has a solution in $\mathbb{Z}$ – namely, $b - a$. This is no longer true if we replace the plus by times; the equation $ax = b$ may or may not have any solutions in $\mathbb{Z}$. If so, a and b enjoy a special relationship.

Def'n 3.1 (Divides)**.** Given integers a and b, we say that a *divides* b, written $a|b$, if there is an integer c such that $ac = b$. In this case we say that a is a *divisor* of b.

Divisibility satisfies several basic properties.

Prop'n 3.2. The following hold for all integers a, b, and c.

(i) $a|0$.

(ii) $1|a$.

(iii) $a|a$.

(iv) If $a|b$ and $b|c$ then $a|c$.

(v) If $a|b$ and $b \neq 0$ then $|a| < |b|$.

Note that 3.2(ii) and 3.2(iii) say that every integer has at least two divisors, though typical integers have many more. Likewise 3.2(i) says that every integer is a divisor of at least one other. If one integer is simultaneously a divisor of two others, we say it is a *common divisor*, and this is very special.

Def'n 3.3. Let a, b, c, and d be integers.

(i) We say that c is a *common divisor* of a and b if $c|a$ and $c|b$.

(ii) We say that d is a *greatest common divisor* of a and b if d is a common divisor, and if c is another common divisor, then $c \leq d$.

In fact any two integers (not both zero) *have* a greatest common divisor, which is unique.

Prop'n 3.4. Any two integers (not both zero) have a unique greatest common divisor, which we denote $\mathsf{gcd}(a, b)$. We also define $\mathsf{gcd}(0, 0) = 0$ as a special case.

Proof. First we show existence. Let $M = \max(|a|, |b|)$ and let

$$S = \{M - c \mid c \text{ is a common divisor of } a \text{ and } b\}.$$

Now S is not empty since $M - 1 \in S$, and in fact $S \subseteq \mathbb{N}$ since $c \leq M$ for all common divisors c of a and b. By WOP, then, S has a least element; say $t = M - d$. Now d is a common divisor of a and b by definition. If c is a common divisor of a and b, then $M - d \leq M - c$, which rearranges as $c \leq d$. So d is a greatest common divisor of a and b. To see uniqueness, note that if d_1 and d_2 are both greatest common divisors of a and b then we have $d_1 \leq d_2$ and $d_2 \leq d_1$ so that $d_1 = d_2$. □

Because it is unique, we can think of the greatest common divisor as a function of two integers.

Prop'n 3.5. The following hold for all integers a and b.

(i) $\mathsf{gcd}(a, b) = \mathsf{gcd}(b, a)$.

(ii) $\mathsf{gcd}(a, a) = |a|$.

(iii) If $b|a$, then $\mathsf{gcd}(a, b) = |b|$.

(iv) $\mathsf{gcd}(a, 1) = 1$.

(v) $\mathsf{gcd}(a, 0) = |a|$.

Our proof that GCDs always exist is not constructive since it depends on the Well-Ordering Property. We know that GCDs exist, but actually finding them is difficult! For example, suppose we want to find the greatest common divisor of 2 and 3. From the definition, we can restrict our search to the common divisors of 2 and 3. Using 3.2(v), we can show that the divisors of 2 are ± 1 and ± 2 and the divisors of 3 are ± 1 and ± 3. The largest number which appears in both lists is 1, so that $\mathsf{gcd}(2, 3) = 1$. This works well enough, but what if we want to find $\mathsf{gcd}(1337, 1007)$? The same brute force strategy works in principle, but would take a long time.

It is possible to give a constructive, induction based proof of Proposition 3.4 (try it!) but we will avoid doing so here. Such a proof will be longer and more tedious while providing no more insight than this one does – giving more heat than light, so to speak – and the resulting algorithm slow. Instead we give the following result which connects greatest common divisors to the division algorithm.

Prop'n 3.6 (Euclidean Algorithm for $\mathbb{Z}$)**.** If a and b are integers with $b > 0$, and if $a = qb + r$ where $0 \leq r < b$, then $\mathsf{gcd}(a, b) = \mathsf{gcd}(b, r)$.

Proof. Let $d = \mathsf{gcd}(a, b)$ and $e = \mathsf{gcd}(b, r)$. We need to show that $d = e$; to do this, we will show that $d \leq e$ and $e \leq d$.

(i) By definition we have $d|a$ and $d|b$; that is, $a = da'$ and $b = db'$ for some integers a' and b'. Now

$$r = a - qb = da' - qdb' = d(a' - qb'),$$

so that $d|r$. In particular, d is a common divisor of b and r, and so $d \leq e$.

(ii) Similarly, we have $e|b$ and $e|r$, so that $e|a$, and thus $e \leq d$. □

The Euclidean Algorithm gives us a way to explicitly compute the GCD of two integers *as long as* we can compute quotients and remainders as in the Division Algorithm; in fact, it is quite fast. Note that since r is strictly less than b, this recursion must eventually terminate with a statement of the form $\gcd(a, 0)$. This result is named after *the* Euclid, because as far as we know he was the first to write it down – in his book *The Elements of Geometry.* Euclid himself thought of the algorithm in geometric terms, where a and b represent the lengths of line segments and $\gcd(a, b)$ the length of the largest line segment that can "tile" both simultaneously.

Theorem 3.7 (Bezout's Identity)**.** If a and b are integers, then there exist integers u and v such that $\gcd(a, b) = ua + vb$.

Proof. We start with the case $b \geq 0$, proceeding by strong induction.

- **Base Case** ($b = 0$): Note that $\gcd(a, 0) = a = a \cdot 1 + 0 \cdot 0$ as needed. That is, the result holds with $u = 1$ and $v = 0$.
- **Base Case** ($b = 1$): Note that $\gcd(a, 1) = 1 = a \cdot 0 + 1 \cdot 1$ as needed. That is, the result holds with $u = 0$ and $v = 1$.
- **Inductive Step**: Suppose the result holds for all integers b' with $0 \leq b' < b$, where $b > 1$. That is, for all such b' and all integers a there exist integers u and v such that $\gcd(a, b') = au + b'v$. Now consider b. By the division algorithm we have integers q and r such that $a = qb + r$ and $0 \leq r < b$. We have two possibilities to consider.
  - If $r = 0$, then in fact $b|a$, since $a = qb$. So $\gcd(a, b) = b = a \cdot 0 + q \cdot b$. That is, the result holds with $u = 0$ and $v = 1$.
  - If $r > 0$, then by the induction hypothesis there exist integers u' and v' such that $\gcd(b, r) = bu' + rv'$. By the euclidean algorithm, we have

$$\begin{aligned} \gcd(a, b) &= \gcd(b, r) \\ &= bu' + rv' \\ &= bu' + (a - qb)v' \\ &= av' + b(u' - qv'). \end{aligned}$$

That is, the result holds with $u = v'$ and $v = u' - qv'$.

By Strong Induction, for all $b \geq 0$ and all integers a there exist integers u and v such that $\gcd(a, b) = au + bv$.

Now suppose $b < 0$, so that $-b > 0$. By the previous discussion, there exist integers u' and v' such that $\gcd(a, -b) = au' + (-b)v'$. Now

$$\gcd(a, b) = \gcd(a, -b) = au' + (-b)v' = au' + b(-v').$$

That is, the result holds with $u = u'$ and $v = -v'$. □

Similar to the Euclidean Algorithm, this proof of Bezout's Identity provides us with a strategy for actually finding the coefficients u and v recursively.

Def'n 3.8 (Relatively Prime)**.** We say that integers a and b are *relatively prime* if $\gcd(a, b) = 1$.

Theorem 3.9 (Euclid's Lemma)**.** If a and b are relatively prime integers and c an integer such that $a|bc$, then $a|c$.

Proof. By Bezout's Identity, we have $1 = au + bv$ for some integers u and v; so $c = auc + bvc$. Since $a|bc$, we have $bc = at$ for some integer t. Thus

$$c = auc + bvc = auc + atv = a(uc + tv),$$

and so $a|c$ as claimed. □

* * Exercises * *

3.1. Let a and b be integers. Show that if $ac = b$ and $ad = b$ then $c = d$.

3.2. Suppose a and b are integers such that $a|b$ and $b|a$. Show that $b = \pm a$.

3.3. Suppose a and b are integers such that $a|b$. Show that $(-a)|b$, $a|(-b)$, and $(-a)|(-b)$.

3.4. Use the Euclidean Algorithm to compute the following.

(i) $\gcd(12, 5)$

(ii) $\gcd(57, 17)$

(iii) $\gcd(210, 115)$

(iv) $\gcd(1337, 1007)$

4 Factorization

Recall that every integer except 1 and -1 has at least two divisors (four if we count negatives): itself and 1. The integers which have *only* these divisors are somehow simpler than the rest, so we give them a name.

Def'n 4.1 (Prime)**.** We say an integer $p \notin \{1, 0, -1\}$ is *prime* if whenever $p = ab$, either $a = \pm 1$ or $b = \pm 1$.

Equivalently, p is prime if it is not 0, 1, or -1, and the only divisors of p are ± 1 and $\pm p$. You might wonder why we have excluded 0, 1, and -1 from our definition of "prime". After all, 1 has no positive integer divisors other than itself and 1, right? This is true, but it turns out that 1 and -1 are special for another reason: we can "divide" by them in the integers in the sense that $1 \cdot x = a$ and $-1 \cdot x = a$ have solutions in $\mathbb{Z}$ for all integers a. For reasons we'll explain in a bit, this makes it inconvenient to think of 1 and -1 as prime. On the other hand there is a very good argument to be made that we should think of 0 as prime, even though it is divisible by every integer! The point (for now) is that the whole business of prime numbers becomes much simpler if we just exclude 1, 0, and -1 from consideration, and so we do.

Prop'n 4.2. Let p and a be integers.

(i) 2 and 3 are prime.

(ii) If p is prime, then $-p$ is prime.

(iii) If p and q are prime such that $p|q$, then $q = \pm p$.

(iv) If p is prime then $\gcd(a, p)$ is $|p|$ if $p|a$ and 1 otherwise.

The following result is an alternative way to characterize the prime integers which will turn out to be interesting.

Prop'n 4.3. An integer p is prime if and only if whenever $p|ab$, either $p|a$ or $p|b$.

Proof. This is an "if and only if" statement; we proceed by proving the "only if" part and then the "if" part.

($\Rightarrow$) Suppose p is prime and that $p|ab$. Consider $\gcd(a, p)$. Since p is prime, by 4.2(iv) there are two possibilities.

(i) If $\gcd(a, p) = |p|$, then $|p|$ divides a, so that $p|a$.

(ii) If $\gcd(a, p) = 1$, then by Euclid's Lemma, $p|b$.

($\Leftarrow$) Suppose p has the property that if $p|ab$ then either $p|a$ or $p|b$. Suppose further that $p = ab$. In particular $p|ab$ (since $1 \cdot p = ab$) and so, without loss of generality, $p|a$. Say $a = pa'$. Now $p = ab = pa'b$, and thus $1 = a'b$. Now $|b| = 1$, and thus $|p| = |a|$, so that $p = \pm a$ as needed. □

(By the way, this is where the argument that 0 should be prime comes in: it satisfies Proposition 4.3.)

Cor. 4.4. If p is a prime and a_i integers such that $p|a_1a_2 \cdots p_n$, then $p|a_k$ for some k.

The next two results are the reason why primes are interesting, and ultimately the reason why we don't consider 1 and -1 to be primes. The primes are the building blocks of all other integers in a very precise sense: every integer can be written as a product of primes in essentially one way. This

Theorem 4.5 (Fundamental Theorem of Arithmetic: Existence)**.** Every integer other than 0, 1, and -1 can be written as a product of primes. That is, every such n can be expressed as $n = p_1p_2 \cdots p_k$, where the p_i are prime. This is called a *prime factorization* of n.

Proof. Suppose first that $b > 0$. We proceed by strong induction.

(i) **Base Case** ($b = 2$): If $d|2$, then $0 < |d| \leq |2|$. Thus the only possible divisors of 2 are ± 1 and ± 2, and so 2 is prime by definition.

(ii) **Inductive Step**: Suppose that for some n, every integer $2 \leq n' < n$ can be written as a product of primes, and consider n. If n is itself prime, then $n = n$ is its own prime factorization. If n is not prime, then by definition there exist integers a and b such that $n = ab$ and $n \neq \pm a$ and $n \neq \pm b$. Since $n > 0$, we can assume that $a > 0$ and $b > 0$. In fact we have $a > 1$ and $b > 1$, so that $a, b < n$. By the inductive hypothesis, a and b have prime factorizations; say $a = p_1p_2 \cdots p_h$ and $b = q_1q_2 \cdots q_k$. Now

$$n = ab = p_1p_2 \cdots p_hq_1q_2 \cdots q_k$$

has a prime factorization.

Thus by strong induction every integer $n \geq 2$ has a prime factorization. If $n < 0$, then $-n > 0$, so that $-n = p_1p_2 \cdots p_k$ has a prime factorization; then $n = (-p_1)p_2 \cdots p_k$ also has a prime factorization. □

Theorem 4.6 (Fundamental Theorem of Arithmetic: Uniqueness)**.** The prime factorization of an integer is unique in the following sense. If n has two prime factorizations

$$n = p_1p_2 \cdots p_k = q_1q_2 \cdots q_\ell,$$

then $k = \ell$ and, after relabeling the q_i, we have $p_i = \pm q_i$ for each $1 \leq i \leq k$.

Proof. We saw in FTA Part 1 that every such n has at least one prime factorization, which consists of at least one prime factor. To show uniqueness we will proceed by strong induction on the *length* of the shortest prime factorization of n.

(i) **Base Case:** Suppose $n = p$ has a prime factorization of length 1; that is, n itself is prime. Suppose $p = q_1 q_2 \dots q_\ell$ is another prime factorization of n. Since p is prime, we have (rearranging the q_i if necessary) $p|q_1$. Since p and q_1 are prime, we have $q_1 = \pm p$. Now if $\ell > 1$ we have

$$1 = |q_2 \cdots q_\ell|,$$

so that $|q_i| = 1$ for each q_i, a contradiction. So in fact $\ell = 1$ and $q_1 = \pm p$, as claimed.

(ii) **Inductive Step:** Suppose that every integer having a shortest prime factorization of length at most k has a unique prime factorization, and suppose $n = p_1 p_2 \cdots p_{k+1}$ is an integer with a shortest prime factorization of length $k+1$. Suppose further that $n = q_1 q_2 \cdots q_\ell$ is another prime factorization of n. In particular, $p_1 | q_1 q_2 \cdots q_\ell$, so that, rearranging the q_i if necessary, $p|q_1$. Now $q_1 = \pm p$, and we have

$$p_2 \cdots p_{k+1} = q_2 \cdots q_\ell.$$

Note that these are two prime factorizations of an integer having a shortest prime factorization of length at most k. By the Inductive Hypothesis, we have $\ell = k+1$ and, relabeling the q_i if necessary, $q_i = \pm p_i$ for each $2 \le i \le k+1$. So the prime factorization of n is unique. □

We've established that every integer has an essentially unique prime factorization. But **how do we prove that a given integer is prime**? The definition suggests one way to do it, known as *trial division*: an integer n is prime if and only if n is not divisible by any integer t with $1 < t < n$. This works, but is extremely time consuming. A better method is suggested by the following.

Prop'n 4.7. Let $n > 1$ be an integer, and let t be the largest integer such that $t^2 < n$. (Such an integer exists, since the set of all t with $t^2 < n$ is bounded above by n.) (Also, this t is $\lfloor \sqrt{n} \rfloor$, but we don't know what $\sqrt{\cdot}$ means.) Then n is prime if and only if n is not divisible by any prime p with $2 \le p \le t$.

Proof. Certainly if n is prime it is not divisible by any such p. We prove the "only if" part by contraposition. Suppose that n is *not* prime; say $n = ab$, where $n \ne a$ and $n \ne b$. (We can assume positive signs here since $n > 0$.) If a and b are both strictly larger than t, then we have $n > t^2 > ab > n$, a contradiction. Without loss of generality, then, $a \le t$. In particular, all prime factors of a are less than t in absolute value, and so n has a prime factor p such that $2 \le p \le t$. □

For example, consider $n = 5$. The largest t such that $t^2 < 5$ is $t = 2$, and the only prime p such that $2 \leq p \leq 2$ is $p = 2$. Since (by the Division Algorithm) we have $5 = 2 \cdot 2 + 1$, with remainder $1 \neq 0$, 2 does not divide 5. So **5 is prime**.

This result gives us a strategy for finding prime numbers, but it only works if we have a complete list of primes up to $\sqrt{n}$ to begin with. In other words, we can find a prime if we start with a list of primes. This can be made into a reasonably efficient algorithm for finding all the primes up to some bound, which we will explore in the exercises. There are some interesting questions left unanswered, though. Suppose we have an integer n.

- Is n prime?
- What is the smallest prime factor of n?
- What is the prime factorization of n?
- How many prime factors does n have?
- How many *distinct* prime factors does n have?
- How difficult is it to answer these questions?
- How difficult is it to *verify* the answers to these questions?

* * Exercises * *

Def'n 4.8 (Euler Totient)**.** Let n be a positive integer. We define the *totient* of n to be the cardinality of the set

$$\{a \mid 0 \leq a < n, \mathsf{gcd}(a, n) = 1\}.$$

We denote this number by $\mathsf{tot}\,(n)$.

4.1. **Sieve of Eratosthenes.** (@@@)

5 Modular Arithmetic

Def'n 5.1 (Congruence Modulo n)**.** Let n be a positive integer. If a and b are integers such that $n|(b-a)$, then we say a and b are *congruent modulo* n, denoted $a \equiv b \mod n$ or $a \equiv_n b$.

Prop'n 5.2. If n is a fixed positive integer then congruence modulo n is an equivalence relation.

Since $\equiv_n$ is an equivalence, it induces a partition on the set $\mathbb{Z}$ of integers, $\mathbb{Z}/\equiv_n$. We will denote this partition using $\mathbb{Z}/(n)$ and refer to this set as the set of *modular integers*.

Prop'n 5.3. The elements of $\mathbb{Z}/(n)$ are sets of the form $[r]_n$, where $0 \leq r < n$; such r are called *residues* mod n. Moreover, any two such sets are distinct. In particular, $\mathbb{Z}/(n)$ is a finite set with precisely n elements, which are represented by the set of residues $\{0, 1, \ldots, n-1\}$.

Proof. First we show that every class in $\mathbb{Z}/(n)$ has a representative r with $0 \leq r < n$. To this end, let $[a] \in \mathbb{Z}/(n)$. By the Division Algorithm, we have $a = qn + r$, where $0 \leq r < n$, and since $a - r = qn$, we have $a \equiv r \mod n$. Thus $[a] = [r]$ as needed.

Next we show that two such classes are distinct. To this end, suppose we have $[r_1] = [r_2]$, where $0 \leq r_1, r_2 < n$. By definition, we have that n divides $r_2 - r_1$; say $r_2 - r_1 = qn$. In particular, $r_2 = qn + r_1$. Note also that $r_2 = 0 \cdot n + r_2$. By the uniqueness of positive remainders given by the Division Algorithm in 2.3, we have $r_1 = r_2$. □

Prop'n 5.4. Let n be a positive integer. If a_1, a_2, b_1, and b_2 are integers such that $a_1 \equiv a_2 \mod n$ and $b_1 \equiv b_2 \mod n$, then we have the following.

(i) $a_1 + b_1 \equiv a_2 + b_2 \mod n$.

(ii) $a_1 b_1 \equiv a_2 b_2 \mod n$.

Cor. 5.5. Let n be a positive integer. Then the operations $+$ and $\cdot$ on $\mathbb{Z}/(n)$ given by

$$[a] + [b] = [a + b] \quad \text{and} \quad [a] \cdot [b] = [ab]$$

are well-defined.

Prop'n 5.6 (Modular Arithmetic)**.** Let n be a positive integer. Then $\mathbb{Z}/(n)$, with the operations $+$ and $\cdot$ defined as above, satisfies the following properties.

A1. $([a]+[b])+[c] = [a]+([b]+[c])$ for all a, b, and c.

A2. There is a modular integer 0 with the property that $[a]+0 = 0+[a] = [a]$ for all a.

A3. For every residue $[a]$, there is a unique residue $[b]$ with the property that $[a]+[b] = [b]+[a] = 0$. We denote this residue by $-[a]$.

A4. $[a]+[b] = [b]+[a]$ for all a and b.

M. $([a]\cdot[b])\cdot[c] = [a]\cdot([b]\cdot[c])$ for all a, b, and c.

D. $[a]\cdot([b]+[c]) = [a]\cdot[b]+[a]\cdot[c]$ and $([b]+[c])\cdot[a] = [b]\cdot[a]+[c]\cdot[a]$ for all a, b, and c.

C. $[a]\cdot[b] = [b]\cdot[a]$ for all a and b.

U. There is a modular integer 1 with the property that $[a]\cdot 1 = 1\cdot[a] = [a]$ for all a.

Prop'n 5.7 (Modular Inverses)**.** Let n be a positive integer, and a an integer. Then the congruence $ax \equiv 1 \mod n$ has a solution x if and only if $\gcd(a,n) = 1$. In this case, the solution x is unique mod n.

Proof. First suppose $\gcd(a,n) = 1$. To see existence, note that by Bezout's Identity, we have $au+nv = 1$ for some integers u and v. In particular, $n|(au-1)$, so that $au \equiv 1 \mod n$ as needed. Conversely, suppose $ax \equiv 1 \mod n$ has a solution u. By definition we have that n divides $au-1$, so that $1 = au+nv$ for some integer v. Now let $d = \gcd(a,n)$, with $a = da'$ and $n = dn'$. Then $1 = d(a'u+n'v)$, so that $d = 1$ as claimed. To see uniqueness, suppose we have two solutions of this equation, u_1 and u_2. Note that $au_1 \equiv au_2 \mod n$, so that n divides $au_1 - au_2 = a(u_1-u_2)$. Since $\gcd(a,n) = 1$ we have $n|(u_1-u_2)$ by Euclid's Lemma, so that $u_1 \equiv u_2 \mod n$ as claimed. □

Cor. 5.8. Let $p > 1$ be a prime. If $ab \equiv 0 \mod p$, then either $a \equiv 0 \mod p$ or $b \equiv 0 \mod p$.

Cor. 5.9. Let p be a prime. If $a \in [1,p)$, then there is a unique $b \in [1,p)$ such that $ab \equiv 1 \mod p$. Moreover, a and b are distinct unless $a = 1$ or $a = p-1$.

Proof. The existence and uniqueness of b follows from the previous result. Now suppose $a = b$; that is, $a^2 \equiv 1 \mod p$. Then $(a-1)(a+1) \equiv 0 \mod p$. Since p is prime, we must have either $a - 1 \equiv 0 \mod p$ or $a + 1 \equiv 0 \mod p$; in the first case, $a = 1$, and in the second case, $a = p - 1$. □

Cor. 5.10 (Wilson's Theorem)**.** Let $n > 2$ be an integer. Then n is prime if and only if $(n-1)! \equiv -1 \mod n$.

Proof. Suppose $n = p$ is prime, and consider the residues

$$1, 2, 3, \ldots, p-2, p-1.$$

All such residues *except* 1 and $p-1$ come in inverse pairs. So after rearranging, we have

$$(p-1)! = 1 \cdot (p-1) \cdot (t_1 \cdot u_1) \cdot \cdots \cdot (t_k \cdot u_k),$$

where $t_i \cdot u_i \equiv 1 \mod p$. Thus $(p-1)! \equiv p - 1 \equiv -1 \mod p$ as claimed.

Conversely, suppose n is not prime; then we have $1 < a < n$ and $1 < b < n$ such that $n = ab$. But now a and b both appear among the factors of $(n-1)!$ so that $(n-1)! \equiv 0 \mod n$. □

Prop'n 5.11 (Simultaneous Linear Congruences)**.** Let a and b be relatively prime positive integers. Then for any integers u and v, the system of congruences

$$\begin{cases} x \equiv u \mod a \\ x \equiv v \mod b \end{cases}$$

has a unique solution mod n.

Proof. First we show existence. Since $\gcd(a, b) = 1$, by Bezout's Identity there exist integers h and k such that $1 = ah + bk$. Multiplying by $v - u$, we have

$$v - u = ah(v-u) + bk(v-u),$$

and rearranging, we let

$$t = u + ah(v-u) = v - bk(v-u).$$

Clearly $t \equiv u \mod a$ and $t \equiv v \mod b$.

Next we show uniqueness. To this end, suppose t and s are both solutions of this system. In particular, we have $t \equiv u \mod a$ and $t \equiv u \mod b$. Say $q_1 a = u - t = q_2 b$. Now a divides $q_2 b$, and since a and b are relatively prime, by Euclid's Lemma we have $a | q_2$. Thus $u - t = q_2' ab$, so that $t \equiv u \mod ab$ as needed. □

Summary of Chapter 0

— I —

Rings: The Basics

You've been using integer arithmetic since before you were in school. By *arithmetic*, I mean not only the basic arithmetic operations – plus, times, and powers – on the integers, but also the *properties* that these operations satisfy. For instance, it is simple enough to verify that

$$(2+3)+6=11 \qquad \text{and} \qquad 2+(3+6)=11.$$

The fact that these two addition problems simplify to the same result is just a specific example of the more general fact that, if a, b, and c are any three integers at all, then

$$(a+b)+c=a+(b+c).$$

This is called the *associative property* of integer addition and is just one of several such properties you probably use without even thinking; properties with names like *commutativity* and *distributivity*. This combination of **objects** with **operations** and **properties** – **arithmetic** – is quite powerful.

More recently you've also learned to use *modular arithmetic* in $\mathbb{Z}/(n)$. This is a strange view of $\mathbb{Z}$ where we only care about remainders when integers are divided by some fixed *modulus* n. It turns out that the plus and times in $\mathbb{Z}$ have counterparts in $\mathbb{Z}/(n)$ which satisfy many (but not all!) of the same properties. We still have distributivity, for instance, but while there do not exist nonzero integers a and b such that $ab=0$, the same is not true in $\mathbb{Z}/(n)$. For instance the residues 2 and 3 are both nonzero in $\mathbb{Z}/(6)$, but $2\cdot 3\equiv 0 \pmod 6$.

That is, $\mathbb{Z}$ and $\mathbb{Z}/(n)$ are different in some ways, but alike in others. Very often when this occurs in mathematics – when we have two or more useful gadgets with similar behavior – it is useful to distill the common behavior into an abstract definition. This will be our project in Chapter I. The common behavior of $\mathbb{Z}$ and $\mathbb{Z}/(n)$ (and some other examples) will be singled out in the definition of a class of objects we call *rings*.

6 Rings

So far we've studied two kinds of numbers: the integers, $\mathbb{Z}$, that we know and love, and the integers modulo n, $\mathbb{Z}/(n)$, which are a little strange. These kinds of numbers differ in some crucial ways. For example, $\mathbb{Z}$ comes with a useful order relation $\leq$ while $\mathbb{Z}/(n)$ does not, and in $\mathbb{Z}/(n)$ it may be possible to find "nonzero" numbers a and b such that $ab \equiv 0$, which cannot happen in $\mathbb{Z}$. However both $\mathbb{Z}$ and $\mathbb{Z}/(n)$ have an arithmetic – plus and times – which behave very similarly. Addition is associative and commutative, there is a zero element with certain properties, and so on.

In mathematics, when different concrete objects have behavior in common it is frequently useful to "factor out" the common behavior as an abstract definition, like this.

Def'n 6.1 (Ring)**.** A *ring* is a set R equipped with a special element $0_R \in R$ (called "zero") and two binary operations $+$ (pronounced "plus") and $\cdot$ (pronounced "times", and usually left implicit) which together satisfy the following properties.

A1. $(a + b) + c = a + (b + c)$ for all $a, b, c \in R$.

A2. $a + 0_R = 0_R + a = a$ for all $a \in R$.

A3. For every $a \in R$ there is an element $-a \in R$ (called a *negative* of a) such that $a + (-a) = (-a) + a = 0_R$.

A4. $a + b = b + a$ for all $a, b \in R$.

M. $(ab)c = a(bc)$ for all $a, b, c \in R$.

D. $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ for all $a, b, c \in R$.

For example, both $\mathbb{Z}$ and $\mathbb{Z}/(n)$ are rings with their corresponding plus and times. We will refer to this list of six properties as "the ring axioms". It is important to remember that the symbols $+$ and $\cdot$ will, from now on, depend on context: each specific ring has its own arithmetic, which generally has nothing to do with numbers. Many of the basic properties of arithmetic in $\mathbb{Z}$ can be derived from these axioms alone and thus hold in any ring.

Prop'n 6.2. The following hold in any ring R.

(i) The zero element of R is unique in the following sense: if $a, b \in R$ such that $a + b = a$, then $b = 0_R$.

(ii) Negative elements in R are unique in the following sense: if $a, b \in R$ such that $a + b = 0_R$, then $b = -a$.

(iii) $-(-a) = a$ for all $a \in R$.

(iv) $0_R \cdot a = a \cdot 0_R = 0_R$ for all $a \in R$.

(v) $(-a)b = a(-b) = -(ab)$ for all $a, b \in R$.

(vi) $(-a)(-b) = ab$ for all $a, b \in R$.

Proof. (i) Suppose $a + b = a$. Now $-a + (a + b) = -a + a$, and by A1 we have $(-a + a) + b = -a + a$. Using A3 we have $0_R + b = 0_R$, and by A2 we have $b = 0_R$.

(ii) Suppose $a + b = 0_R$. Now $-a + (a + b) = -a + 0_R$, and by A1 we have $(-a + a) + b = -a + 0_R$. By A3 we have $0_R + b = -a + 0_R$, and using A2 (twice) we have $b = -a$.

(iii) By definition, $(-a) + a = 0_R$, so by the uniqueness of negatives we have $a = -(-a)$.

(iv) Let $a \in R$. Now $a \cdot a + 0_R \cdot a = (a + 0_R) \cdot a = a \cdot a$, and so $0_R \cdot a = 0_R$. The other equality is proved similarly.

(v) Let $a, b \in R$. Now $(-a)b + ab = (-a + a)b = 0_R \cdot b = 0_R$, so that $(-a)b = -(ab)$. The other equality is proved similarly.

(vi) Using ((v)) (twice) and ((iii)), we have

$$(-a)(-b) = -(a(-b)) = -(-(ab)) = ab. \qquad \square$$

We can interpret 6.2(i) as saying, "if it acts like zero, it is zero", and likewise 6.2(ii) says "if it acts like $-a$, it is $-a$". These two properties are particularly handy in practice.

Abstract definitions like 6.1 are great – they can make writing proofs easier, for instance, by throwing away unnecessary details. But abstract definitions are only useful if they represent the behavior of more concrete objects which we care about for some reason. Here is a short list of some basic examples.

Examples

1. **Rings of numbers.** The integers $\mathbb{Z}$ and the modular integers $\mathbb{Z}/(n)$ are our prototypical examples of rings, using the usual plus and times. The rational numbers $\mathbb{Q}$ are also a ring under the usual plus and times; we will prove this later. We will define $\mathbb{Q}$ in Section 20. The real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$ are also rings under the usual plus and times. However, even defining these sets of "numbers" is complicated, so we will avoid using $\mathbb{R}$ and $\mathbb{C}$ as examples for as long as possible. For the curious, $\mathbb{R}$ and $\mathbb{C}$ are defined in Section 39.

2. **The trivial ring.** What is the smallest possible ring? Every ring must (by definition) have at least one element, the zero. Suppose this is *all* we have. Now the arithmetic is necessarily pretty boring: $0 + 0 = 0$ and

$0 \cdot 0 = 0$. It is straightforward to check that these operations make the set $\{0\}$ into a ring. This example isn't very interesting, so we call it the *trivial ring*. (Later on we will see that this ring isn't totally useless.)

3. **Rings of functions.** Suppose we have a ring R, and let A be any nonempty set. Then the set $R^A = \{\varphi \mid \varphi : A \to R\}$ of all mappings $A \to R$ is a ring under the "pointwise" operations

$$(\alpha + \beta)(x) = \alpha(x) + \beta(x) \quad \text{and} \quad (\alpha\beta)(x) = \alpha(x)\beta(x).$$

4. **Matrix rings.** Let R be a ring, and consider the set

$$\mathsf{Mat}_2(R) = \left\{ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \mid a_{11}, a_{12}, a_{21}, a_{22} \in R \right\}$$

of all 2×2 matrices with entries in R. The usual matrix addition and multiplication make $\mathsf{Mat}_2(R)$ into a ring. Specifically, we define

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}$$

and

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}.$$

5. **Even integers.** Consider the set $2\mathbb{Z} = \{2k \mid k \in \mathbb{Z}\}$ consisting of only the even integers. It is not too difficult to show that this set is a ring under the usual plus and times.

6. **Rings of sets.** Let X be any nonempty set. The powerset $\mathcal{P}(X)$ is a ring under the operations $A + B = (A \setminus B) \cup (B \setminus A)$ and $A \cdot B = A \cap B$. This is called a *ring of sets*. Similarly, the sets $\mathcal{P}_{\neq}(X)$ of proper subsets and $\mathcal{P}_F(X)$ of finite subsets are rings under these operations.

This is just the beginning of our list of rings; we will see many more. Two of the most important examples of rings are rings of polynomials and rings of matrices of arbitrary (fixed) size. We will look at these in more depth later.

* * Exercises * *

6.1. Let R be a ring. Show that $-0_R = 0_R$. (Hint: Use 6.2(ii).)

6.2. Let R be a ring. Show that $\mathsf{Mat}_2(R)$ is a ring by verifying that each of the properties in Definition 6.1 hold. This will be tedious.

6.3. Let X be a set.

(i) Show that the ring of sets $\mathcal{P}(X)$ is a ring under the operations given in 6 by verifying that each of the properties in Definition 6.1 hold. This will be tedious.

(ii) Show also that $\mathcal{P}_{\neq}(X)$ and $\mathcal{P}_F(X)$ are rings. Note that you will need to verify that the plus and times given in 6 are total; that is, the sum and product of proper or finite sets is again proper or finite.

6.4. Let R be a ring and let $a \in R$. Show that the set $aR = \{ar \mid r \in R\}$ is a ring under the usual plus and times by verifying that each of the properties in Definition 6.1 hold. Note that you must also show that these operations are total; that is, that the sum and product of multiples of a are again multiples of a. This will be tedious.

Def'n 6.3 (Big Sigma)**.** Let R be a ring. We define the *big sigma* operator on finite lists r_i of elements of R inductively as follows:

$$\sum_{i=1}^{0} r_i = 0_R \qquad \text{and} \qquad \sum_{i=1}^{n+1} r_i = \left(\sum_{i=1}^{n} r_i\right) + r_{n+1}.$$

Then if $a, b \in \mathbb{Z}$ such that $a \leq b$ we define

$$\sum_{i=a}^{b} r_i = \sum_{i=1}^{b-a+1} r_{a+i-1}.$$

6.5. Let R be a ring and let $r_i \in R$. Show that for all integers $a \leq b \leq c$ we have

$$\left(\sum_{i=a}^{b} r_i\right) + \left(\sum_{i=b+1}^{c} r_i\right) = \sum_{i=a}^{c} r_i.$$

(Hint: Use induction.)

6.6. Let R be a ring and $r_i, s \in R$. Show that for all integers n we have

$$s \cdot \left(\sum_{i=1}^{n} r_i\right) = \sum_{i=1}^{n} sr_i \qquad \text{and} \qquad \left(\sum_{i=1}^{n} r_i\right) \cdot s = \sum_{i=1}^{n} r_i s.$$

(Hint: Use induction.)

6.7. Let R be a ring, let $r_i \in R$ for $1 \leq i \leq n$, and let σ be a permutation of $[1, n]$. Show that

$$\sum_{i=1}^{n} r_{\sigma(i)} = \sum_{i=1}^{n} r_i.$$

(Hint: Use induction and Exercise 6.5.)

Def'n 6.4 (Multiples of an element)**.** Let R be a ring and $a \in R$ a fixed

element. Given $n \in \mathbb{Z}$, we define

$$na = \sum_{i=1}^{n} a$$

if $n \geq 0$ and $na = -(-n)a$ if $n < 0$. That is,

$$na = \underbrace{a + a + \cdots + a}_{n \text{ times}}$$

for positive n. We will call the ring elements na the *multiples* of a in R.

6.8. For the following rings R and elements $x \in R$, compute $2x$, $3x$, and $4x$.

(i) $R = \mathbb{Z}/(6)$ and $x = [2]$

(ii) $R = \mathsf{Mat}_2(\mathbb{Z})$ and $x = \begin{bmatrix} 1 & 0 \\ 2 & -4 \end{bmatrix}$

(iii) $R = \mathcal{P}(\{1, 2, 3\})$ and $x = \{1, 2\}$

6.9. Show that the following properties hold for all $a, b \in R$ and $m, n \in \mathbb{Z}$.

(i) $0a = 0_R$.

(ii) $n0_R = 0_R$.

(iii) $n(a + b) = na + nb$.

(iv) $n(ab) = (na)b = a(nb)$.

(v) $n(-a) = -(na)$.

(vi) $(m + n)a = ma + na$.

(vii) $(mn)a = m(na)$.

Def'n 6.5 (Powers of an element)**.** Let R be a ring and $a \in R$ a fixed element. We define a mapping $a^* : \mathbb{N} \setminus \{0\} \to R$ inductively as follows: $a^1 = a$ and $a^{n+1} = a^n a$. We will call the ring elements a^n the *powers* of a in R.

6.10. Let $R = \mathbb{Z}/(7)$ and $x = [2]$. Compute x^2, x^3, and x^4.

6.11. Show that the following properties hold for all $a \in R$ and $m, n \in \mathbb{N} \setminus \{0\}$.

(i) $a^{m+n} = a^m a^n$.

(ii) $a^{mn} = (a^m)^n$.

(iii) $(-a)^m = a^m$ if m is even and $-a^m$ if m is odd.

Def'n 6.6 (Idempotent)**.** We say that an element r in a ring R is *idempotent* if $r^2 = r$. For instance, 0_R is idempotent by 6.2(iv).

6.12. Determine which elements (if any) of the following rings are idempotent.

(i) $\mathbb{Z}/(5)$

(ii) $\mathbb{Z}/(12)$

(iii) $\mathbb{Z}/(30)$

6.13. Show that if a and b are idempotent elements such that $ab = ba$, then ab is also idempotent.

Def'n 6.7 (Nilpotent)**.** We say that an element r in a ring R is *nilpotent* if $r^n = 0_R$ for some positive natural number n. For instance, 0_R is nilpotent in any ring since $0_R^2 = 0_R$. If $r \in R$ is nilpotent, the *smallest* positive integer k such that $r^k = 0_R$ is called the *index of nilpotency* of r and is denoted $\mathsf{nilp}\,(r)$.

6.14. Determine which elements of the following rings are nilpotent.

(i) $\mathbb{Z}/(18)$

(ii) $\mathsf{Mat}_2(\mathbb{Z}/(2))$

Def'n 6.8 (Boolean Ring)**.** A ring R is called *boolean* if all of its elements are idempotent; that is, for all $a \in R$ we have $a^2 = a$.

6.15. Show that the ring $\mathbb{Z}/(n)$ is boolean if and only if $n = 2$.

6.16. Show that if R is a boolean ring, then $-a = a$ for all $a \in R$. (Hint: Meditate upon $(a + a)^2$.)

6.17. Let X be a nonempty set.

(i) Show that the ring of sets $\mathcal{P}\,(X)$ is boolean. (cf. Exercise 6.)

(ii) Show that $\mathcal{P}_{\neq}(X)$ and $\mathcal{P}_F\,(X)$ are boolean.

6.18. Let R be a ring and A a set. Show that R^A is boolean if and only if R is boolean.

6.19. Show that $\mathsf{Mat}_2(R)$ is boolean if and only if R is the trivial ring.

Def'n 6.9 (Integer Annihilator)**.** Let R be a ring with $a \in R$ and let $n \in \mathbb{Z}$. We say that n is an *annihilator* of a if $na = 0_R$. We say that n is an *annihilator* of R if n annihilates every element of R. For instance, the integer 0 is an annihilator of every ring.

6.20. Show that n is an annihilator of $\mathbb{Z}/(n)$ for all $n > 1$.

6.21. Let R be a ring. Show that if m and n are annihilators of R, then $m+n$, mn, and $-m$ are also annihilators of R.

Def'n 6.10 (Additive Order)**.** Let R be a ring and let $a \in R$. If a has a positive integer annihilator, then by the well-ordering property of $\mathbb{N}$ it has a *least* positive annihilator. The *additive order* of a, denoted $\mathsf{ord}\,(a)$, is the least positive annihilator of a if it exists and is 0 otherwise.

6.22. (@@@) additive order of elements of $\mathbb{Z}/(n)$

Def'n 6.11 (Characteristic)**.** Let R be a ring. If R has a positive annihilator, then by the well-ordering property of $\mathbb{N}$ it has a *least* positive annihilator. The *characteristic* of R, denoted $\mathsf{char}(R)$, is the least positive annihilator of R if it exists and is 0 otherwise.

6.23. Show that $\mathsf{char}(\mathbb{Z}) = 0$. (Hint: Note that $na = na$, where on the left we have the nth multiple of a as defined in Exercise 6.9 and on the right we have ordinary integer multiplication.)

6.24. Show that $\mathsf{char}(\mathbb{Z}/(n)) = n$. (Hint: Show that no k with $0 < k < n$ can be an annihilator.)

6.25. Show that if n is an annihilator of the ring R then $\mathsf{char}(R)$ divides n.

6.26. Show that $\mathsf{char}(R) = 1$ if and only if R is the zero ring.

6.27. Let R be a ring. Show that $\mathsf{char}(\mathsf{Mat}_2(R)) = \mathsf{char}(R)$.

6.28. Show that every boolean ring has characteristic 2. Give an example of a ring with characteristic 2 which is not boolean. (cf. Exercise 6.16.)

6.29. Let R be a ring and A a nonempty set. Show that $\mathsf{char}(R^A) = \mathsf{char}(R)$.

Def'n 6.12 (Null Ring)**.** We say that a ring R is *null* if $xy = 0_R$ for all $x, y \in R$.

6.30. Show that the trivial ring is null.

6.31. Let $R = \mathbb{Z}/(4)$ and let $a = [2]$.

(i) Show that the ring aR contains 2 elements. (cf. Exercise 6.4.)

(ii) Show that aR is null.

6.32. Show that if R is a null ring and A a nonempty set, then the ring R^A is null.

6.33. Show that if R is a null ring, then the ring $\mathsf{Mat}_2(R)$ is null.

One of the main activities in $\mathbb{Z}$ or $\mathbb{Q}$ is to solve *equations*. An equation over a ring R is an expression involving elements of R, the arithmetic operations, one or more *variables*, and a single equals sign. To *solve* an equation means to find all the elements of R which can be substituted in for the variables to yield a true statement. Solving equations in a general ring may be difficult.

6.34. Find all elements $x \in \mathbb{Z}/(3)$ which satisfy the equation $x^3 + 1 = 0$.

6.35. Find all the matrices $x \in \mathsf{Mat}_2(\mathbb{Z}/(5))$ which satisfy the equation $x^2 = 0$. This will be tedious.

7 Families of Rings

In Section 6 we defined a kind of abstract structure, called a *ring*, which generalizes the arithmetic of the integers. We also gave a list of several examples which will be amended as we move along. To help us make sense of a growing menagerie of rings we now introduce some useful special properties that a ring may or may not possess.

Def'n 7.1. Let R be a ring. We say that R is *commutative* if it satisfies the following property C (otherwise R is *noncommutative*).

C. $ab = ba$ for all $a, b \in R$.

We say that R is *unital* if it satisfies the following property U (otherwise R is *nonunital*).

U. There is an element $1_R \in R$ (called a *one*), different from 0_R, such that $1_R \cdot a = a \cdot 1_R = a$ for all $a \in R$.

We say that R is *finite* if the set R has finite cardinality (otherwise R is *infinite*).

Just as the zero element of a ring is not literally the number zero, a one in a ring is not literally the number one, but rather a distinguished element that *acts like* one.

Any given ring can have any combination of these properties. Many of the rings we are used to, such as $\mathbb{Z}$, are commutative and unital. In fact many rings which are useful in practice are both commutative and unital, and so in this text we will refer to such rings as *CU rings*. However it is important to remember that there are plenty of interesting rings which are not commutative or not unital.

Examples

1. The number rings $\mathbb{Z}$, $\mathbb{Z}/(n)$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all both commutative and unital. Of these, only $\mathbb{Z}/(n)$ is finite, when $n > 0$.

2. If R is unital then $\mathsf{Mat}_2(R)$ is also unital, with

$$1_{\mathsf{Mat}_2(R)} = \begin{bmatrix} 1_R & 0_R \\ 0_R & 1_R \end{bmatrix}.$$

   Is this ring ever commutative? Is it ever finite?

3. If R is unital, then (like every other element of R) the element 1_R has a unique negative, called -1_R. These two elements need not be distinct! For example in $\mathbb{Z}/(2)$ we have $1 = -1$ since $1 + 1 = 2 \equiv 0$.

4. The ring $2\mathbb{Z}$ of even integers is *not* unital. Why?

As we might expect, some of the nice properties of the integer 1 are shared by the one element of any ring.

Prop'n 7.2. Let R be a unital ring.

(i) The one element of R is unique in the following sense: if $u \in R$ such that $u \cdot a = a$ for all $a \in R$, then $u = 1_R$.

(ii) $(-1_R) \cdot a = -a$ for all $a \in R$.

Proof. (i) Suppose u is such an element; then we have $1_R = u \cdot 1_R = u$.

(ii) Let $a \in R$. Then $a + (-1_R)a = 1_R a + (-1_R)a = (1_R + (-1_R))a = 0a = 0$, so that $(-1_R)a = -a$ by 6.2(ii).

□

If a ring is finite and small enough we can visualize its structure using two *cayley tables* - one for plus and one for times. These are simply addition and multiplication tables like the ones elementary school students learn, with the arithmetic done in R rather than in $\mathbb{Z}$. For example, the cayley tables of $\mathbb{Z}/(5)$ are shown in Figure 7.1. The entry in row a and column b is $a + b$ in the addition table and is ab in the multiplication table. (It is important to keep this convention straight if R happens to be noncommutative!) Cayley tables quickly

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Figure 7.1: Cayley tables of $\mathbb{Z}/(5)$.

become unwieldy as the size of a ring increases. However if a ring is small enough its cayley tables give us a nice way to "see" some important concepts. For instance, the defining property of zero is reflected in the row and column of zero in the additive cayley table, and similarly for one in the multiplicative table. As another example, Figure 7.2 gives the cayley tables of $\mathcal{P}(\{a, b, c\})$, with $\varnothing$ denoted by a dash. The fact that this ring is boolean is reflected in the main diagonal of the multiplication table, and that the characteristic is 2 is reflected in the main diagonal of the addition table. Both $\mathbb{Z}(5)$ and $\mathcal{P}(\{a, b, c\})$ are commutative, which we can see because their multiplication tables are symmetric about the main diagonal. Also, note that the multiplication table of $\mathcal{P}(\{a, b, c\})$ has lots of zeros, but that of $\mathbb{Z}/(5)$ only has zeros in the zero row and column.

+	-	a	b	c	ab	ac	bc	abc
-	-	a	b	c	ab	ac	bc	abc
a	a	-	ab	ac	b	c	abc	bc
b	b	ab	-	bc	a	abc	c	ac
c	c	ac	bc	-	abc	a	b	ab
ab	ab	b	a	abc	-	bc	ac	c
ac	ac	c	abc	a	bc	-	ab	b
bc	bc	abc	c	b	ac	ab	-	a
abc	abc	bc	ac	ab	c	b	a	-

·	-	a	b	c	ab	ac	bc	abc
-	-	-	-	-	-	-	-	-
a	-	a	-	-	a	a	-	a
b	-	-	b	-	b	-	b	b
c	-	-	-	c	-	c	c	c
ab	-	a	b	-	ab	a	b	ab
ac	-	a	-	c	a	ac	c	ac
bc	-	-	b	c	b	c	bc	bc
abc	-	a	b	c	ab	ac	bc	abc

Figure 7.2: Cayley tables of $\mathcal{P}(\{a, b, c\})$.

Commutativity, finiteness, and the existence of a one are just three of many properties we will use to understand the different kinds of rings. Figure 7.3 shows these families as a Venn diagram with examples in each region.

* * EXERCISES * *

7.1. Show that $\mathbb{Z}/(n)$ is a finite, commutative, unital ring for all $n \geq 2$.

7.2. Let X be a nonempty set.

(i) Show that $\mathcal{P}(X)$ is commutative and unital, and that $\mathcal{P}(X)$ is finite if and only if X is finite.

(ii) Show that $\mathcal{P}_{\neq}(X)$ is commutative and *not* unital, and that $\mathcal{P}_{\neq}(X)$ is finite if and only if X is finite.

7.3. Draw the cayley tables of the following rings. This will be tedious.

(i) $\mathbb{Z}/(7)$

(ii) $\mathcal{P}_{\neq}(\{a, b\})$

(iii) $\mathbb{Z}/(8)$

(iv) $\mathcal{P}(\{a, b\})$

(v) aR, where $R = \mathbb{Z}/(4)$ and $a = [2]$

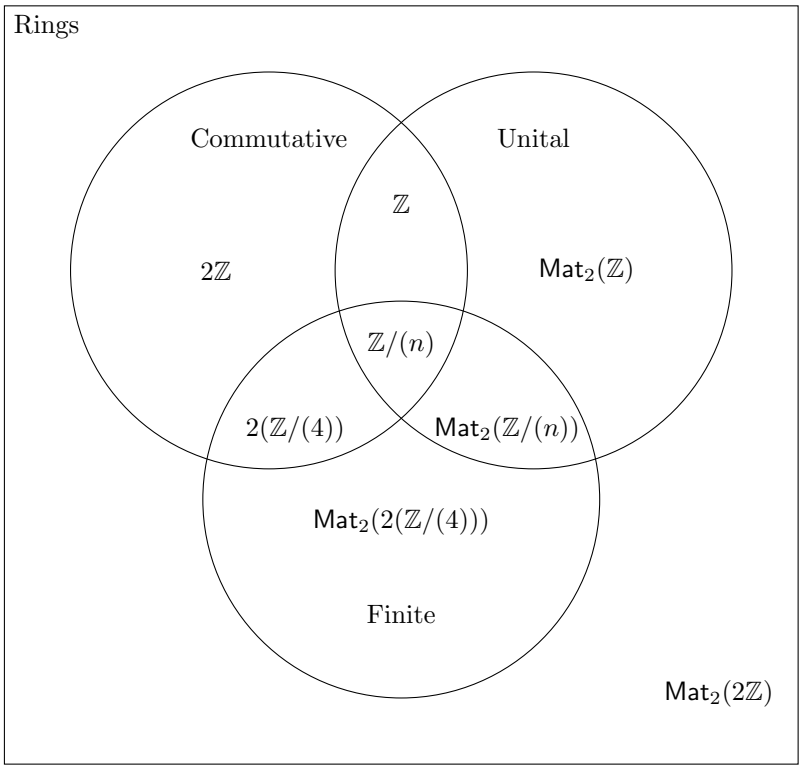


Figure 7.3: A Venn diagram of ring families.

7.4. Let R be a ring.

(i) Show that $\mathsf{Mat}_2(R)$ is unital if and only if R is unital.

(ii) Show that $\mathsf{Mat}_2(R)$ is finite if and only if R is finite.

(iii) Show that if R is not the zero ring, then $\mathsf{Mat}_2(R)$ is not commutative. (Hint: find two specific matrices A and B such that $AB \neq BA$.)

7.5. Let R be a ring and A a nonempty set.

(i) Show that R^A is commutative if and only if R is commutative.

(ii) Show that R^A is unital if and only if R is unital.

7.6. Show that the ring $k\mathbb{Z}$, with $k \geq 2$, is a commutative ring which is not unital. (cf. Exercise 6.4.)

7.7. If R is a unital ring, we can extend the power function a^* (cf. 6.5) to zero exponents by defining $a^0 = 1_R$. Show that if R is a unital ring, then the following properties hold for all $a \in R$ and $m, n \in \mathbb{N}$.

(i) $a^{m+n} = a^m a^n$.

(ii) $a^{mn} = (a^m)^n$.

7.8. Show that if R is commutative, then $(ab)^n = a^n b^n$ for all $a, b \in R$ and all $n \geq 1$. If R is unital, show that this identity also holds for the case $n = 0$.

7.9. **Binomial Theorem.** Let R be a commutative ring and let $a, b \in R$. Show that for all natural numbers $n \geq 1$ we have

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

If R is unital, show that this identity also holds for the case $n = 0$.

7.10. Let R be a unital ring.

(i) Suppose $n1_R = 0_R$ for some $n \geq 2$. Show that $\mathsf{char}(R)$ divides n.

(ii) Show that $\mathsf{char}(R) = \mathsf{ord}\,(1_R)$.

7.11. Show that every finite ring has positive characteristic.

7.12. Let R be a null ring.

(i) Show that R is commutative.

(ii) Show that R is not unital.

7.13. Show that every boolean ring is commutative. (Hint: Meditate upon $(a+b)^2$.)

7.14. Exhibit a boolean ring which is unital, and another which is not unital.

8 Homomorphisms

The ring axioms represent a kind of abstract *structure*. When two objects have the same kind of structure the *structure-preserving mappings* between them are often useful. For example, consider the ring $\mathbb{Z}/(5)$ of integers mod 5. As a set, $\mathbb{Z}/(5)$ contains five equivalence classes of integers. We have a natural mapping $\pi : \mathbb{Z} \to \mathbb{Z}/(5)$, namely the mapping that sends an integer k to its remainder (equivalence class) mod 5. Importantly this mapping respects the arithmetic on $\mathbb{Z}$ in the sense that $\pi(a+b) = \pi(a) + \pi(b)$ and $\pi(ab) = \pi(a)\pi(b)$ – the sum (product) of residues is the residue of the sum (product). This mapping can be used to show, for instance, that the equation $a^4 + b^4 + c^4 = 4d^4$ has no nontrivial solutions in $\mathbb{Z}$ by transporting any such solution to the finite ring $\mathbb{Z}/(5)$. We can formalize the notion of "structure-preserving map" as follows.

Def'n 8.1 (Ring Homomorphism)**.** Let R and S be rings. A mapping $\varphi : R \to S$ is called a *ring homomorphism* if the following are satisfied.

(i) $\varphi(a+b) = \varphi(a) + \varphi(b)$ for all $a, b \in R$.

(ii) $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in R$.

If R and S are both unital rings, we say that φ is *unital* if, in addition to the above, $\varphi(1_R) = 1_S$.

A homomorphism $\varphi : R \to S$ "transports" the structure of R into S; in a concrete sense there is a "shadow" of R inside S. If R and S are both unital rings then the one element is an extra bit of structure which may be preserved.

Prop'n 8.2.

(i) If R is a ring then the *identity map* $\mathsf{id}_R : R \to R$ given by $\mathsf{id}_R(x) = x$ is a ring homomorphism. If R is unital, then id is unital.

(ii) If $\varphi : R \to S$ and $\psi : S \to T$ are ring homomorphisms, then the *composite map* $\psi \circ \varphi : R \to T$ is a homomorphism. If φ and ψ are unital, then $\psi \circ \varphi$ is unital.

Proof. (i) Letting $x, y \in R$, we have $\mathsf{id}_R(x+y) = x + y = \mathsf{id}_R(x) + \mathsf{id}_R(y)$ and likewise $\mathsf{id}_R(xy) = xy = \mathsf{id}_R(x)\mathsf{id}_R(y)$, so that id_R is a ring homomorphism. If R is unital, then $\mathsf{id}_R(1_R) = 1_R$ by definition.

(ii) Let $x, y \in R$. We have

$$\begin{aligned}(\psi \circ \varphi)(x+y) &= \psi(\varphi(x+y)) \\ &= \psi(\varphi(x) + \varphi(y)) \\ &= \psi(\varphi(x)) + \psi(\varphi(y)) \\ &= (\psi \circ \varphi)(x) + (\psi \circ \varphi)(y)\end{aligned}$$

so that $\psi \circ \varphi$ preserves plus. Similarly, $(\psi \circ \varphi)(xy) = (\psi \circ \varphi)(x)(\psi \circ \varphi)(y)$ for all $x, y \in R$, so that $\psi \circ \varphi$ preserves times. Finally, if φ and ψ are unital, we have $(\psi \circ \varphi)(1_R) = \psi(\varphi(1_R)) = \psi(1_S) = 1_T$ and thus $\psi \circ \varphi$ is unital.

□

Although by definition a homomorphism preserves only plus and times, we can show that the zero and negation are also preserved.

Prop'n 8.3. If $\varphi : R \to S$ is a ring homomorphism then we have the following.

(i) $\varphi(0_R) = 0_S$.

(ii) $\varphi(-a) = -\varphi(a)$ for all $r \in R$.

(iii) $\varphi(a - b) = \varphi(a) - \varphi(b)$ for all $a, b \in R$.

Proof. (i) Note that $\varphi(0_R) + \varphi(0_R) = \varphi(0_R + 0_R) = \varphi(0_R)$. By 6.2(i), we have $\varphi(0_R) = 0_S$.

(ii) Note that $\varphi(a) + \varphi(-a) = \varphi(a + (-a)) = \varphi(0_R) = 0_S$, using (i). By 6.2(ii) we have $\varphi(-a) = -\varphi(a)$.

(iii) Follows from (ii).

□

Examples

1. The natural projection $\pi : \mathbb{Z} \to \mathbb{Z}/(n)$ is a surjective unital ring homomorphism for any n.

2. If R is any ring, then there is exactly one ring homomorphism $\varphi : R \to 0$, and exactly one homomorphism $\psi : 0 \to R$. Neither of these is ever unital.

3. Let R be any unital ring. Then $\varphi : R \to \mathsf{Mat}_2(R)$ given by

$$\varphi(r) = \begin{bmatrix} 0 & 0 \\ -r & r \end{bmatrix}$$

is a ring homomorphism. Although R (and hence $\mathsf{Mat}_2(R)$) is unital, φ is *not* a unital homomorphism. (Why?)

When studying algebra, it is often the case that we are concerned with several objects at once along with some of the homomorphisms among them. In this case we can make our lives easier by using a *commutative diagram*. A *diagram* is a collection of rings and homomorphisms among them, represented graphically using labeled arrows. For example, suppose we are concerned with

rings A, B, C, and D, as well as homomorphisms $\varphi : A \to B$, $\psi : C \to D$, $\chi : A \to C$, and $\theta : B \to D$. We might represent these data using the following diagram.

$$\begin{array}{ccc} A & \xrightarrow{\varphi} & B \\ {\scriptstyle\chi}\downarrow & & \downarrow{\scriptstyle\theta} \\ C & \xrightarrow[\psi]{} & D \end{array}$$

The arrows are meant to suggest the flow of elements through a network of conduits. A diagram is said to *commute* if any two composite chains of homomorphisms with the same start and end points are equal. For instance, the above diagram is commutative precisely when $\theta \circ \varphi = \psi \circ \chi$. This is a powerful idea; when a diagram commutes we can prove things about mappings using substitution.

It turns out that lots of important theorems can be stated in the form "there is a unique map which makes such-and-such diagram commute". It also turns out that, once we get used to thinking in terms of homomorphisms, this can be a powerful way to prove theorems. (Diagrams can also look pretty.) We will see several commutative diagrams in this text, but as an initial example, 8.2 can be understood to assert that the two diagrams in Figure 8.1 commute for any φ and ψ. (Why?)

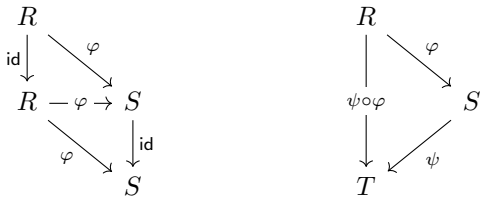


Figure 8.1: The punchline of 8.2.

* * Exercises * *

8.1. Show that the projection map $\pi : \mathbb{Z} \to \mathbb{Z}/(n)$ given by $\pi(k) = [k]$ is a unital ring homomorphism.

8.2. Let $\varphi : R \to S$ be a ring homomorphism.

(i) Show that $\varphi(a^n) = \varphi(a)^n$ for all $a \in R$ and all positive n.

(ii) Show that $\varphi(a^0) = \varphi(a)^0$ for all $a \in R$ if and only if φ is unital.

8.3. Show that if $\varphi : R \to S$ is a homomorphism, then

$$\varphi\left(\sum_{i=a}^{b} r_i\right) = \sum_{i=a}^{b} \varphi(r_i)$$

for all finite sequences r_i.

8.4. Let R and S be rings. Show that the map $\varphi : R \to S$ given by $\varphi(x) = 0_S$ for all $x \in R$ is a ring homomorphism. Is this map ever unital?

If we wish to understand a ring, we must also understand the homomorphisms to and from that ring. The next three exercises characterize some of the homomorphisms involving 0, $\mathbb{Z}$, and $\mathbb{Z}/(n)$.

8.5. **Homomorphisms to and from 0.** Let R be a ring.

(i) Show that there is a *unique* ring homomorphism $R \to 0$.

(ii) Show that there is a *unique* ring homomorphism $0 \to R$.

8.6. **Homomorphisms from $\mathbb{Z}$.** Let R be a ring.

(i) Let $\varphi : \mathbb{Z} \to R$ be a ring homomorphism. Show that φ is uniquely determined by its image at 1 in the following sense: if ψ is another ring homomorphism $\mathbb{Z} \to R$ such that $\psi(1) = \varphi(1)$, then $\psi = \varphi$.

(ii) Let $a \in R$ be a fixed element. Show that there exists a ring homomorphism $\varphi : \mathbb{Z} \to R$ such that $\varphi(1) = a$. (Hint: There's only one way to do it.)

(iii) Show that if R is unital then there is a *unique* unital ring homomorphism $\mathbb{Z} \to R$.

8.7. **Homomorphisms from $\mathbb{Z}/(n)$.** Let $n \geq 2$ be a positive integer and let R be a ring.

(i) Show that every ring homomorphism $\varphi : \mathbb{Z}/(n) \to R$ is uniquely determined by its image of 1 in the following sense: if ψ is another ring homomorphism $\mathbb{Z}/(n) \to R$ such that $\psi(1) = \varphi(1)$, then $\psi = \varphi$.

(ii) Let $a \in R$ be a fixed element. Show that there exists a ring homomorphism $\varphi : \mathbb{Z}/(n) \to R$ such that $\varphi(1) = a$ if and only if a is annihilated by a divisor of n. Take care that your map is well-defined. (Hint: There's only one way to do it.)

(iii) Show that if R is unital and $\mathsf{char}(R)$ divides n, then there is a *unique* unital ring homomorphism $\mathbb{Z}/(n) \to R$.

8.8. (@@@) characterize the homomorphisms $\mathbb{Z}/(m) \to \mathbb{Z}/(n)$.

8.9. Show that if $\varphi : R \to S$ is a unital ring homomorphism, then $\mathsf{char}(S)$ divides $\mathsf{char}(R)$. (@@@ is this true)

8.10. Let R be a commutative ring and $e \in R$ idempotent. Show that the "multiplication map" $\mu_e : R \to R$ given by $\mu_e(x) = ex$ is a ring homomorphism. If R is also unital, is μ_e ever unital?

8.11. Let A be a nonempty set with $a \in A$ a fixed element, and let R be a ring. Show that the *evaluation map* $\varepsilon_a : R^A \to R$ given by $\varepsilon_a(f) = f(a)$ is a ring homomorphism. Show that if R is a unital ring then ε_a is a unital homomorphism.

8.12. Let $\varphi : R \to S$ be a ring homomorphism and let $a \in R$.

(i) Show that if a is idempotent, then $\varphi(a)$ is idempotent, but the converse does not always hold.

(ii) Show that if a is nilpotent, then $\varphi(a)$ is nilpotent, but the converse does not always hold.

8.13. Let X and Y be sets and $f : X \to Y$ an injective map. Denote by Φ_f the induced map $\mathcal{P}(X) \to \mathcal{P}(Y)$ given by $\Phi_f(A) = f[A]$.

(i) Show that Φ_f is a ring homomorphism.

(ii) Show that Φ_f is unital if and only if f is surjective.

Def'n 8.4 (Endomorphism)**.** Let R be a ring. A ring homomorphism $R \to R$ is called an *endomorphism*. We denote by $\mathsf{End}(R)$ the set of all ring endomorphisms of R, and by $\mathsf{End}_1(R)$ the set of all *unital* ring endomorphisms of R. Certainly we have $\mathsf{End}_1(R) \subseteq \mathsf{End}(R)$, and neither set is empty, since id_R is a unital endomorphism.

8.14. **Endomorphism Arithmetic.** Given a ring R, the set $\mathsf{End}(R)$ comes with a sort of arithmetic. We define the sum of two endos pointwise and the product of two endos by composition.

(i) Show that if $\alpha, \beta \in \mathsf{End}(R)$ then $\alpha\beta \in \mathsf{End}(R)$.

(ii) Show that if $\alpha, \beta \in \mathsf{End}(R)$ then $\alpha + \beta$ need not be in $\mathsf{End}(R)$. (Give an example.)

(iii) Show that if $\alpha, \beta \in \mathsf{End}_1(R)$ then $\alpha + \beta$ is not in $\mathsf{End}_1(R)$.

8.15. Show that if R is a null ring then $\mathsf{End}(R)$ is a ring. (It turns out that this ring is extremely important.)

9 Direct Sums

How do we find concrete examples of rings? Most sets don't have obvious choices for the plus and times operations lying around. But if we have some *existing* rings, equipped with plus and times, we might hope to use that existing arithmetic to construct a new one. So a major motivating question throughout this text is what we will call *The Construction Problem*.

The Construction Problem

How can we build new rings out of the "parts" of old ones?

We will see that, while this question is nebulous enough to avoid a complete answer, there are many useful partial answers. The first, and perhaps the simplest, is to construct *direct sums*: two (or more) rings glued together as a cartesian product.

Def'n 9.1 (Direct Sum of Two Rings)**.** Let R and S be rings. Then the *direct sum* of R and S, denoted $R \oplus S$, is the cartesian product

$$R \oplus S = \{(r, s) \mid r \in R, s \in S\}$$

equipped with the "componentwise" operations defined as follows.

$$\begin{aligned} (r_1, s_1) + (r_2, s_2) &= (r_1 + r_2, s_1 + s_2) \\ (r_1, s_1) \cdot (r_2, s_2) &= (r_1 \cdot r_2, s_1 \cdot s_2) \end{aligned}$$

The cartesian product is one of the least sophisticated ways to bolt two sets together, and the least sophisticated way to attempt a ring structure on a cartesian product is to perform arithmetic coordinatewise. Of course we have to verify that these operations make $R \oplus S$ into a ring; the proof of this is left to the exercises.

Prop'n 9.2. Given rings R and S we have the following.

(i) The direct sum $R \oplus S$ with the componentwise operations is a ring.

(ii) The *coordinate projections* $\pi_1 : R \oplus S \to R$ and $\pi_2 : R \oplus S \to S$, given by $\pi_1(r, s) = r$ and $\pi_2(r, s) = s$, are surjective ring homomorphisms.

(iii) The *coordinate injections* $\iota_1 : R \to R \oplus S$ and $\iota_2 : S \to R \oplus S$, given by $\iota_1(r) = (r, 0_S)$ and $\iota_2(s) = (0_R, s)$, are injective ring homomorphisms.

Although (or perhaps because) the direct sum is such a simple way to build new rings out of old ones, it is a very important tool in our quest to understand the structure of rings. We also get some mappings for free: the

coordinate projections π_1 and π_2 and injections ι_1 and ι_2. These mappings are simple, but they play a very important structural role for the direct sum.

Examples

1. The ring $\mathbb{Z} \oplus \mathbb{Z}$ is, as a set,

$$\mathbb{Z} \oplus \mathbb{Z} = \{(a, b) \mid a, b \in \mathbb{Z}\}.$$

   In this ring we have, for example,

$$(1, 2) + (4, -3) = (1 + 4, 2 - 3) = (5, -1)$$

   and

$$(2, -1) \cdot (5, 0) = (2 \cdot 5, -1 \cdot 0) = (10, 0).$$

2. The ring $\mathbb{Z}/(2) \oplus \mathbb{Z}/(2)$ has 4 elements. The cayley table of this ring is shown in Figure 9.1.

+	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	(0,0)	(0,1)	(1,0)	(1,1)
(0,1)	(0,1)	(0,0)	(1,1)	(1,0)
(1,0)	(1,0)	(1,1)	(0,0)	(0,1)
(1,1)	(1,1)	(1,0)	(0,1)	(0,0)

·	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
(0,1)	(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(0,0)	(0,0)	(1,0)	(1,0)
(1,1)	(0,0)	(0,1)	(1,0)	(1,1)

Figure 9.1: Cayley tables of $\mathbb{Z}/(2) \oplus \mathbb{Z}/(2)$.

We can think of $\oplus$ like an operation on *rings*, which takes two rings and builds a new one. It isn't an operation quite like plus on $\mathbb{Z}$, though, because operations can only be defined on *sets*, but there is no set of all rings. But it is still reasonable to hope that as an operation-like thing $\oplus$ behaves nicely. For instance, we might hope that for two rings R and S, we have $R \oplus S = S \oplus R$. Of course this cannot possibly be true; as sets, $R \oplus S$ and $S \oplus R$ will generally be disjoint. In the next section, we will see how to get around this restriction.

There are other ways to put a ring structure on a cartesian product of rings; some of these are explored in the exercises, and another is in Section 13.

* * Exercises * *

9.1. Construct the cayley tables of the following rings.

(i) $\mathbb{Z}/(2) \oplus \mathbb{Z}/(2)$

(ii) $\mathbb{Z}/(2) \oplus \mathbb{Z}/(3)$

(iii) $2R \oplus \mathbb{Z}/(3)$, where $R = \mathbb{Z}/(4)$

9.2. Let R and S be rings.

(i) Show that $R \oplus S$ is commutative if and only if R and S are commutative.

(ii) Show that $R \oplus S$ is boolean if and only if R and S are boolean.

(iii) Show that $R \oplus S$ is unital if and only if R and S are unital, and in this case $1_{R \oplus S} = (1_R, 1_S)$.

(iv) Show that $R \oplus S$ is null if and only if R and S are null.

(v) Show that $R \oplus S$ is finite if and only if R and S are finite.

9.3. Show that if R is unital then the coordinate projection $\pi_1 : R \oplus S \to R$ is a unital homomorphism.

9.4. Prove Proposition 9.2.

9.5. Let R and S be rings and let $(r, s) \in R \oplus S$.

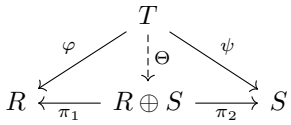
(i) Show that $(r, s)^n = (r^n, s^n)$ for all $n \in \mathbb{N} \setminus \{0\}$. (cf. Exercise 6.11.)

(ii) If R and S are both unital, show that this identity also holds for $n = 0$.

9.6. Let R and S be rings and let $(r, s) \in R \oplus S$. Show that $n(r, s) = (nr, ns)$ for all $n \in \mathbb{Z}$. (cf. Exercise 6.9.)

9.7. Show that $\mathsf{char}(R \oplus S) = \mathsf{lcm}(\mathsf{char}(R), \mathsf{char}(S))$.

9.8. **Universal Property of Binary Direct Sums.** Let R, S, and T be rings, and suppose we have ring homomorphisms $\varphi : T \to R$ and $\psi : T \to S$. Show that there is a unique ring homomorphism $\Theta : T \to R \oplus S$ such that the following diagram commutes.



That is, there is a unique Θ such that $\pi_1 \circ \Theta = \varphi$ and $\pi_2 \circ \Theta = \psi$.

Before we give the proof of this result, a note on its importance. The punchline of Exercise 9.8 asserts the existence of a unique homomorphism Θ through which any pair (φ, ψ) must "factor". This is the first of several results about rings known as *universal properties*. The precise meaning of the term "universal property" is not important for us at the moment; it is enough to think of them as giving us unique homomorphisms for free – in this case, homomorphisms into a direct sum. They usually also involve a kind of unique factorization of homomorphisms as we have here.

9.9. When is the induced map in Exercise 9.8 unital?

9.10. Let R, S, and T be rings with homomorphisms $\varphi : T \to R$ and $\psi : T \to S$. Show that if φ and ψ are both injective, then the induced map $\Theta : T \to R \oplus S$ is injective.

9.11. Let $\varphi : R_1 \to R_2$ and $\psi : S_1 \to S_2$ be ring homomorphisms.

(i) Show that the mapping $\Theta : R_1 \oplus S_1 \to R_2 \oplus S_2$ given by $\Theta(r, s) = (\varphi(r), \psi(s))$ is a homomorphism. We refer to this mapping as $\varphi \oplus \psi$.

(ii) Show that $\varphi \oplus \psi$ is surjective if and only if φ and ψ are both surjective.

(iii) Show that $\varphi \oplus \psi$ is injective if and only if φ and ψ are both injective.

(iv) (@@@) unital

9.12. **A Zero-Square Ring.** The direct sum is one way to make the cartesian product of rings into a ring; here is another. Let S be a commutative ring and let $R = S \times S \times S$. Define addition on R coordinatewise, and define multiplication by

$$(a_1, a_2, a_3) \cdot (b_1, b_2, b_3) = (0, 0, a_1 b_2 - a_2 b_1).$$

Show that R is a ring, and that $x^2 = 0$ for all $x \in R$.

9.13. **Direct Products.** Let Λ be a nonempty set and let $\mathcal{R} = \{R_i \mid i \in \Lambda\}$ be a family of rings indexed by Λ. We denote by

$$\prod_{i \in \Lambda} R_i$$

the set of all mappings $r : \Lambda \to \bigcup_{i \in \Lambda} R_i$ such that $r(i) \in R_i$ for all $i \in \Lambda$.

(i) Show that $\prod_{i \in \Lambda} R_i$ is a ring with respect to pointwise plus and times.

(ii) Show that $\prod_{i \in \Lambda} R_i$ is commutative if and only if R_i is commutative for each $i \in \Lambda$.

(iii) Show that $\prod_{i \in \Lambda} R_i$ is unital if and only if R_i is unital for each $i \in \Lambda$.

9.14. (@@@) move to subrings **Direct Sums.** Let Λ be a nonempty set and let $\mathcal{R} = \{R_i \mid i \in \Lambda\}$ be a family of rings indexed by Λ. We say that an element $\alpha \in \prod_{i \in \Lambda} R_i$ has *finite support* if $\alpha(i) = 0$ for all but finitely many $i \in \Lambda$. We denote by

$$\bigoplus_{i \in \Lambda} R_i$$

the set of all $\alpha \in \prod_{i \in \Lambda} R_i$ of finite support.

(i) Show that $\bigoplus_{i \in \Lambda} R_i$ is a subring of $\prod_{i \in \Lambda} R_i$.

(ii) Show that $\bigoplus_{i \in \Lambda} R_i$ is commutative if and only if R_i is commutative for each $i \in \Lambda$.

(iii) Show that $\bigoplus_{i \in \Lambda} R_i$ is unital if and only if R_i is unital for each $i \in \Lambda$ and Λ is finite.

Def'n 9.3 (Null Extensions of Rings.)**.** Let R and S be rings and let $\theta : R \to \mathsf{End}\,(S)$ be a mapping. Note that $\mathsf{End}\,(R)$ is not (generally) a ring under pointwise addition and composition, so θ cannot possibly be a ring homomorphism. Instead, we say that θ is a *prehomomorphism* if $\theta(a+b) = \theta(a) + \theta(b)$ and $\theta(ab) = \theta(a)\theta(b)$ for all $a, b \in R$. In this case, note that $\mathsf{im}(\theta)$ is a ring.

If θ is a prehomomorphism such that $\mathsf{im}(theta)$ is commutative, we can define an alternative arithmetic on the cartesian product $R \times S$ as follows. Addition is performed coordinatewise, while

$$(r_1, s_1) \cdot (r_2, s_2) = (r_1 r_2, \theta(r_1)(s_2) + \theta(r_2)(s_1)).$$

We call this the *null extension* of S by R via θ, denoted $R \rtimes_\theta S$.

9.15. Show that $R \rtimes_\theta S$ is a ring.

9.16. Let S be a ring.

(i) Show that the map $\theta : 0 \to \mathsf{End}\,(S)$ given by $\theta(0)(s) = 0_S$ is a prehomomorphism and that $\mathsf{im}(\theta)$ is commutative.

(ii) Show that $0 \rtimes_\theta S$ is a null ring which has the same additive structure as S. We call this the *nullification* of S and denote it S^0.

9.17. Let R be a ring and N a null ring, and suppose we have a prehomomorphism θ with $\mathsf{im}(\theta)$ commutative. (@@@ what was I going to say here?)

9.18. Let R be a commutative ring. We define operations on the set $\mathbb{H}(R) = R \times R \times R \times R$ by (@@@)

(i) These operations make $\mathbb{H}(R)$ a noncommutative ring.

(ii) The map $\iota : R \to \mathbb{H}(R)$ given by $\iota(r) = (r, 0, 0, 0)$ is an injective ring homomorphism.

(iii) If R is a field, then $\mathbb{H}(R)$ is a skew field.

10 Isomorphisms

Just for fun, let's consider the cayley tables of the ring $\mathbb{Z}/(6)$ as shown in Figure 10.1. Let's also consider the cayley tables of the ring $\mathbb{Z}/(2) \oplus \mathbb{Z}/(3)$,

$+$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\cdot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Figure 10.1: Cayley tables of $\mathbb{Z}/(6)$.

with the row and column headings written in a particular order (which will be explained in a moment) as shown in Figure 10.2. Drawing these tables is a little tedious, but straightforward. It may be tricky to spot at first, but these

$+$	(0,0)	(1,1)	(0,2)	(1,0)	(0,1)	(1,2)
(0,0)	(0,0)	(1,1)	(0,2)	(1,0)	(0,1)	(1,2)
(1,1)	(1,1)	(0,2)	(1,0)	(0,1)	(1,2)	(0,0)
(0,2)	(0,2)	(1,0)	(0,1)	(1,2)	(0,0)	(1,1)
(1,0)	(1,0)	(0,1)	(1,2)	(0,0)	(1,1)	(0,2)
(0,1)	(0,1)	(1,2)	(0,0)	(1,1)	(0,2)	(1,0)
(1,2)	(1,2)	(0,0)	(1,1)	(0,2)	(1,0)	(0,1)

$\cdot$	(0,0)	(1,1)	(0,2)	(1,0)	(0,1)	(1,2)
(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
(1,1)	(0,0)	(1,1)	(0,2)	(1,0)	(0,1)	(1,2)
(0,2)	(0,0)	(0,2)	(0,1)	(0,0)	(0,2)	(0,1)
(1,0)	(0,0)	(1,0)	(0,0)	(1,0)	(0,0)	(1,0)
(0,1)	(0,0)	(0,1)	(0,2)	(0,0)	(0,1)	(0,2)
(1,2)	(0,0)	(1,2)	(0,1)	(1,0)	(0,2)	(1,1)

Figure 10.2: Cayley tables of $\mathbb{Z}/(2) \oplus \mathbb{Z}/(3)$.

two pairs of cayley tables are *essentially the same*. Why? Note that the only real difference between the addition tables for $\mathbb{Z}/(6)$ and $\mathbb{Z}/(2) \oplus \mathbb{Z}/(3)$ is the **labels of the elements**. Specifically, we can relabel them as follows.

$$0 \leftrightarrow (0,0) \quad 1 \leftrightarrow (1,1) \quad 2 \leftrightarrow (0,2)$$
$$3 \leftrightarrow (1,0) \quad 4 \leftrightarrow (0,1) \quad 5 \leftrightarrow (1,2)$$

The same is true of the multiplication tables. It may help to draw these cayley tables in color, so that corresponding elements are the same color. Thinking

of this correspondence between $\mathbb{Z}/(6)$ and $\mathbb{Z}/(2) \oplus \mathbb{Z}/(3)$ as a mapping, the sameness of the cayley tables gives us a little more: the correspondence is a *homomorphism*.

Note that we can arrange the row and column labels of Figure 10.1 and Figure 10.2 however we like, but using this particular ordering makes it easier to see the correspondence. Changing the order does not destroy the correspondence, it only makes it harder to see.

This relabeling raises an interesting question: in what sense are $\mathbb{Z}/(6)$ and $\mathbb{Z}/(2) \oplus \mathbb{Z}/(3)$ different rings? Of course they are different as *sets*. But from a ring-theoretic perspective we only really care about *structure* – that is, the arithmetic, as embodied in the cayley tables. These two rings have exactly the same structure, and we can translate between them via the relabeling without losing any information. So in a sense $\mathbb{Z}/(6)$ and $\mathbb{Z}/(2) \oplus \mathbb{Z}/(3)$ are really two different manifestations of the same abstract ring structure, obtained by giving the elements different names. When two rings are related in this way, we say they are *isomorphic* to one another.

Def'n 10.1. Let R and S be rings and $\varphi : R \to S$ a ring homomorphism. If φ is also bijective as a mapping, we say φ is an *isomorphism*. In this case we say that R is *isomorphic to* S, denoted $R \cong S$.

Remember: to show that one ring is isomorphic to another, we must find a bijective homomorphism from one to the other. Conversely, if we know in advance that $R \cong S$, then *there exists* an isomorphism $R \to S$; in general there may be many such mappings. Isomorphisms formalize and generalize the "structure preserving relabeling" we saw between $\mathbb{Z}/(6)$ and $\mathbb{Z}/(2) \oplus \mathbb{Z}/(3)$. In a very concrete sense, if $R \cong S$, then R and S are really "the same" ring. *Isomorphism is structural equality.* It is not too surprising, then, that "is isomorphic to" behaves very much like equality.

Prop'n 10.2. For all rings R, S, and T, the following hold.

(i) $R \cong R$.

(ii) If $R \cong S$ then $S \cong R$.

(iii) If $R \cong S$ and $S \cong T$ then $R \cong T$.

A note before we prove this result: 10.2 appears to be saying that $\cong$ is an equivalence relation. We have to be careful here, though, because equivalence relations only exist *on sets*. There is no set of all rings, for the same reason that there is no set of all sets. And so we cannot say that isomorphism "is" an equivalence relation, even though it "acts like" and equivalence relation.

Proof. (i) The identity mapping $\mathsf{id} : R \to R$ is a bijective homomorphism by 8.2(i).

(ii) Suppose we have an isomorphism $\varphi : R \to S$. Since φ is a bijection, it has an inverse $\varphi^{-1} : S \to R$, with the property that $\varphi \circ \varphi^{-1} = \mathsf{id}_S$ and $\varphi \circ \varphi^{-1} = \mathsf{id}_R$ and which is also a bijection. It thus suffices to show that φ^{-1} is also a ring homomorphism. To this end, let $x, y \in S$. We then have the following.

$$\begin{aligned}
\varphi^{-1}(x+y) &= \varphi^{-1}\left(\mathsf{id}_S(x) + \mathsf{id}_S(y)\right) \\
&= \varphi^{-1}\left((\varphi \circ \varphi^{-1})(x) + (\varphi \circ \varphi^{-1})(y)\right) \\
&= \varphi^{-1}\left(\varphi(\varphi^{-1}(x)) + \varphi(\varphi^{-1}(y))\right) \\
&= \varphi^{-1}\left(\varphi\left(\varphi^{-1}(x) + \varphi^{-1}(y)\right)\right) \\
&= (\varphi^{-1} \circ \varphi)\left(\varphi^{-1}(x) + \varphi^{-1}(y)\right) \\
&= \mathsf{id}_R\left(\varphi^{-1}(x) + \varphi^{-1}(y)\right) \\
&= \varphi^{-1}(x) + \varphi^{-1}(y).
\end{aligned}$$

Thus φ^{-1} preserves plus. A similar argument shows that φ^{-1} preserves times, and so φ^{-1} is a ring homomorphism.

(iii) If $R \cong S$ and $S \cong T$, then we have bijective homomorphisms $\varphi : R \to S$ and $\psi : S \to T$. Now $\psi \circ \varphi : R \to T$ is a homomorphism by 8.2(ii), and since the composite of bijections is bijective, $\psi \circ \varphi$ is bijective. Thus $R \cong T$.

□

Given two mathematical objects, one of the most basic questions we can ask about them is whether or not they are the same. To emphasize the importance of this question, we'll give it a fancy name.

The Distinction Problem

Given rings R and S, can we detect whether or not $R \cong S$?

The Distinction Problem is one of the big questions of algebra. Big enough, in fact, that we cannot hope for a complete answer; the best we can hope for is a sequence of partial answers of increasing power and sophistication. To this end it is useful to have on hand several properties of rings which are *preserved* by isomorphisms, so that two rings which differ must be nonisomorphic. For instance, the property "contains the number 3 as an element" *is not* preserved by isomorphisms: the elements of isomorphic rings may have nothing to do with each other. However, the property "consists of three elements" *is* preserved by isomorphisms. Here are some others.

Prop'n 10.3. Let R and S be rings. If $R \cong S$, then the following hold.

(i) R and S have the same cardinality.

(ii) R is commutative if and only if S is commutative.

(iii) R is unital if and only if S is unital.

(iv) $\mathsf{char}(R) = \mathsf{char}(S)$.

(v) R is boolean if and only if S is boolean.

Just as important as determining when two rings are different is determining when two rings are the same, or determining what all the different rings are (up to isomorphism).

The Classification Problem

Given a property P, what are the rings (up to isomorphism) with property P?

For example, property P may be "is commutative", "is finite", "is boolean", or something more complicated.

* * EXERCISES * *

10.1. Prove 10.3.

10.2. Are the rings $\mathbb{Z}/(m)$ and $\mathbb{Z}/(n)$ ever isomorphic? Why or why not?

10.3. Show that if R is a finite unital ring of prime order p then $R \cong \mathbb{Z}/(p)$. (cf. Exercise 8.7.)

10.4. Show that the following properties hold for all rings R, S, T, and U.

(i) $R \oplus 0 \cong R$.

(ii) $R \oplus S \cong S \oplus R$.

(iii) $(R \oplus S) \oplus T \cong R \oplus (S \oplus T)$.

(iv) If $R \cong T$ and $S \cong U$, then $R \oplus S \cong T \oplus U$.

10.5. Let R be a ring. Show that $\mathsf{Diag}_2(R) \cong R \oplus R$. (cf. 11.8.)

10.6. Let $2 = \{0, 1\}$ denote a set with two elements, and let R be a ring. Show that $R^2 \cong R \oplus R$, where R^2 is the ring of functions $2 \to R$.

10.7. Let R be a ring. Is the ring of sets $R^{\varnothing}$ isomorphic to any other named rings?

10.8. Show that $\mathbb{Z}$ and $\mathsf{Mat}_2(\mathbb{Z})$ are not isomorphic.

10.9. Classify the rings of order 2 up to isomorphism.

10.10. **Rings of prime order.** Let $p \in \mathbb{N}$ be a prime and let R be a finite ring of order p.

(i) Let $\alpha \in R$ be nonzero. Show that every element of R is of the form $k\alpha$ for some $0 \leq k < p$.

(ii) Show that R is commutative.

(iii) Show that if R contains zero divisors, then R is null.

11 Subrings

So far we've seen several examples of rings which are subsets of other rings, like $2\mathbb{Z}$ in $\mathbb{Z}$ and $\mathcal{P}_{\neq}(X)$ in $\mathcal{P}(X)$. This suggests an alternative answer to the Construction Problem: take a *subset* of an existing ring R, and restrict the arithmetic to that subset. There is a potential obstacle to making this work, though; given a subset $S \subseteq R$ and two elements $x, y \in S$, *a priori* we expect their sum $x + y$ and product xy to be in R but not necessarily in S – see Figure 11.1. This is a problem! The plus and times on a ring must be *closed*.

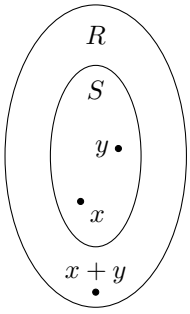


Figure 11.1: Arithmetic in a subset

To avoid this, we single out exactly those subsets of R for which this does not happen. That is, the subsets which are closed under the arithmetic on R.

Def'n 11.1 (Subring)**.** Let R be a ring and $S \subseteq R$ a subset. We say that S is a *subring* of R if S is closed under the operations in R as follows.

(i) $0_R \in S$,

(ii) If $x, y \in S$ then $x + y \in S$,

(iii) If $x \in S$ then $-x \in S$, and

(iv) If $x, y \in S$ then $xy \in S$.

If R is unital, we say that a subring S is *unital* if in addition $1_R \in S$.

Prop'n 11.2. If R is a ring and $S \subseteq R$ a subring, then S is itself a ring under the restricted operations on R. If S is a unital subring of R then S is a unital ring.

Using the definition, showing that a given sub*set* of a ring is a sub*ring* requires verifying four properties (or five if we want a unital subring). We have a slightly more efficient way to achieve this called the Subring Criterion.

Prop'n 11.3 (Subring Criterion)**.** Let $S \subseteq R$ be a subset. Then S is a subring of R if and only if S is not empty and is closed under subtraction and multiplication. That is, S is a subring of R if and only if the following hold.

(i) $S \neq \varnothing$.

(ii) If $x, y \in S$ then $x - y \in S$.

(iii) If $x, y \in S$ then $xy \in S$.

Proof. First suppose $S \subseteq R$ is a subring. Then $S \neq \varnothing$, since $0_R \in S$. Now if $x, y \in S$, we have $xy \in S$, and $-y \in S$, and $x + (-y) \in S$. So S satisfies the Subring Criterion. Conversely, suppose S satisfies the Subring Criterion. Since S is not empty, it contains an element s, and since S is closed under subtraction we have $0_R = s - s \in S$. Now if x is any element of S we have $-x = 0_R - x \in S$, so that S is closed under negation. If $x, y \in S$, then $x + y = x - (-y) \in S$ and $xy \in S$. Thus S is a subring of R. □

The Subring Criterion is a slick way to show that a given subset of a ring is a subring. We have a yet more efficient way to characterize *unital* subrings.

Prop'n 11.4 (Unital Subring Criterion)**.** Let R be a unital ring. Then $S \subseteq R$ is a unital subring if and only if $1_R \in S$ and for all $x, y, z \in S$, $x - yz \in S$.

Proof. Certainly if S is a unital subring of R then it satisfies the Unital Subring Criterion. Conversely, suppose S satisfies the Unital Subring Criterion. Now S is not empty since $1_R \in S$. If $x, y \in S$, then $x - y = x - 1_R \cdot y \in S$, so that S is closed under subtraction. Now $0_R = 1 - R - 1_R 1_R \in S$ and $-x = 0_R - 1_R \cdot x \in S$ whenever $x \in S$, so if $x, y \in S$ we have $xy = 0_R - (-x)y \in S$. By the Subring Criterion S is a subring of R, and since $1_R \in S$, it is a *unital* subring of R. □

Examples

1. **The zero subring.** Let R be any ring. The subset $0 = \{0_R\} \subseteq R$ is a subring, called the *zero subring*. (Show it!)

2. Let R be a ring and k a positive integer, and define $kR = \{kr \mid r \in R\}$. Then $kR \subseteq R$ is a subring, but is *not* a unital subring. (@@@)

3. Let R be any ring and $a \in R$. Then $aR = \{ar \mid r \in R\}$ is a subring of R. (This is the content of Exercise 6.4.) Similarly, $Ra = \{ra \mid r \in R\}$ is a subring of R. These subrings do not have to be the same!

4. If $S_1, S_2 \subseteq R$ are (unital) subrings of R, then $S_1 \cap S_2 \subseteq R$ is also a (unital) subring of R.

There are two special subrings associated to every ring homomorphism: the *image* and the *kernel*.

Prop'n 11.5 (Image and Kernel)**.** Let $\varphi : R \to S$ be a ring homomorphism. Then we have the following.

(i) The set
$$\ker(\varphi) = \{r \in R \mid \varphi(r) = 0\},$$
called the *kernel* of φ, is a subring of R.

(ii) The set
$$\mathsf{im}(\varphi) = \{s \in S \mid s = \varphi(r) \text{ for some } r \in R\},$$
called the *image* of φ, is a subring of S.

Proof. (@@@) □

The image of a homomorphism is everything in the codomain that "gets hit", while the kernel is everything in the domain that is "sent to zero". We can visualize these subsets as in Figure 11.2. The kernels of ring homomorphisms

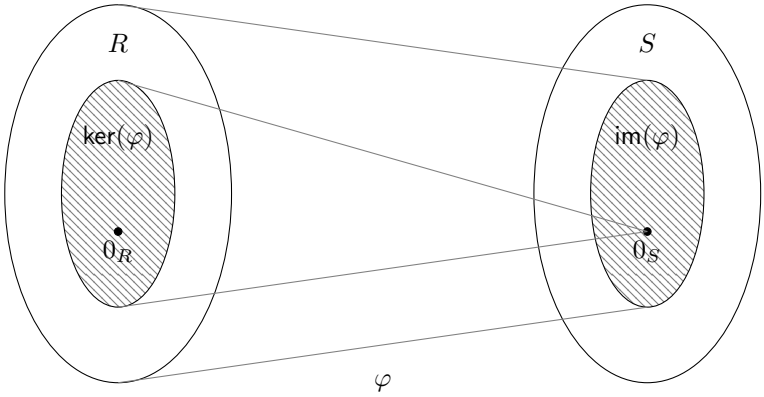


Figure 11.2: The image and kernel of a homomorphism.

turn out to be very important, as we will see later in Chapter IV. As a preview, we can think of homomorphisms $\varphi : R \to S$ as projecting a "shadow" of R into S such that parts of R get collapsed down in the shadow. The kernel is the part of R which is collapsed to zero. As a consequence, the kernel measures how badly a homomorphism fails to be injective as in the following result.

Prop'n 11.6. A ring homomorphism φ is injective if and only if $\ker(\varphi) = 0$.

Proof. Suppose $\varphi : R \to S$ is an injective homomorphism. If $x \in \ker(\varphi)$, then (by definition) we have $\varphi(x) = 0_S = \varphi(0_R)$. Since φ is injective, we have $x = 0$. Thus $\ker(\varphi) = 0$. Conversely, suppose $\ker(\varphi) = 0$. Now suppose we have $x, y \in R$ such that $\varphi(x) = \varphi(y)$. Then $\varphi(x-y) = \varphi(x) - \varphi(y) = 0$, so that $x - y \in \ker(\varphi) = 0$. So $x - y = 0$, and thus $x = y$. Hence φ is injective. □

We have yet another way to characterize subrings: they are precisely the images of injective ring homomorphisms.

Prop'n 11.7. Let R be a ring and $S \subseteq R$ a subset. The following are equivalent.

(i) S is a subring of R.

(ii) There is an injective ring homomorphism $\varphi : T \to R$ such that $\mathsf{im}(\varphi) = S$.

Note that if $\varphi : R \to S$ is an injective ring homomorphism, then we have $R \cong \mathsf{im}(\varphi)$. This allows us to think of R as a "subring" of S, even though it isn't really, by identifying $r \in R$ with $\varphi(r) \in S$. At first this may seem like a trivial observation; we can think of R as sitting inside S, okay. But suppose R is a particularly nasty or inconvenient ring to compute in, while S is very nice – say, a ring of matrices or numbers. Then φ gives us a way to think of R in much more convenient terms. An injective mapping $\varphi : R \to S$ is sometimes called a *representation* of R, as it literally allows us to "represent" elements of R by elements of S.

The Representation Problem

What are all the representations $\varphi : R \to S$
of a ring R by rings S having property P?

Property P here might be "is a ring of sets" or "is a direct sum of rings of matrices"; typically, any property of rings which makes them particularly nice to compute in.

* * Exercises * *

11.1. Let $R = \mathbb{Z}$ and $S \subseteq R$ the set of all prime integers. Show that S is *not* a subring of R.

11.2. Let $a, b > 1$ be relatively prime integers. Show that $\mathbb{Z}/(a) \oplus \mathbb{Z}/(b) \cong \mathbb{Z}/(ab)$. (Hint: Use Exercise 8.7 and 11.6)

11.3. Show that every subring of $\mathbb{Z}$ is of the form $k\mathbb{Z}$ for some k. (@@@) is this true?

11.4. Let R be a ring, and let $e \in R$ be idempotent (that is, $e^2 = e$).

(i) Show that $eRe = \{ere \mid r \in R\}$ is a subring of R.

(ii) Show that as a ring, $S = eRe$ is unital with $1_S = e$.

In particular, if R is a unital ring and $e \neq 1_R$, then S is not a *unital subring*, even though it is a *subring which is unital*, since in a unital subring S we have $1_S = 1_R$.

11.5. Let R be a ring and let $\mathcal{S}$ be an arbitrary family of subrings of R. Show that $\bigcap_{S \in \mathcal{S}} S$ is a subring of R.

Def'n 11.8 (Triangular Matrix, Diagonal Matrix)**.** Let R be a ring. The set of 2×2 *upper triangular* matrices over R is defined to be

$$\mathsf{UT}_2(R) = \left\{ \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} \mid a_{11}, a_{12}, a_{22} \in R \right\}.$$

Similarly, the set of 2×2 *lower triangular* matrices is defined to be

$$\mathsf{LT}_2(R) = \left\{ \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \mid a_{11}, a_{12}, a_{22} \in R \right\},$$

and the set of 2×2 *diagonal* matrices is defined to be $\mathsf{Diag}_2(R) = \mathsf{UT}_2(R) \cap \mathsf{LT}_2(R)$.

11.6. Let R be a ring. Show that $\mathsf{UT}_2(R)$ and $\mathsf{LT}_2(R)$ are subrings of $\mathsf{Mat}_2(R)$. Show also that if R is unital, then $\mathsf{UT}_2(R)$ and $\mathsf{LT}_2(R)$ are unital subrings.

11.7. Find the cayley tables of the ring $\mathsf{UT}_2(\mathbb{Z}/(2))$.

Def'n 11.9 (Center)**.** Let R be a ring. We define a subset $Z(R) \subseteq R$ called the *center* as follows.

$$Z(R) = \{a \in R \mid ax = xa \text{ for all } x \in R\}$$

That is, the center is the set of all ring elements which commute with every other element of R. For example, $0_R \in Z(R)$, since if $x \in R$ we have $0 \cdot x = 0 = x \cdot 0$.

11.8. Let R be a ring. Show that the center $Z(R)$ is a subring of R. Show also that if R is unital, then $Z(R)$ is a unital subring of R.

11.9. Let R be a unital ring. Show that

$$Z(\mathsf{Mat}_2(R)) = \left\{ \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \mid a \in Z(R) \right\}.$$

(Hint: consider products of

$$\begin{bmatrix} r & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

by an arbitrary element in the center.)

11.10. Let R and S be rings. Show that $Z(R \oplus S) = Z(R) \oplus Z(S)$.

11.11. Let R be a ring and A a nonempty set. Show that $Z\left(R^A\right) = Z(R)^A$.

> **Def'n 11.10** (Nilradical)**.** If R is a ring, then the set $N(R) \subseteq R$ consisting of all nilpotent elements is called the *nilradical* of R.

11.12. Show that the nilradical of a commutative ring is a subring.

11.13. Exhibit two matrices $A, B \in \mathsf{Mat}_2(\mathbb{Z}/(2))$ such that A and B are both nilpotent but $A+B$ is not nilpotent. (Thus illustrating that commutativity is necessary for the nilradical to be a subring.)

11.14. Compute the nilradical of the following rings.

(i) $\mathbb{Z}/(5)$

(ii) $\mathbb{Z}/(8)$

(iii) $\mathbb{Z}/(12)$

(iv) $\mathbb{Z}/(30)$

11.15. **Internal Direct Sums.** Let R be a ring, and suppose $A, B \subseteq R$ are subrings such that $A + B = R$ and $A \cap B = 0$.

(i) Show that every element of $A+B$ can be written as $a+b$ for some *unique* $a \in A$ and $b \in B$.

(ii) Show that $R \cong A \oplus B$. (In this case we say that R is an *internal* direct sum of A and B.)

11.16. If R is commutative and $\varphi : R \to S$ a homomorphism, then the subring $\mathsf{im}(\varphi) \subseteq S$ is commutative.

11.17. Let R be a ring. Definition 11.1 gives four conditions which a sub*set* $S \subseteq R$ must satisfy in order to be a sub*ring*. In this exercise we will show that, in general, none of these conditions is redundant. A given subset either does or does not satisfy each condition, giving 16 possibilities. For instance, a given subset may satisfy 11.1(i) and 11.1(ii), but not 11.1(iii) or 11.1(iv).

(i) For each of the 12 possible combinations of properties in Definition 11.1 in which either 11.1(ii) or 11.1(iii) is not satisfied, find a subset $S \subseteq \mathbb{Z}$ which satisfies exactly those properties.

(ii) Show that $\varnothing \subseteq R$ satisfies 11.1(ii), 11.1(iii), and 11.1(iv), but not 11.1(i).

(iii) Show that $\{k\sqrt{2} \mid k \in \mathbb{Z}\} \subseteq \mathbb{Z}[\sqrt{2}]$ satisfies 11.1(i), 11.1(ii), and 11.1(iii), but *not* 11.1(iv).

(iv) Show that no subset $S \subseteq R$ can satisfy 11.1(ii) and 11.1(iii) but not 11.1(i) and 11.1(iv).

Def'n 11.11 (Exact Sequence)**.** A *sequence* is a list of ring homomorphisms

$$R_1 \xrightarrow{\varphi_1} R_2 \xrightarrow{\varphi_2} \cdots \xrightarrow{\varphi_{n-1}} R_n$$

We say that a sequence is *exact* if $\mathsf{im}(\varphi_i) = \mathsf{ker}(\varphi_{i+1})$ for each i. As a special case, if the sequence

$$0 \longrightarrow A \xrightarrow{\varphi} B \xrightarrow{\psi} C \longrightarrow 0$$

is exact, we say that the pair (φ, ψ) is a *short exact sequence*.

11.18. Let $\varphi : R \to S$ be a ring homomorphism.

(i) Show that $0 \to R \xrightarrow{\varphi} S$ is exact if and only if φ is injective.

(ii) Show that $R \xrightarrow{\varphi} S \to 0$ is exact if and only if φ is surjective.

11.19. Let $k \in \mathbb{Z}$. Show that the following sequence is short exact.

$$0 \longrightarrow k\mathbb{Z} \xrightarrow{\iota} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/(k) \longrightarrow 0$$

11.20. Let R and S be rings. Show that the following sequence is short exact.

$$0 \longrightarrow R \xrightarrow{\iota_1} R \oplus S \xrightarrow{\pi_2} S \longrightarrow 0$$

11.21. **Every ring is contained in a unital ring.** Let R be a ring. We define a plus and times on the set $R^{(1)} = \mathbb{Z} \times R$ by

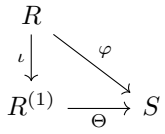
$$\begin{aligned} (a, r) + (b, s) &= (a + b, r + s) \\ (a, r) \cdot (b, s) &= (ab, as + br + rs) \end{aligned}$$

Show the following.

(i) These operations make $R^{(1)}$ a unital ring.

(ii) The map $\iota : R \to R^{(1)}$ given by $\iota(r) = (0, r)$ is an injective ring homomorphism.

(iii) If S is a unital ring and $\varphi : R \to S$ a ring homomorphism, then there is a unique unital ring homomorphism $\Theta : R^{(1)} \to S$ such that the following diagram commutes.



11.22. **A representation theorem for unital rings.** Let R be a ring, and let R^0 be the nullification of R (cf. (@@@)). Denote by $*$ the null product on R^0; that is, $r * s = 0$ for all $r, s \in R$. Recall (cf. (@@@)) that $\mathsf{End}\left(R^0\right)$ is a ring under pointwise addition and function composition.

(i) Given $r \in R$, define $\mu_r : R^0 \to R^0$ by $\mu_r(a) = ra$. Show that μ_r is a ring homomorphism.

(ii) Show that $\Psi : R \to \mathsf{End}\left(R^0\right)$ given by $\Psi(r) = \mu_r$ is a ring homomorphism.

(iii) Show that if R is unital then Ψ is injective.

That is, every unital ring is isomorphic to a ring of endomorphisms of a null ring. Combined with Exercise 11.21, *every* ring is isomorphic to a ring of endomorphisms of a null ring.

(Extended example - Generated Subrings?)

12 Domains and Fields

In College Algebra, how might we go about solving an equation like $2x = 6$ in the rational numbers? There are essentially two different, but related, strategies.

The simplest idea is to multiply both sides of the equation by $1/2$. Why $1/2$? Because that is precisely the number which, when multiplied by 2, yields 1, the multiplicative identity. The numbers 2 and $1/2$ enjoy this relationship with one another, just like 3 and $1/3$ or $5/7$ and $7/5$. But if r is an arbitrary element of an arbitrary ring R, there may be no way to find a special element like $\frac{1}{r}$ with the property that $r \cdot \frac{1}{r} = 1_R$. Such elements are very precious, so they get a name.

Def'n 12.1 (Unit)**.** Let R be a unital ring.

(i) We say that $u \in R$ is a *unit* if there is an element $u^{-1} \in R$, called an *inverse* of u, such that $uu^{-1} = u^{-1}u = 1_R$.

(ii) We say that R is a *field* if R is commutative and every nonzero element of R is a unit.

Examples

1. In any unital ring, 1_R is a unit.
2. In $\mathbb{Z}$ the only units are 1 and -1.

Prop'n 12.2. Let R be a unital ring.

(i) If $u \in R$ is invertible, then u^{-1} is unique in the following sense: if $v \in R$ such that $uv = vu = 1_R$, then $v = u^{-1}$.

(ii) If $\varphi : R \to S$ is a unital homomorphism and $u \in R$ is a unit, then $\varphi(u)$ is a unit and $\varphi(u)^{-1} = \varphi(u^{-1})$.

Proof. (i) Note that $v = v \cdot 1_R = v \cdot u \cdot u^{-1} = 1 \cdot u^1 = u^{-1}$.

(ii) Note that $1_S = \varphi(1_R) = \varphi(uu^{-1}) = \varphi(u)\varphi(u^{-1})$, and similarly $1_S = \varphi(u^{-1})\varphi(u)$. Thus $\varphi(u)$ is invertible, and we have $\varphi(u^{-1}) = \varphi(u)^{-1}$ by ((i)).

□

Let's revisit our motivating equation $2x = 6$. Another way to solve this equation is to rewrite it as $2x - 6 = 0$, and then factor as $2(x - 3) = 0$. Now the integers have the following very nice "zero product property":

If a and b are integers and $ab = 0$, then either $a = 0$ or $b = 0$.

From here we can break our equation into two: either $2 = 0$ or $x - 3 = 0$. Of course the first case is absurd (in $\mathbb{Z}$!) so we conclude that $x - 3 = 0$, so $x = 3$. The zero product property is also what allows the factorization method of solving equations. For instance, to solve an equation like $x^2 - 5x - 6 = 0$ over the integers, we factor the left hand side as $(x - 2)(x - 3) = 0$. Using the zero product property, either $x - 2 = 0$ (so $x = 2$) or $x - 3 = 0$ (so $x = 3$).

A given ring may or may not have this property. While $\mathbb{Z}$ does, $\mathbb{Z}/(n)$, for instance, may not; in $\mathbb{Z}/(6)$ we have $2 \neq 0$ and $3 \neq 0$, but $2 \cdot 3 = 0$. In this case we say that 2 and 3 are zerodivisors in $\mathbb{Z}/(6)$.

Def'n 12.3 (Zerodivisor)**.** Let R be a commutative ring.

(i) We say that a nonzero element $r \in R$ is a *zerodivisor* if there is a nonzero element $s \in R$ such that $rs = 0$.

(ii) We say that a commutative unital ring R is an *integral domain*, or simply *domain*, if R does not contain any zero divisors.

Note that this definition of zerodivisor applies only to *commutative* rings. Generalization to noncommutative rings is left to the exercises.

Prop'n 12.4 (Cancellation)**.** Let R be a domain with $r, s, t \in R$. If $rs = rt$ and $r \neq 0$, then $s = t$.

Proof. If $rs = rt$, then $rs - rt = 0$, so that $r(s - t) = 0$. Since $r \neq 0$, we must have $s - t = 0$, so that $s = t$. □

Prop'n 12.5. Every field is also an integral domain.

Proof. Let R be a field; in particular R is commutative. We need to show that R does not contain any zero divisors. To this end, suppose $r, s \in R$ such that $rs = 0$. If $r \neq 0_R$, then r is a unit in R. Now $r^{-1}rs = 0_R$, so that $1_R \cdot s = 0_R$, and thus $s = 0_R$. That is, either $r = 0_R$ or $s = 0_R$. So R does not contain any zero divisors. □

* * Exercises * *

12.1. Show that if R and S are nontrivial rings, then $R \oplus S$ is *not* a domain.

12.2. Show that every subring of a domain is a domain. In particular, every subring of a field is a domain.

12.3. Show that every domain (hence every field) has characteristic 0 or a prime. (cf. 6.11)

12.4. Show that $\mathbb{Z}/(n)$ is a field if and only if n is prime.

Def'n 12.6 (Group of Units)**.** Let R be a unital ring. The *group of units* of R is the set $\mathcal{U}(R) = \{u \in R \mid u \text{ is a unit}\}$.

12.5. Let R be a unital ring. Show that the following hold.

(i) 1_R is a unit.

(ii) If $u, v \in \mathcal{U}(R)$, then $uv \in \mathcal{U}(R)$.

(iii) If $u \in \mathcal{U}(R)$, then $u^{-1} \in \mathcal{U}(R)$.

12.6. Show that $\mathcal{U}(\mathbb{Z}) = \{1, -1\}$.

12.7. Show that $\mathcal{U}(\mathbb{Z}/(n)) = \{[a] \in \mathbb{Z}/(n) \mid \mathsf{gcd}(a, n) = 1\}$.

12.8. Let X be a nonempty set. Show that $\mathcal{U}(\mathcal{P}(X)) = \{X\}$.

12.9. Let A be a CU ring. Show that

$$\mathcal{U}(\mathsf{Mat}_2(A)) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid ad - bc \text{ is a unit in } A \right\}.$$

12.10. Show that if R is a domain, then the nilradical of R is 0. (cf. 11.10)

12.11. **Every finite domain is a field.** Let R be a *finite* integral domain. In this exercise we will show that R must be a field.

(i) Let $r \in R$ be a nonzero element and define a mapping $\varphi_r : R \to R$ by $\varphi_r(x) = rx$. Show that φ_r must be injective.

(ii) Deduce that φ_r must be bijective.

(iii) Deduce that r must be a unit in R, and then that R must be a field.

Def'n 12.7 (Fractional Integers)**.** Let $k \in \mathbb{Z}$ with $k \neq 0$. We define the set of *fractional integers* with denominator k to be

$$k^{-1}\mathbb{Z} = \left\{ \frac{a}{k^t} \mid a \in \mathbb{Z}, t \in \mathbb{N} \right\} \subseteq \mathbb{Q}.$$

12.12. Show that $k^{-1}\mathbb{Z}$ is a unital subring of $\mathbb{Q}$.

12.13. Show that the units in $k^{-1}\mathbb{Z}$ are of the form $\frac{\pm 1}{2^t}$ or $\pm 2^t$ where $t \in \mathbb{N}$.

(@@@) one-sided zerodivisors and units
(@@@) zero divisor graph
Skew fields.

13 Quadratic Extensions

In this section we will pause to construct an important family of example rings.

Def'n 13.1 (Quadratic Extension)**.** Let R be a commutative ring, and let $D \in R$. We define the *quadratic extension* of R by D to be $R[\sqrt{D}] = R \times R$ as a set, and define operations on $R[\sqrt{D}]$ as follows.

$$\begin{aligned}(r_1, r_2) + (s_1, s_2) &= (r_1 + s_1, r_2 + s_2) \\ (r_1, r_2) \cdot (s_1, s_2) &= (r_1 s_1 + D r_2 s_2, r_1 s_2 + r_2 s_1)\end{aligned}$$

We will write the element $(a, b) \in R[\sqrt{D}]$ as $a + b\sqrt{D}$.

The notation $R[\sqrt{D}]$ is intended to suggest that this structure behaves much like R, but now the element D has a "square root". It's important to remember that the plus in $a + b\sqrt{D}$ is a *structural* symbol, not an operation symbol.

Prop'n 13.2. Let R be a ring and $D \in R$.

(i) $R[\sqrt{D}]$ is a commutative ring.

(ii) If R is unital, then $R[\sqrt{D}]$ is unital.

(iii) The map $\iota : R \to R[\sqrt{D}]$ given by $\iota(r) = (r, 0_R)$ is a ring homomorphism. If R is unital, then ι is unital.

The most important use of quadratic extensions is to construct new domains and fields out of old ones.

Prop'n 13.3.

(i) If R is a domain, then $R[\sqrt{D}]$ is a domain if and only if the equation $Dy^2 = x^2$ has no nonzero solutions in R.

(ii) If R is a field, then $R[\sqrt{D}]$ is a field if and only if the equation $D = x^2$ has no nonzero solutions in R.

Proof. (i) (@@@)

(ii) (@@@)

□

Prop'n 13.4. Let D be an integer. We say that D not equal to 0 or 1 is *squarefree* if D is not divisible by a^2 for any $a > 1$.

(i) If D is squarefree, then the equation $Dy^2 = x^2$ has no nonzero solutions in $\mathbb{Z}$, and thus $\mathbb{Z}[\sqrt{D}]$ is a domain.

(ii) If D is squarefree, then the equation $D = x^2$ has no solutions in $\mathbb{Q}$, and thus $\mathbb{Q}[\sqrt{D}]$ is a field.

Proof. (i) We consider two cases: if $|D| \leq 1$ and if $|D| > 1$. In the first case, $D = -1$. Now if $Dy^2 = x^2$, then we have $-y^2 = x^2$; by sign considerations we have $x = y = 0$. In the second case, D is divisible by a prime, say p. If $x = 0$, then $y = 0$, and likewise if $y = 0$ then $x = 0$, so we can assume x and y are both nonzero. By the fundamental theorem of arithmetic, p appears the same number of times in the factorizations of Dy^2 and x^2. But note that p must appear an even number of times in x^2, but an odd number of times in Dy^2, a contradiction. Thus in either case the equation $Dy^2 = x^2$ has no nonzero solutions.

(ii) If $D = x^2$ has a rational solution $x = p/q$, then $Dq^2 = p^2$, which contradicts (i).

□

There are plenty of squarefree integers; in fact these are precisely the integers with no repeated prime factors. Each such integer yields a different domain $\mathbb{Z}[\sqrt{D}]$ and field $\mathbb{Q}[\sqrt{D}]$. These rings are particularly nice for a few reasons. They are easy to compute in, and while similar to $\mathbb{Z}$ and $\mathbb{Q}$, can exhibit wildly different behavior, as we will see in the next chapter.

* * Exercises * *

13.1. Prove Proposition 13.2.

13.2. Let $R = \mathbb{Z}/(3)$.

(i) Show that $2x^2 = y^2$ has no nonzero solutions in R. (Hint: When in doubt, use brute force.)

(ii) Conclude that $R[\sqrt{2}]$ is a field with 9 elements.

13.3. Elements of $\mathbb{Z}[\sqrt{-1}]$ are sometimes called *gaussian integers*.

(i) Determine which elements of $\mathbb{Z}[\sqrt{-1}]$ are units.

13.4. Show that $\mathsf{char}(R) = \mathsf{char}(R[\sqrt{D}])$.

13.5. Let R be a ring and $S \subseteq R$ a subring, with $D \in S$. Show that $S[\sqrt{D}]$ is a subring of $R[\sqrt{D}]$.

13.6. Let $\varphi : R \to S$ be a ring homomorphism with $D \in R$ and $E = \varphi(D)$. Show that there is a unique ring homomorphism $\Phi : R[\sqrt{D}] \to S[\sqrt{E}]$ such that the following diagram commutes.

$$\begin{array}{ccc} R & \xrightarrow{\varphi} & S \\ {\scriptstyle\iota}\downarrow & & \downarrow{\scriptstyle\iota} \\ R[\sqrt{D}] & \xrightarrow[\Phi]{} & S[\sqrt{E}] \end{array}$$

14 Divisibility

In addition to the arithmetic structure, the ring of integers has a *divisibility structure* which we are able to use to some effect. This is the language we use in $\mathbb{Z}$ to say things about greatest common divisors and unique factorization. And so we carry the "divides" concept to any commutative ring.

Def'n 14.1 (Divides)**.** Let R be a commutative ring with $a, b \in R$. We say a *divides* b, denoted $a|b$, if there is an element $c \in R$ such that $b = ac$. In this case we say that a is a *divisor* of b.

Divisibility captures some of the multiplicative structure on R and, as a relation, enjoys some of the familiar properties from $\mathbb{Z}$.

Prop'n 14.2. Let R be a commutative ring with $a, b, c \in R$. Then we have the following.

(i) $a|a$.

(ii) If $a|b$ and $b|c$ then $a|c$.

(iii) $a|0$.

Proof. (i) We have $a = a \cdot 1$, so $a|a$.

(ii) Suppose $a|b$ and $b|c$; then there exist $h, k \in R$ such that $b = ah$ and $c = bk$. Now $c = a(hk)$, so that $a|c$.

(iii) $0 = a \cdot 0$ as needed.

□

If R is unital, we can say a bit more.

Prop'n 14.3. Let R be a CU ring and $a, u \in R$ with u a unit.

(i) $1|a$ for all $a \in R$.

(ii) If $u \in R$ is a unit, then $u|a$.

(iii) If $u \in R$ is a unit and $a|u$, then a is a unit.

Proof. (i) Note that $a \cdot 1 = a$.

(ii) First note that $a = 1 \cdot a$, so that $1|a$. Now if $u \in R$ is a unit, then by definition we have an element v such that $1 = uv$, so that $u|1$. By 14.2(ii) we have $u|a$.

(iii) Say $ab = u$. Now $abu^{-1} = 1_R$, and thus a is a unit.

□

We should not let our intuition about divisibility in $\mathbb{Z}$ go too far, though, as divisibility in an arbitrary commutative ring can be very strange. For example, in $\mathbb{Z}$ if $a|b$ then in a specific sense a is "smaller than" b, and we cannot have never-ending chains of integers, each divisible by the next. This alone makes $\mathbb{Z}$ very special. In fact Chapter II is devoted to understanding how divisibility in R is different from divisibility in $\mathbb{Z}$, at least when R is a domain.

Two ring elements which differ only by a unit factor are indistinguishable from the perspective of divisibility; we call such pairs *associates*.

Def'n 14.4 (Associate)**.** Let R be a CU ring. If $a, b \in R$, we say a *is associate to* b, denoted $a \approx b$, if there is a *unit* $u \in R$ such that $b = ua$.

We will say quite a bit more about associate elements in Chapter II. For now, to help get a handle on the concept, we consider only two special cases.

Prop'n 14.5. Let R be a CU ring.

(i) If $0_R \approx a$ then $a = 0_R$.

(ii) $1_R \approx a$ if and only if a is a unit.

(iii) If $a \approx b$, then $ac \approx bc$.

Proof. (i) There is a unit u with $a = u0_R = 0_R$.

(ii) There is a unit u with $a = u1_R = u$.

(iii) If $b = ua$ for some unit u; then also $bc = uac$.

□

In a CU ring, every element is divisible by (1) units and (2) its associates. These are called *trivial divisors*. In general, a ring element will have more divisors. Some ring elements, however, have *only* the trivial divisors. These are special. In $\mathbb{Z}$, such elements are called *prime*. However, for some very good reasons that will be made clear later, in a general ring such elements are called *irreducible*.

Def'n 14.6 (Irreducible)**.** Let R be a CU ring and $x \in R$ a nonzero nonunit. We say that x is *irreducible* in R if, whenever $a, b \in R$ such that $x = ab$, either a or b is a unit.

For example, in $\mathbb{Z}$ the irreducible elements are precisely the prime integers in the usual sense. If R is a field, then by a quirk of logic R has no irreducible elements. (There are no nonzero nonunits!)

It is somewhat unfortunate that our most immediate examples of irreducible elements are the prime integers – you may be wondering why we use the word "irreducible" for the generalized concept. But recall that there are in fact two

different ways to characterize the prime integers; these are discussed in 4.1 and 4.3. In $\mathbb{Z}$ these two are equivalent, but in a general ring they are not – so we split the familiar concept of *primeness* into two.

Def'n 14.7 (Prime)**.** Let R be a CU ring and $p \in R$ a nonzero nonunit. We say that $p \in R$ is *prime* if whenever $p|ab$, either $p|a$ or $p|b$ (or both).

The distinction between "irreducible" and "prime" is a subtle one; at first glance we've simply replaced the equality by divisibility. Roughly, *irreducible* means *has no nontrivial divisors*, while *prime* means *cannot be nontrivially decomposed*. These two ideas are equivalent in $\mathbb{Z}$, as we know, but different in general. They are related, however, as in the following result.

Prop'n 14.8. If R is a domain, then every prime element is also irreducible.

Proof. Suppose $p \in R$ is prime, and factor p as $p = ab$. In particular, $p|ab$, and since p is prime, WLOG we have $p|a$. Say $a = pt$. Now $p = ab = ptb$, and by cancellation, $tb = 1$. In particular b is a unit. Thus p is irreducible. □

* * Exercises * *

14.1. Find all the divisors of 4 in $\mathbb{Z}$. Prove that your list is complete.

14.2. Let X be a nonempty set, and let $A, B \in \mathcal{P}(X)$. Show that $A|B$ if and only if $B \subseteq A$.

14.3. Let R and S be rings, with $(a, b), (r, s) \in R \oplus S$. Show that $(a, b)|(r, s)$ if and only if $a|r$ and $b|s$.

14.4. (@@@) if R is not a domain, yadda yadda converse of last exercise.

14.5. Let $\varphi : R \to S$ be a ring homomorphism with $a, b \in R$. Show that if $a|b$, then $\varphi(r)|\varphi(s)$.

14.6. Let $R = \mathcal{P}(\mathbb{N})$ and consider the elements of R of the form $r_k = [1, k]$. Show that $r_k|r_\ell$ if and only if $k \geq \ell$.

14.7. Let R be a CU ring. Show that "is associate to" is an equivalence relation.

14.8. Let R be a CU ring. Show that the following hold.

(i) If $x \approx y$, then $z|x$ if and only if $z|y$ for all $z \in R$.

(ii) If $x \approx y$, then $x|z$ if and only if $y|z$ for all $z \in R$.

14.9. Let R be a domain with $a, b \in R$. Show that if $a|b$ and $b|a$ then $a \approx b$.

14.10. Let R be a domain with $a, b, c \in R$. Show that if $ac|bc$ then $a|b$.

14.11. Let R be a domain and $a, b \in R$ such that $a \approx b$. Show that the following hold.

(i) If a is irreducible, then b is irreducible.

(ii) If a is prime, then b is prime.

14.12. Let R be a domain with $a, b, c \in R$ and $c \neq 0$. Show that if $ac \approx bc$, then $a \approx b$.

14.13. Let R and S be CU rings. Show that $(r_1, s_1) \approx (r_2, s_2)$ in $R \oplus S$ if and only if $r_1 \approx r_2$ in R and $s_1 \approx s_2$ in S.

14.14. Definition 14.1 requires R to be a commutative ring, but is this really necessary? (Divisibility in noncommutative rings.)

15 Norms

In $\mathbb{Z}$ there are lots of neat things we can do with irreducibles. Every integer is a product of irreducibles, $\mathbb{Z}/(n)$ is a field precisely when n is irreducible, and algorithms for detecting irreducibles have practical applications. In a general ring we might hope that some of these neat things generalize. First, we need to be able to detect whether or not an element is irreducible. In $\mathbb{Z}$ we have a brute-force algorithm for this thanks to induction, but most rings do not have an obvious analogue of this strategy.

In some rings (not all!) we can make progress on the problem of finding irreducibles by mapping the multiplicative structure of R to the $\mathbb{N}$ – doing this we can take advantage of what we know about natural numbers and, sometimes, recover the benefits of induction.

Def'n 15.1 (Norm)**.** Let R be a domain. A mapping $N : R \to \mathbb{N}$ is called a *norm* if the following hold.

N1. $N(\alpha) = 0$ iff $\alpha = 0$.

N2. If $\alpha, \beta \in R$ are nonzero then $N(\alpha) \leq N(\alpha\beta)$.

We say N is *multiplicative* if it satisfies the additional constraint that

N3. $N(\alpha\beta) = N(\alpha)N(\beta)$ for all $\alpha, \beta \in R$.

Examples

1. $N : \mathbb{Z} \to \mathbb{N}$ given by $N(a) = |a|$ is a multiplicative norm.
2. $N : \mathbb{Z}[i] \to \mathbb{N}$ given by $N(a + bi) = a^2 + b^2$ is a multiplicative norm.
3. More generally, $N : \mathcal{O}(\sqrt{D}) \to \mathbb{N}$ given by $N(a + b\sqrt{D}) = |a^2 + Db^2|$ if $D \equiv 2, 3 \mod 4$ and $N(a + b\frac{1+\sqrt{D}}{2}) = |a^2 + ab + b^2\frac{1-D}{4}|$ if $D \equiv 1 \mod 4$ is a multiplicative norm.
4. If R has a norm N, then the quadratic extension of R by D has a norm $M(a + b\sqrt{D}) = N(a^2 - Db^2)$. (@@@)

If a ring has a multiplicative norm, we can use it to detect (some) irreducible elements.

Prop'n 15.2. Let R be a domain and $N : R \to \mathbb{N}$ a multiplicative norm. If $\alpha \in R$ such that $N(\alpha)$ is prime in $\mathbb{N}$, then α is irreducible in R.

For example, consider $\mathbb{Z}[i]$. Applying this result here, we see that $a \pm bi$ is irreducible if $a^2 + b^2$ is prime. In particular $1 \pm i$, $1 \pm 2i$, $2 \pm 3i$, and many other Gaussian integers are irreducible (since $1^2 + 1^2 = 2$, $1^2 + 2^2 = 5$, and

$2^2 + 3^2 = 13$ are prime). This leads to a natural question about the natural numbers: for which primes p does the equation $a^2 + b^2 = p$ have a solution?

As this example shows, a good multiplicative norm can turn questions in R into number theory problems. This turns out to be a useful technique more generally: given a problem about some object, look for a way to map the relevant structure of that object to some other object which either is well-understood or with which we can compute things. A good strategy for solving algebraic problems is to try to reduce to number theory or to linear algebra.

* * EXERCISES * *

15.1. Let F be a field. Show that the map $N : F \to \mathbb{N}$ given by

$$N(x) = \begin{cases} 0 & \text{if } x = 0_F \\ 1 & \text{otherwise} \end{cases}$$

is a multiplicative, unit-preserving norm.

15.2. Let D be a squarefree integer.

(i) Show that (@@@) given by (@@@) is a multiplicative norm.

(ii) Show that if $D < -1$, then the units in (@@@) are ± 1.

15.3. Let R be a domain and N a multiplicative norm on R.

(i) Show that $N(1) = 1$.

(ii) Show that $N(u) = 1$ for any unit u.

15.4. Draw the Hasse diagram of divisors of $2\sqrt{-2}$ in zee adjoin sqrt minus 2. (@@@) make this one multipart

Def'n 15.3. We say a norm is *unit-separating* if $N(x) = 1$ implies that x is a unit.

Summary of Chapter I

— II —

The Domain Hierarchy

In Chapter I we defined a class of structures, called *rings*, which generalize the basic arithmetic on $\mathbb{Z}$ and $\mathbb{Z}/(n)$. The ring of integers comes equipped with some extra technology which turns out to be useful: specifically, the **division algorithm**, **unique factorization**, and the existence of **greatest common divisors**. In this chapter we will see how this technology generalizes (or not!) to other rings – to integral domains, in particular. At the risk of giving away the punchline, we will see that the technology on $\mathbb{Z}$, when reimagined in arbitrary domains, gives rise to a **hierarchy** of families of domains of increasing algorithmic power.

16 Associates

Recall that if R is a CU ring, we say that two elements $a, b \in R$ are *associate*, denoted $a \approx b$, if there is a unit $u \in R$ such that $a = ub$. It turns out that the divisibility structure of R cannot distinguish associates from each other. Considering associate elements to be "the same", we have a nice way to visualize the divisibility structure of a ring.

Recall that if σ is an equivalence relation on a set X, then we have a canonical *partition* of X induced by σ given by the σ-classes. This partition, considered as a set, is called the *quotient* of X by σ, and acts very much like X does except that elements of X which are σ related become equal in X/σ. This is a powerful idea which we will use again in Chapter IV. For now, we will use the classes of the associate relation to reason about divisibility.

Def'n 16.1 (Associate Structure)**.** Let R be a CU ring. Recall by Exercise 14.7 that $\approx$ is an equivalence relation on R. The set $\mathcal{A}(R) = R/\approx$ of all $\approx$-equivalence classes of R is called the *associate structure* of R.

Note that, by definition, the associate class of a is all the elements of the form ua where u is a unit. If R is small enough we can find its associate classes by hand. For example, the associate classes of $\mathbb{Z}/(4)$ are $[0] = \{0\}$, $[1] = \{1, 3\}$, and $[2] = \{2\}$. The ring $\mathbb{Z}$ is not very small, but we can see that its associate classes are $[0]$ and $[n] = \{n, -n\}$ where $n > 0$. As promised, the divisibility relation on R can be bumped up to a kind of relation on $\mathcal{A}(R)$.

Prop'n 16.2. Let R be a CU ring. We define a relation $\lesssim$ on $\mathcal{A}(R)$ as follows: given two associate classes A and B, we say $A \lesssim B$ if there exist $a \in A$ and $b \in B$ such that $a|b$. Then the following hold for all $A, B, C \in \mathcal{A}(R)$.

(i) If $A \lesssim B$, then in fact $a|b$ for *any* $a \in A$ and $b \in B$.

(ii) $A \lesssim A$.

(iii) If $A \lesssim B$ and $B \lesssim C$, then $A \lesssim C$.

(iv) If R is a domain, then if $A \lesssim B$ and $B \lesssim A$ then $A = B$.

Proof. (i) Since $A \lesssim B$, there exist $x \in A$ and $y \in B$ such that $x|y$. Now $a \approx x$ and $b \approx y$, so using Exercise 14.8, we have $x|y$ if and only if $x|b$, if and only if $a|b$.

(ii) Let $a \in A$; now $a|a$ as needed.

(iii) Let $a \in A$, $b \in B$, and $c \in C$. We have $a|b$ and $b|c$, so that $a|c$.

(iv) Let $a \in A$ and $b \in B$. We have $a|b$ and $b|a$; say $b = xa$ and $a = yb$. Substituting, we have $1_r b = b = xyb$; by cancellation, $xy = 1_R$, and thus x and y are units as needed.

□

For any ring R the $\lesssim$ relation is both reflexive and transitive. Such relations are sometimes called *preorders*, and can be visualized using a kind of picture called a *Hasse diagram*. To draw the Hasse diagram of a preorder like $\mathcal{A}(R)$, we draw each element on the page so that if $A \lesssim B$ then A is *below* B. Then we connect A and B by a line if and only if there are no other elements C such that $A \lesssim C \lesssim B$. In the special case that A and B are distinct but $A \lesssim B$ and $B \lesssim A$, we simply write A and B adjacent to each other. We call the Hasse diagram of $\mathcal{A}(R)$ the *divisor diagram* of R.

As an example, consider the ring $\mathbb{Z}/(12)$. The associate classes of this ring are $[0] = \{0\}$, $[1] = \{1, 5, 7, 11\}$, $[2] = \{2, 10\}$, $[3] = \{3, 9\}$, $[4] = \{4, 8\}$, and $[6] = \{6\}$; the divisor diagrams of $\mathbb{Z}/(12)$ and two other rings are shown in Figure 16.1.

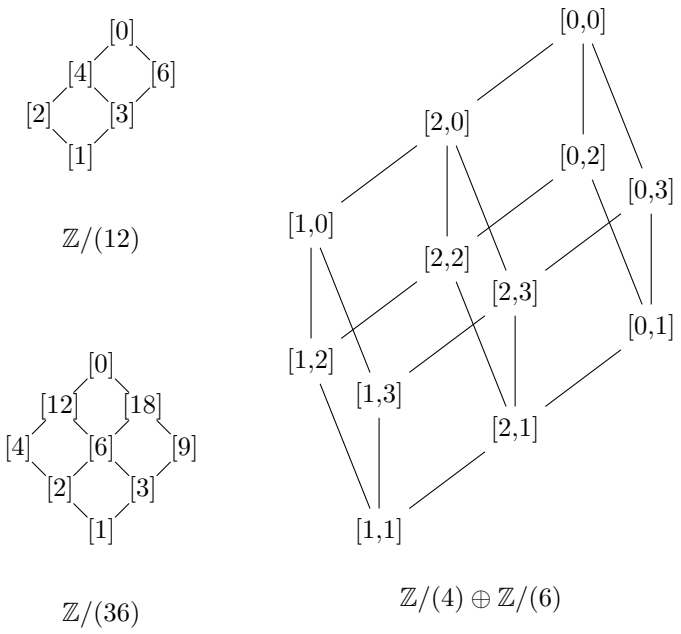


Figure 16.1: Some divisor diagrams.

Even if the divisor diagram of R is infinite it may be helpful to see small parts of it at a time. If $a \in R$ is an element, the divisor diagram of a in R is simply the hasse diagram of the divisors of a. Divisor diagrams provide an important answer to what we might call the Visualization Problem.

The Visualization Problem

Given a ring, can we draw a picture of it?

In addition to divisor diagrams both cayley tables and commutative diagrams are also useful visualization tools, and we will see others. For infinite rings, or even just large finite rings, the complete divisor diagram may not be very helpful. But even just looking at a portion of the diagram can give insight. Seeing a ring in this way can also lead us to make conjectures. For instance, perhaps R is infinite, but $\mathcal{A}(R)$ is finite, or has only finite "height", or "width" (whatever that means), or has only finitely many minimal or maximal elements. Does that imply anything about R? Do divisor diagrams always have the boxy shape of the examples in Figure 16.1?

* * EXERCISES * *

16.1. Compute the associate classes of $\mathbb{Z}$.

16.2. Draw the divisor diagrams of the following rings.

(i) $\mathbb{Z}/(8)$

(ii) $\mathcal{P}(\{a, b, c\})$

(iii) $\mathbb{Z}/(4) \oplus \mathbb{Z}/(4)$

(iv) $\mathbb{Z}/(36)$

16.3. Show that R is a field precisely when $\mathcal{A}(R)$ consists only of $[0]$ and $[1]$.

16.4. (@@@) find a CU nondomain and a, b where $a|b$ and $b|a$ but a, b not associate.

16.5. Show that divisibility on $\mathbb{Z}/(4)$ is symmetric.

16.6. Let R be a domain. Show that $a \in R$ is irreducible if and only if $[a]$ is $\lesssim$-minimal.

16.7. Let R be a CU ring. We define a kind of setwise product on associate classes as follows: given $A, B \in \mathcal{A}(R)$, we define

$$A * B = \{ab \mid a \in A, b \in B\}.$$

Show that the following hold.

(i) $A * B \in \mathcal{A}(R)$, so that $*$ is a binary operation on $\mathcal{A}(R)$.

(ii) $(A * B) * C = A * (B * C)$ for all $A, B, C \in \mathcal{A}(C)$.

(iii) $A * B = B * A$ for all $A, B \in \mathcal{A}(R)$.

(iv) If R is a domain and $A * C = B * C$, then $A = B$.

Def'n 16.3 (Monotone Map)**.** Recall that a *preorder* is a set P equipped with a relation $\lesssim$ which is reflexive and transitive. If P and Q are preorders, a mapping $\varphi : P \to Q$ is called *monotone* if whenever $a \lesssim b$ in P we have $\varphi(a) \lesssim \varphi(b)$ in Q. Similarly, we say φ is *antitone* if whenever $a \lesssim b$ in P we have $\varphi(b) \lesssim \varphi(a)$ in Q.

16.8. Let R be a CU ring.

(i) Show that R is a preorder under the divisibility relation.

(ii) Show that the mapping $\varphi : R \to \mathcal{A}(R)$ given by $\varphi(r) = [r]$ is a preorder homomorphism.

16.9. Let $\varphi : R \to S$ be a ring homomorphism, where R and S are commutative. Show that, if we consider R and S to be preorders under divisibility, then φ is monotone.

16.10. Let $\varphi : R \to S$ be a unital ring homomorphism.

(i) Show that if $a, b \in R$ and $a \approx b$, then $\varphi(a) \approx \varphi(b)$.

(ii) Show that the relation $\Phi \subseteq \mathcal{A}(R) \times \mathcal{A}(S)$ given by

$$\Phi = \{([a], [\varphi(a)]) \mid a \in R\}$$

is well-defined and total.

(iii) Show that Φ is monotone.

(iv) Show that if φ is surjective, then Φ is surjective.

17 Greatest Common Divisors and GCD Rings

In $\mathbb{Z}$ any two elements have a greatest common divisor, and this concept turns out to have some interesting uses. We generalize GCDs to arbitrary CU rings as follows.

Def'n 17.1 (Greatest Common Divisor)**.** Let R be a CU ring with $a, b \in R$. We say $d \in R$ is a *greatest common divisor* of a and b if the following hold.

(i) d is a *common divisor* of a and b; that is, $d|a$ and $d|b$.

(ii) d is greatest among the common divisors of a and b; that is, if $c \in R$ such that $c|a$ and $c|b$, then $c|d$.

The set of all greatest common divisors of a and b is denoted $\mathsf{gcd}(a, b)$. We say that a and b are *relatively prime*, denoted $a \perp b$, if $1_R \in \mathsf{gcd}(a, b)$.

Watch out: sometimes in $\mathbb{Z}$ the greatest common divisor is defined to be the *largest* common divisor with respect to the $\leq$ relation. But in an arbitrary domain we may not have a meaningful order relation, so the word "largest" is interpreted here with respect to divisibility. An important consequence of this is that in a general domain, greatest common divisors need not exist, and if they do, they need not (and generally won't) be unique. For example by our definition we have $\mathsf{gcd}(4, 6) = \{2, -2\}$ in $\mathbb{Z}$.

Prop'n 17.2. Let R be a domain with $a, b, c \in R$. Then we have the following.

(i) $\mathsf{gcd}(a, b)$ is either empty or an associate class.

(ii) If $a \approx b$, then $\mathsf{gcd}(a, c) = \mathsf{gcd}(b, c)$.

Proof. (i) Let $a, b \in R$ and suppose $\mathsf{gcd}(a, b)$ is not empty; say x is a greatest common divisor of a and b. It suffices to show that if $y \approx x$, then y is also a greatest common divisor of a and b. By Exercise 14.8 we have that $y|a$ and $y|b$, and if $c|a$ and $c|b$, then $c|x$, so that $c|y$ by Exercise 14.8.

(ii) Say $b = ua$, and suppose we have $d \in \mathsf{gcd}(a, c)$. Now $d|a$, so that $d|b$ by Exercise 14.8. That is, d is a common divisor of b and c. Suppose now that $e|b$ and $e|c$. We have $e|a$ by Exercise 14.8, so that $e|d$. Thus $d \in \mathsf{gcd}(b, c)$. Similarly, if $d \in \mathsf{gcd}(b, c)$ then $d \in \mathsf{gcd}(a, c)$. That is, if *either* $\mathsf{gcd}(a, c)$ or $\mathsf{gcd}(b, c)$ is nonempty, then the two are equal. □

It is important to remember that $\mathsf{gcd}(*, *)$ is a *subset* of R, not a specific element. However, Proposition 17.2 says that nonempty GCDs are associate classes, and that GCD cannot distinguish elements in the same associate class.

This suggests that we should really think of $\gcd(*,*)$ as a binary operation on the associate classes of R; this is explored further in Exercise 17.8.

In $\mathbb{Z}$, we had some nice technology for computing GCDs – namely, the Euclidean Algorithm. In an arbitrary domain the situation is not so nice; computing $\gcd(a,b)$ is generally difficult. The next proposition handles some special cases.

Prop'n 17.3. Let R be a domain with $a, b \in R$. Then we have the following.

(i) $\gcd(a,b) = \gcd(b,a)$.

(ii) $a|b$ if and only if $a \in \gcd(a,b)$.

(iii) $a \in \gcd(a,0)$.

(iv) If $u \in R$ is a unit then $a \perp u$.

(v) If $a \perp b$ and $c|a$, then $c \perp b$.

Proof. (i) Follows from the definition of GCD.

(ii) Suppose $a|b$. Now $a|a$ and $a|b$, so a is a common divisor of a and b. Moreover if $c|a$ and $c|b$, then $c|a$, so a is a greatest common divisor of a and b. The converse follows from the definition of common divisor.

(iii) Note that $a|a$ and $a|0$ by 14.2, so a is a common divisor of a and 0. Certainly if $c|a$ and $c|0$ then $c|a$, as needed.

(iv) It suffices to show that 1_R is a greatest common divisor of a and u. Certainly $1_R|a$ and $1_R|u$. Now if $c|a$ and $c|u$, then c is a unit by 14.2**??**, and thus $c|1_R$ as needed.

(v) If $x \in \gcd(c,b)$, then we have $x|a$ and $x|b$, so that $x|1_R$ as needed.

□

The next result is very important: it roughly says that if a divides a product but is relatively prime to one of the factors, then it must divide the other factor.

Prop'n 17.4 (Euclid's Lemma)**.** Let R be a domain with $a, b, c \in R$.

(i) If $d \in \gcd(a,b)$ and $e \in \gcd(ac,bc)$ then $e \approx dc$.

(ii) If $a \perp b$, $\gcd(ac,bc)$ exists, and $a|bc$, then $a|c$.

Proof. (i) Note that $d|a$ and $d|b$, so that $dc|ac$ and $dc|bc$, and thus $dc|e$. Say $e = dcx$. Now $dcx|ac$ and $dcx|bc$, so that $dx|a$ and $dx|b$. Now $dx|d$, so that $x|1_R$. That is, x is a unit, and we have $e \approx dc$.

(ii) Say $e \in \gcd(ac, bc)$. Now $a|ac$ and $a|bc$, so $a|e$. By (i), we have $e \approx c$, thus $a|c$ as needed.

□

In fact many nice properties hold for greatest common divisors, all of which are contingent on GCDs existing in the first place – and this is not guaranteed! That is, there are some domains, like $\mathbb{Z}$, where $\gcd(a, b)$ is always nonempty, but there are other domains containing elements which do not have any greatest common divisors. (We explore some examples in the exercises.) Those rings, like $\mathbb{Z}$, where GCDs always exist are special enough to warrant a definition.

Def'n 17.5 (GCD Ring)**.** Let R be a domain. We say that R is a *GCD ring* if any two elements of R have a greatest common divisor. If R is also a domain, we say it is a *GCD domain*.

Of course our motivating example $\mathbb{Z}$ is a GCD domain, and we will see many more. First, we establish that several of the nice things we can say about GCDs in $\mathbb{Z}$ hold in an arbitrary GCD domain.

Prop'n 17.6. Let R be a GCD domain with $a, b, c \in R$.

(i) If $a \perp b$ and $a \perp c$, then $a \perp bc$.

(ii) If $a \perp b$, then $\gcd(a, bc) = \gcd(a, c)$.

(iii) If $d \in \gcd(a, b)$ with $a = dx$ and $b = dy$, then $x \perp y$.

(iv) If $a \perp b$, $d \in \gcd(ab, c)$, $e \in \gcd(a, c)$, and $f \in \gcd(b, c)$, then $d \approx ef$.

Proof. (i) Suppose $d|a$ and $d|bc$. Now $d|ab$ and $d|cb$, so that $d|e$, where $e \in \gcd(ab, cb)$. By 17.4(i) we have that $e|b$. But now $e|a$ and $e|b$, so that $e|1_R$; thus 1_R is a greatest common divisor of a and bc.

(ii) Let $d \in \gcd(a, bc)$ and $e \in \gcd(a, c)$; it suffices to show that $d \approx e$. To this end, note that $e|a$ and $e|c$, so that $e|bc$; thus $e|d$. Now $d|a$, so that $d|ac$, and $d|bc$. Letting $f \in \gcd(ac, bc)$, we have $f \approx 1_R \cdot c$, and thus $d|c$. So we have $d|e$, and thus $d \approx e$ as needed.

(iii) Let $e \in \gcd(x, y)$. Now by 17.4(i) we have $d \approx ed$, and thus $1_R \approx e$ by **????**, as needed.

(iv) Say $a = ex$ and $c = ey$ where $x \perp y$. Note that $b \perp e$ by 17.3(v), so that

$$\gcd(b, c) = \gcd(b, ey) = \gcd(b, y) = \gcd(xb, y)$$

by 17.6(ii). In particular, we have $d \approx ef$ by 17.4(i).

□

Recall that in a general domain we can characterize the "indivisible" elements in two slightly different ways: the irreducibles and the primes. In any domain, every prime is also irreducible. In a GCD domain, the converse also holds.

Prop'n 17.7. If R is a GCD domain, then every irreducible element of R is also prime.

Proof. Let p be irreducible and suppose $p|ab$. Let $d \in \mathsf{gcd}(a,p)$, and write $a = da'$ and $p = dp'$. Since p is irreducible, either d or p' is a unit. If d is a unit, then we have $p|b$ by Euclid's Lemma. If p' is a unit, then $p|a$. □

Remember that unlike $\mathbb{Z}$, in a GCD domain we only know that greatest common divisors *exist* – we don't necessarily have a way to actually compute them. In the exercises we will compute GCDs (@@@)

* * EXERCISES * *

17.1. Let R and S be domains, and let $x, y \in R \oplus S$. Show that if $x|y$ and $y|x$ then $x \approx y$.

17.2. Let R and S be GCD domains with $r_1, r_2 \in R$ and $s_1, s_2 \in S$, and let $d_r \in \mathsf{gcd}(r_1, r_2)$ and $d_s \in \mathsf{gcd}(s_1, s_2)$. Show that $(d_r, d_s) \in \mathsf{gcd}((r_1, s_1), (r_2, s_2))$.

17.3. Let X be a nonempty set and let $a, b \in \mathcal{P}(X)$. Show that if $a|b$ and $b|a$ then $a = b$.

17.4. Let X be a nonempty set, and let $A, B \in \mathcal{P}(X)$. Show that $A \cup B \in \mathsf{gcd}(A, B)$.

17.5. (@@@) Let $n > 1$.

(i) Show that $k \approx \mathsf{gcd}(k, n)$.

(ii) Show that the associate classes of $\mathbb{Z}/(n)$ are of the form $\mathsf{gcd}(k, n)$ for $0 \leq k < n$.

(iii) Show that if $a|b$ and $b|a$ then $a \approx b$.

17.6. $\mathbb{Z}/(n)$ **has greatest common divisors.** (@@@) Let $n > 1$, let $a, b \in \mathbb{Z}/(n)$, and let d be the greatest common divisor of a, b, and n in $\mathbb{Z}$. Show that $d \in \mathsf{gcd}(a, b)$ in $\mathbb{Z}/(n)$.

17.7. (@@@) find an example of two ring elements which divide each other but are not associate.

17.8. Let R be a CU ring, and let $\mathcal{A}(R)$ be the set of $\approx$-equivalence classes of R (see 16.1). Define a relation $\wedge \subseteq (\mathcal{A}(R) \times \mathcal{A}(R)) \times \mathcal{A}(R)$ as follows.

$$\wedge = \{(([a],[b]), \mathsf{gcd}(a,b)) \mid a, b \in R \text{ such that } \mathsf{gcd}(a,b) \neq \varnothing\}.$$

(i) Show that $\wedge$ is well-defined.

(ii) Show that $\wedge$ is total if and only if R is a GCD domain.

That is, $\wedge$ is a partial function $\mathcal{A}(R) \times \mathcal{A}(R) \to \mathcal{A}(R)$ which is total if and only if R is a GCD domain.

17.9. Let R be a GCD domain, and let $\wedge$ be defined on $\mathcal{A}(R)$ as in Exercise 17.8. Show that the following hold for all $A, B, C \in \mathcal{A}(R)$.

(i) $A \wedge A = A$

(ii) $A \wedge B = B \wedge A$

(iii) $(A \wedge B) \wedge C = A \wedge (B \wedge C)$

17.10. **A domain which is not a GCD domain.** In this exercise we will show that $\mathbb{Z}[\sqrt{-3}]$ is a domain which is not a GCD domain.

(i) Show that the equation $a^2 + ab + b^2 = 2$ has no solutions in $\mathbb{Z}$.

(ii) Show that no element of $\mathbb{Z}[\sqrt{-3}]$ has norm 2 (@@@).

(iii) Show that 2 is irreducible in $\mathbb{Z}[\sqrt{-3}]$.

(iv) Show that 2 divides $4 = (1+\sqrt{-3})(1-\sqrt{-3})$ in $\mathbb{Z}[\sqrt{-3}]$, but that 2 does not divide $1+\sqrt{-3}$ or $1-\sqrt{-3}$. In particular, 2 is not prime in $\mathbb{Z}[\sqrt{-3}]$.

(v) Conclude that $\mathbb{Z}[\sqrt{-3}]$ is not a GCD domain.

Def'n 17.8 (Least Common Multiple)**.** Let R be a domain, with $a, b \in R$. We say that $m \in R$ is a *least common multiple* of a and b in R if the following hold.

(i) m is a *common multiple* of a and b; that is, $a|m$ and $b|m$.

(ii) m is least among the common multiples of a and b; that is, if $n \in R$ such that $a|n$ and $b|n$, then $m|n$.

The set of all least common multiples of a and b is denoted $\mathsf{lcm}(a, b)$.

17.11. (@@@) example where euclid's lemma fails

17.12. (@@@) Show that $k^{-1}\mathbb{Z}$ is a GCD domain.

17.13. (@@@) is this true: if R is a domain in which every irreducible is prime, then R is a GCD domain.

18 Factorization and UFDs

In $\mathbb{Z}$ we have the Fundamental Theorem of Arithmetic: every integer can be written as a product of prime integers in essentially one way, in the sense that any two factorizations have the same length and can be rearranged so that corresponding factors are associate. For example,

$$30 = (-2) \cdot 5 \cdot (-3) \quad \text{and} \quad 30 = 3 \cdot (-5) \cdot (-2)$$

are different factorizations of 30, but we can rearrange them as

$$30 = (-2) \cdot (-3) \cdot 5 \quad \text{and} \quad 30 = (-2) \cdot 3 \cdot (-5),$$

where -2 and -2 are associate, -3 and 3 are associate, and 5 and -5 are associate.

FTA is merely a statement about the *existence* of these unique factorizations, but if we examine the usual proof, we see that it can be turned into an effective algorithm which actually finds a factorization. The most basic (but correct!) algorithm to actually *compute* the prime factorization of a given integer n goes something like this.

To find a prime factorization of n:
Search among all a with $1 < a < |n|$ for a divisor of n.

- If no such divisor is found, then n is prime and thus $n = n$ is a prime factorization of n.
- If such a divisor is found, with $n = ab$, then *recursively* find prime factorizations $a = p_1 \cdots p_h$ and $b = q_1 \cdots q_k$. Then $n = p_1 \cdots p_h q_1 \cdots q_k$ is a prime factorization of n.

There are some minor optimizations we can make to this procedure, say by only looking for *prime* divisors up to $\sqrt{n}$. In the end this **trial division** procedure is slow, but correct.

We might try to generalize this procedure to any domain by replacing the word "prime" with "irreducible". However, there are many subtle ways it can fail. For a particular domain R and a particular element $x \in R$,

- Maybe there is no way to restrict the possible divisors of x to a finite list of candidates.
- Maybe R has no irreducible elements.
- Maybe R does have irreducible elements, but the factorization procedure never terminates. (Every "factorization" of x has infinite length.)
- Maybe the factorization procedure terminates only for *some* choices of the divisors a and b. (Some "factorizations" of x have infinite length.)

- Maybe the factorization procedure always terminates, but x has arbitrarily long factorizations.
- Maybe the factorizations of x are all of bounded length, but generally have *different* lengths.
- Maybe the factorizations of x are all of the same length, but cannot generally be rearranged so that corresponding factors are associates.

By the way, there are examples of domains where each of these things happens! Evidently, then, a ring like $\mathbb{Z}$ where the Fundamental Theorem of Arithmetic holds is very special, since it manages to avoid all these problems. We single out rings with unique factorization with a definition.

Def'n 18.1 (Unique Factorization Domain)**.** Let R be a domain. We say that R is a *unique factorization domain* (UFD) if every nonzero element of R can be written as a finite product of irreducibles in essentially one way.

More precisely, for each $x \in R$, we have $x = p_1 p_2 \cdots p_m$ where each p_i is irreducible, and if $x = q_1 q_2 \cdots q_\ell$, for some irreducible q_i, then $m = \ell$ and there is a permutation σ of $\{1, \ldots, m\}$ so that p_i and $q_{\sigma(i)}$ are associates for each i.

Of course $\mathbb{Z}$ is our prototypical example of a UFD. The value of a UFD is that irreducibles act like the "building blocks" for all other elements, much like the prime numbers are the building blocks of the integers. For instance, irreducible factorizations give us a nice characterization of divisibility as follows.

Prop'n 18.2. Let R be a UFD, and let $x, y \in R$. Suppose we have factored x and y as

$$x = p_1^{e_1} p_2^{e_2} \cdots p_m^{e_m} \quad \text{and} \quad y = u p_1^{f_1} p_2^{f_2} \cdots p_m^{f_m},$$

where the e_i and f_i are natural numbers. (Note that this is always possible by taking associates and letting e_i or f_i be zero in case some irreducible does not appear in the factorization of x or y.) Then $x|y$ if and only if $e_i \leq f_i$ for each $i \in [1, m]$.

Proof. (@@@) □

From here, we can characterize gcds nicely.

Prop'n 18.3. Every UFD is a GCD Domain. Specifically, if R is a UFD and $x, y \in R$ such that

$$x = p_1^{e_1} p_2^{e_2} \cdots p_m^{e_m} \quad \text{and} \quad y = u p_1^{f_1} p_2^{f_2} \cdots p_m^{f_m}.$$

Then
$$z = p_1^{\min(e_1,f_1)} p_2^{\min(e_2,f_2)} \cdots p_m^{\min(e_m,f_m)}$$
is a maximal common divisor of x and y.

Unique factorization is a very natural thing to want in a ring. In many areas of mathematics, one way to try to understand some object is to decompose it into smaller pieces in a simple way, understand the smaller pieces, and then reassemble into an understanding of the original object. It is especially nice, then, if there is *only one way* that the decomposition process could proceed. UFDs are very rare, but fortunately many rings of practical importance (such as $\mathbb{Z}$) do have unique factorization.

One of the historical threads that led to the development of modern abstract algebra was an attempt by 19th century mathematician Ernst Kummer to prove Fermat's Last Theorem using factorization. Recall that FLT is the assertion that the equation $a^n + b^n = c^n$ has no interesting integer solutions (a, b, c) if $n > 2$. Kummer's idea was to rearrange this equation as $a^n = c^n - b^n$. Now if ζ is a primitive nth root of unity, the right hand side factors as

$$a^n = (c - b)(c - \zeta b)(c - \zeta^2 b) \cdots (c - \zeta^{n-1} b).$$

For example, if $n = 4$ then i is a primitive 4th root of unity and we have

$$a^4 = (c - b)(c - ib)(c + b)(c + ib)$$

This yields two different factorizations of a^n, which may lead to a contradiction.

The problem with Kummer's idea was that it only works if the ring $\mathbb{Z}[\zeta]$ is a UFD, which, at the time, was not known. Nobody had bothered to check! The idea that this line of attack might solve the famous FLT led to an explosion of new ideas in algebraic number theory that continues to this day.

Eventually it was found that $\mathbb{Z}[\zeta]$ is sometimes a UFD, but not always – so Kummer's idea did not work. But it is more useful to judge an idea not by what problems it *can't* solve, but by what problems it *can* and what new ideas it inspires. By this measure Kummer's attempt at FLT is among the most successful failures in mathematics.

* * EXERCISES * *

18.1. Suppose that R is a domain and $N : R \to \mathbb{N}$ a multiplicative norm. Show that if $x \in R$, then the irreducible factorizations of x have bounded length.

Def'n 18.4. Let R be a domain and $p, x \in R$, with $x \neq 0$. We define the p-divisibility order of x, denoted $\mathsf{div}_p(x)$, to be the exponent of the largest power of p which divides x (if it exists). That is, $\mathsf{div}_p(x) = \max\{k \mid p^k | x\}$ if this set has a maximum and is undefined otherwise.

18.2. Let R be a UFD.

(i) Show that $\mathsf{div}_p(x)$ exists for all $p, x \in R$ with $x \neq 0$.

(ii) Show that if p is irreducible, then $\mathsf{div}_p(ab) = \mathsf{div}_p(a) + \mathsf{div}_p(b)$ for all $a, b \in R$.

(iii) Show that $N : R \to \mathbb{N}$ given by $N(x) = 2^{\mathsf{div}_p(x)}$ if $x \neq 0_R$ and 0 if $x = 0_R$ is a multiplicative norm on R.

19 Division with Remainder and Euclidean Domains

In $\mathbb{Z}$, we have the extremely important Division Algorithm. This theorem states that if a and b are integers with $b \neq 0$, then there exists a "quotient" q and a "remainder" r such that $a = qb + r$, and, moreover, the remainder is not too large – $0 \leq r < |b|$. This is the result from which most of the interesting results and algorithms in $\mathbb{Z}$ spring.

We'd like to generalize this property to integral domains. Notice that one problem is the appearance of absolute value in the bound on r: in general, rings do not have anything like absolute value, or a way to compare the "sizes" of two elements. However in Section 15 we did describe such a gadget for some rings: norms. Recall that a map $N : R \to \mathbb{N}$ is called a *norm* if (1) $N(x) = 0$ if and only if $x = 0$ and (2) $N(xy) \geq N(x)$ when $y \neq 0$. These properties do generalize the absolute value.

Def'n 19.1 (Euclidean Norm)**.** Let R be a domain and $N : R \to \mathbb{N}$ a norm.

(i) We say that N is *Euclidean* if for all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = qb + r$ and $0 \leq N(r) < N(b)$.

(ii) A domain which has a Euclidean norm is called a *Euclidean Domain*.

Of course $\mathbb{Z}$ is a Euclidean Domain with norm $N(a) = |a|$. The existence of a Euclidean norm on R is very powerful. For instance, many of the nice properties of $\mathbb{Z}$ which we derived from the Division Algorithm have analogues in any Euclidean Domain. More generally, the norm allows us to recover some of the benefits of mathematical induction.

Prop'n 19.2. Every Euclidean domain is a UFD.

Proof. (@@@) □

As a consequence of this result, every Euclidean domain is also a GCD domain; recall that we have a nice characterization of greatest common divisors in a UFD due to unique factorization. But this characterization of GCDs is *nonconstructive*; we know that given two elements of a UFD, they have a greatest common divisor, but actually computing a GCD requires that we be able to factor and recognize associates. In a Euclidean domain, we can do much better.

Prop'n 19.3 (Euclidean Algorithm)**.** Every Euclidean Domain is also a GCD Domain.

Proof. Let R be a Euclidean domain with norm N. We want to show that for all $a \in R$, for all $b \in R$, the set $\mathsf{gcd}(a, b)$ is not empty. We proceed by strong induction on $N(a)$.

Base case. If $N(a) = 0$, then $a = 0$, and so we have $b \in \mathsf{gcd}(a, b)$ for all b.

Inductive Step. Let $a \in R$ and suppose that the result holds for all a' with $1 \leq N(a') < N(a)$. In particular, note that $a \neq 0$. Now let $b \in R$. By the division algorithm we may decompose b as $b = qa + r$, where $0 \leq N(r) < N(a)$. If $r = 0$ then $a|b$ and we have $a \in \mathsf{gcd}(a, b)$. If $r \neq 0$, then by the inductive hypothesis $\varnothing \neq \mathsf{gcd}(r, a) = \mathsf{gcd}(b - qa, a) = \mathsf{gcd}(b, a)$ as needed. □

Again, the proof of this statement is logically redundant; we already knew that Euclidean domains are GCD domains. But algorithmically this proof gives us something very strong. If we have an effective procedure for computing quotients and remainders in R, then we have an effective procedure for computing GCDs.

Prop'n 19.4. Every field is a Euclidean domain.

Proof. Define a mapping $N : F \to \mathbb{N}$ by $N(x) = 0$ if $x = 0$ and 1 if $x \neq 0$. We can see that N is a Euclidean norm. □

* * Exercises * *

19.1. (@@@) (k-stage Euclidean)

19.2. (@@@) Dropping the domain condition on euclidean domains; see [**?**]

20 Localization and the Field of Fractions

In a general ring with 1, or even a general domain, elements typically do not have multiplicative inverses. Those which do are called *units* and are very special. In this section we will see how a domain can be "extended" to a larger ring so that any given element can be made into a unit. First we need a definition.

Def'n 20.1 (Multiplicative Subset)**.** Let R be a domain and $S \subseteq R$. We say that S is a *multiplicative subset* (or a set of *denominators*) of R if $0 \notin R$ and if S is closed under multiplication.

Domains have plenty of multiplicative sets. For instance, the set of all nonzero elements is multiplicative. If $a \in R$ is not zero, then the set $S = \{1, a, a^2, a^3, \ldots\}$ of powers of a is multiplicative. Here is the punch line of this section.

If $S \subseteq R$ is a multiplicative subset, then we can construct a new ring, T, which contains R as a subset, but in which the elements of S are units.

Prop'n 20.2. Let R be a domain and $S \subseteq R$ a multiplicative subset. We define a relation Φ on the cartesian product $S \times R$ as follows:

$$(s_1, r_1)\Phi(s_2, r_2) \quad \text{iff} \quad r_1 s_2 = r_2 s_1.$$

This relation Φ is an equivalence.

Proof. (i) We have $rs = rs$ for all $r \in R$ and $s \in S$, so that $(s, r)\Phi(s, r)$.

(ii) Suppose $(s_1, r_1)\Phi(s_2, r_2)$. Then $r_1 s_2 = r_2 s_1$, so that $r_2 s_1 = r_1 s_2$, and thus $(s_2, r_1)\Phi(s_1, r_1)$.

(iii) Suppose $(s_1, r_1)\Phi(s_2, r_2)$ and $(s_2, r_2)\Phi(s_3, r_3)$. Now $r_1 s_2 = r_2 s_1$ and $r_2 s_3 = r_3 s_2$. We then have $r_1 s_2 r_2 s_3 = r_2 s_1 r_3 s_2$; rearranging (since R is commutative) and using cancellation, we have $r_1 s_3 = r_3 s_1$. So $(s_1, r_1)\Phi(s_3, r_3)$ as needed.

□

Since Φ is an equivalence, it induces a partition on the set $S \times R$. We will denote this quotient set $S^{-1}R = (S \times R)/\Phi$ and denote the equivalence class of (s, r) by $\frac{r}{s}$.

Prop'n 20.3. Let R be a domain with $S \subseteq R$ a multiplicative subset.

Define operations $+$ and $\cdot$ on $S^{-1}R$ as follows.

$$\frac{r_1}{s_1} + \frac{r_2}{s_2} = \frac{r_1 s_2 + r_2 s_1}{s_1 s_2} \quad \text{and} \quad \frac{r_1}{s_1} \cdot \frac{r_2}{s_2} = \frac{r_1 r_2}{s_1 s_2}.$$

Then we have the following.

(i) $+$ and $\cdot$ are well-defined.

(ii) $S^{-1}R$, with these operations, is an integral domain, which we call the *localization* of R *at* S.

(iii) If $t \in S$, then the mapping $\iota : R \to S^{-1}R$ given by $\iota(r) = \frac{rt}{t}$ is an injective ring homomorphism, and $\iota(t)$ is a unit in $S^{-1}R$.

Proof. (super tedious) □

So $S^{-1}R$ is a new ring which contains a "copy" (homomorphic image) of R, within which the elements of S become units.

Def'n 20.4. Let R be a domain and let $D = \{x \in R \mid x \neq 0\}$ be the multiplicative subset of all nonzero elements of R. Then the localization $D^{-1}R$ is a field, called the *field of fractions* of R.

For example, $\mathbb{Q}$ is properly defined as the field of fractions of $\mathbb{Z}$.

* * Exercises * *

20.1. Let R be a domain and $S, T \subseteq R$ multiplicative sets. Show that $ST = \{st \mid s \in S, t \in T\}$ is multiplicative.

20.2. Let R be a GCD domain, with F its field of fractions. An element $\frac{a}{b} \in F$ is said to be *reduced* (or in *lowest terms*) if $\gcd(a, b) = 1$.

(i) Show that every element of F has a reduced representative.

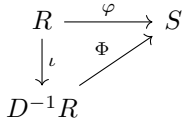
(ii) Show that reduced fractions are unique in the following sense: If $\frac{a}{b} = \frac{c}{d}$ are both reduced, then $c = au$ and $d = bu$ for some unit u.

20.3. (@@@ do stuff in $\mathbb{Z}[\frac{1}{2}]$)

20.4. If R is a UFD and $S \subseteq R$ any multiplicative set, then $S^{-1}R$ is also a UFD.

20.5. If R is a Euclidean domain and $S \subseteq R$ any multiplicative set, then $S^{-1}R$ is a Euclidean domain.

20.6. **Universal Property of Localization** Let R be a ring and let $D \subseteq R$ be a multiplicative subset. Suppose $\varphi : R \to S$ is a ring homomorphism (with S unital) such that $\varphi(d)$ is a unit in S for all $d \in D$. Show that there is a unique ring homomorphism $\Phi : D^{-1}R \to S$ such that the following diagram commutes.



That is, a unique Φ such that $\varphi = \Phi \circ \iota$.

21 Quadratic Numbers

Prop'n 21.1. $\mathbb{Z}[i]$ is a Euclidean domain under the norm $N(a+bi) = a^2+b^2$.

Proof. Let $\alpha = a_1 + a_2 i$ and $\beta = b_1 + b_2 i$ be Gaussian integers, with $\beta \neq 0$. Thinking of α and β as elements of $\mathbb{Q}(i)$, we have

$$\frac{\alpha}{\beta} = t_1 + t_2 i = \frac{a_1 b_1 + a_2 b_2}{b_1^2 + b_2^2} + \frac{a_2 b_1 - a_1 b_2}{b_1^2 + b_2^2} i.$$

Choose integers q_1 and q_2 such that $|q_1 - t_1| \leq \frac{1}{2}$ and $|q_2 - t_2| \leq \frac{1}{2}$. (Note that this is always possible.) Let $\gamma = q_1 + q_2 i$, and let $\delta = \alpha - \gamma\beta$. Note that by construction, γ and δ are in $\mathbb{Z}[i]$.

We now have

$$\begin{aligned} N(\delta) &= N(\alpha - \gamma\beta) = N\left((\frac{\alpha}{\beta} - \gamma)\beta\right) = N(\frac{\alpha}{\beta} - \gamma)N(\beta) \\ &= ((q_1 - t_1)^2 + (q_2 - t_2)^2)N(\beta) \leq \frac{1}{2}N(\beta) < N(\beta), \end{aligned}$$

as needed. □

Cor. 21.2. $\mathbb{Z}[i]$ is a GCD domain and a UFD.

Here is a worked example of the division algorithm in the Gaussian integers. Let $\alpha = 10 + 7i$ and $\beta = 3 + 2i$. Now

$$\frac{\alpha}{\beta} = \frac{44}{13} + \frac{1}{13} i = (3 + \frac{5}{13}) + (0 + \frac{1}{13})i.$$

Let $t_1 = 3$ and $t_2 = 0$, so that $\gamma = 3$. Now $\delta = \alpha - \gamma\beta = 1 + i$. We then have $10 + 7i = 3(3 + 2i) + (1 + i)$ and $N(1 + i) < N(3 + 2i)$.

* * EXERCISES * *

21.1. **A Theorem of Fermat** 5 and 3 have the curious property that $3^3 = 5^2 + 2$. Fermat asked whether or not there are any other pairs like this: a perfect square and a perfect cube separated by 2. (@@@)

(i) Show that $x \equiv y \equiv 1 \pmod 2$.

(ii) Show that $x - \sqrt{-2}$ and $x + \sqrt{-2}$ are relatively prime in (@@@).

(iii) Since blah is a UFD let $x + \sqrt{-2} = (m + n\sqrt{-2})$, show that $x = \pm 5$.

21.2. (@@@) (Factorization in $\mathbb{Z}[i]$)

mordell equations

Summary of Chapter II

— III —

Polynomial Rings

22 Polynomials

We've been working with polynomials since taking algebra in middle school. But what is a polynomial, exactly? In this section, we will extend some of our ideas about rings to sets of polynomials. First, though, we need to have a better idea of what makes a polynomial a polynomial. It is easy enough to come up with some examples of polynomials as we'd see them in College Algebra.

$$x^2 + x - 1 \qquad x^3 - 1 \qquad 7 \qquad \frac{1}{2}x^7 - \frac{2}{3}x^2 + 1 \qquad \pi x^3 + ex + \sqrt{2}$$

Just as important, we can come up with examples of things that sort of look like polynomials but aren't.

$$x^{1/2} + 2x \qquad x^{-1} + x^{\sqrt{2}} \qquad x^2 + x = 7x^3 \qquad 1 + x + x^2 + \cdots$$

From here, let's try to extract a definition that includes all the examples we want but none of the ones we don't. A polynomial in the **variable** x is an **expression** (not an equation) which can be written as a **finite sum** of things of the form cx^k, where c is **some kind of number** and k is a **natural number exponent**. The cs are called the coefficients of the polynomial.

We can add polynomials by "combining like terms", such as

$$\begin{aligned}(x^2 + 2x + 1) + (3x^2 - 4x + 27) &= (1+3)x^2 + (2-4)x + (1+27)\\ &= 4x^2 - 2x + 28.\end{aligned}$$

And if a particular polynomial is "missing" a term, we can pretend it is there with coefficient zero.

$$\begin{aligned}(x^2 + 1) + (x + 1) &= (x^2 + 0x + 1) + (0x^2 + x + 1)\\ &= (1+0)x^2 + (0+1)x + (1+1)\\ &= x^2 + x + 2\end{aligned}$$

We can even multiply polynomials by using the "distributive property" over and over again.

$$\begin{array}{rrrrr}
 & & +2x^2 & +3x & +1 \\
 & \times & +1x^2 & -2x & +2 \\
\hline
 & & +4x^2 & +6x & +2 \\
 & +3x^3 & -6x^2 & -2x & \\
2x^4 & +3x^3 & +1x^2 & & \\
\hline
2x^4 & +6x^3 & -1x^2 & +4x & +2
\end{array}$$

As lazy mathematicians we might start to suspect that the "variable", x, is not so special, and really just serves as a placeholder to keep the coefficients separate. We may even come to think of a polynomial as just a list of coefficients, only using the variables to keep track of what position each coefficient

takes in the list. But then a list is really a mapping from the natural numbers, say $f : \mathbb{N} \to \mathbb{Q}$ (if the coefficients are rational numbers), where $f(i)$ is the coefficient of x^i, and by convention the $f(i)$ are all zero after some point. Now the arithmetic of polynomials corresponds to a funny arithmetic on functions $\mathbb{N} \to \mathbb{Q}$. Note that to make the arithmetic on *polynomials* work, we just need to have an arithmetic on *coefficients* – which is provided by a ring.

Def'n 22.1 (Polynomial)**.** Let R be a ring. A mapping $a : \mathbb{N} \to R$ is called a *polynomial* with *coefficients in* R if there is a natural number M such that $a_i = 0$ whenever $i > M$.

The polynomial X given by $X(k) = 1$ if $k = 1$ and 0 otherwise is called the *indeterminate*. The ring of polynomials over R with indeterminate X is denoted $R[X]$.

That is, a polynomial is an infinite list of elements of R which is "eventually" zero. This may seem like a strange way to think about polynomials at first, but it's not really so different from the way we define matrices. Like matrices, what makes polynomials interesting is that they inherit a natural arithmetic from R.

Prop'n 22.2. Let R be a ring and x an indeterminate. We define operations $+$ and $\cdot$ on $R[x]$ as follows: if $a, b \in R[x]$, then

$$\begin{aligned} (a+b)(k) &= a(k) + b(k) \\ (a \cdot b)(k) &= \sum_{i+j=k} a(i)b(j) \end{aligned}$$

where the arithmetic on the right hand sides takes place in R.

(i) These operations make $R[x]$ into a ring.

(ii) $R[x]$ is commutative if and only if R is commutative.

(iii) $R[x]$ is unital if and only if R is unital. In this case $1_{R[x]}$ is the polynomial whose 0th coefficient is 1_R and whose every other coefficient is 0_R.

That is, the kth coefficient of a sum is the sum of kth coefficients, and the kth coefficient of a product is a linear combination of the coefficients of the factors. To be clear: This is the usual polynomial arithmetic we know and love, but with coefficients coming from any fixed ring rather than from a ring of numbers.

Def'n 22.3 (Degree and Leading Coefficient)**.** Let R be a ring and $a \in R[x]$ a nonzero polynomial. By definition there is some natural number M such that $a_i = 0$ when $i > M$; the *smallest* such M is called the *degree*

of a and is denoted $\deg a$. Now $a_{\deg a}$, which is nonzero if $a \neq 0$, is called the *leading coefficient* of a.

If R is unital and the leading coefficient of a is 1_R, we say that $a(x)$ is *monic*. The degree of the zero polynomial is left undefined.

We can think of the degree of a polynomial as a kind of size. We can use the degree to decompose polynomials into "sum of powers" form; a polynomial p of degree d can be written as

$$p = \sum_{i=0}^{d} p_i x^i$$

(see Exercise 22.2). As an example let $R = \mathbb{Z}/(6)$ and consider the two polynomials

$$p = 1 + 2x^2 \quad \text{and} \quad q = 2 + 3x.$$

Now $p(0) = 1$, $p(2) = 2$, and $p(i) = 0$ otherwise, and $q(0) = 2$, $q(1) = 3$, and $q(i) = 0$ otherwise. So we have

$$\begin{aligned} p + q &= (p_0 + q_0) + (p_1 + q_1)x + (p_2 + q_2)x^2 \\ &= (1 + 2) + (0 + 3)x + (2 + 0)x^2 \\ &= 3 + 3x + 2x^2 \end{aligned}$$

and

$$\begin{aligned} pq &= p_0q_0 + (p_0q_1 + p_1q_0)x + (p_0q_2 + p_1q_1 + p_2q_0)x^2 \\ &\quad + (p_0q_3 + p_1q_2 + p_2q_1 + p_3q_0)x^3 \\ &= (1 \cdot 2) + (1 \cdot 3 + 0 \cdot 2)x + (1 \cdot 0 + 0 \cdot 3 + 2 \cdot 2)x^2 \\ &\quad + (1 \cdot 0 + 0 \cdot 0 + 2 \cdot 3 + 0 \cdot 2)x^3 \\ &= 2 + 3x + 4x^2 \end{aligned}$$

So far this business about polynomials works over any ring. But over a domain, we can say a little more.

Prop'n 22.4. Let R be a domain. Then we have the following.

(i) If a and b are nonzero, then ab is nonzero, and moreover $\deg ab = \deg a + \deg b$ for all nonzero $a, b \in R$.

(ii) $u \in R[x]$ is a unit if and only if $\deg u = 0$ and u_0 is a unit in R.

Proof. (i) Let $a, b \in R[x]$ be nonzero; say $m = \deg a$ and $n = \deg b$. Note that

$$(ab)_{m+n} = \sum_{i+j=m+n} a_i b_j.$$

Now if $i > m$, then $a(i) = 0$, and if $i < m$ then $j > n$ and $b(j) = 0$. So the only possible nonzero term is $a_m b_n$. Moreover, a_m and b_n are

not zero (being the leading terms of a and b) and since R is a domain, $a_m b_n \neq 0$. Thus $(ab)_{m+n} = a_m b_n \neq 0$, and so $ab \neq 0$. Now note that $\deg ab \geq m+n$. To see that this is in fact an equality, let $k > m+n$. Now

$$(ab)_k = \sum_{i+j=k} a(i)b(j).$$

As before, if $i > m$ then $a(i) = 0$, and if $i \leq m$, then $j > n$ and we have $b(j) = 0$. So $(ab)_k = 0$ if $k > m+n$, and thus $\deg ab = m+n = \deg a + \deg b$.

(ii) Let $u \in R[x]$ be a unit. Now $1 = uu^{-1}$, and we have

$$0 = \deg 1 = \deg uu^{-1} = \deg u + \deg u^{-1}.$$

Since the degree of a polynomial is a natural number, we must have $\deg u = \deg u^{-1} = 0$.

□

Cor. 22.5. Let R be a domain. Then the map $N : R[x] \to \mathbb{N}$ given by $N(a) = 2^{\deg a}$ if $a \neq 0$ and $N(0) = 0$ is a multiplicative norm.

In the rest of this chapter we will be concerned mostly with $R[x]$ when R is a domain. In this situation we will consider two basic questions:

1. How is the structure of R reflected in the structure of $R[x]$? (Quite a bit, it turns out.)
2. Given a polynomial $p \in R[x]$, can we detect whether or not p is irreducible? (Sometimes.)

* * Exercises * *

22.1.

(i) In $(\mathbb{Z}/(3))[x]$, let $p(x) = [1] + [2]x$ and $q(x) = [2] + x$. Find $p+q$ and pq.

(ii) In $(\mathbb{Z}/(6))[x]$, let $p(x) = [1] + [2]x$ and $q(x) = [1] + x + [3]x^2$. Compute pq.

(iii) In $\mathsf{Mat}_2(\mathbb{Z})[x]$, let

$$p(x) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} x.$$

Find p^2.

22.2. **The expansion of a polynomial.** Let R be a ring with $c \in R$, p a polynomial over R of degree d, and x an indeterminate.

(i) $(\overline{c}p)(k) = cp(k)$.

(ii) $(xp)(0) = 0_R$ and $(xp)(k+1) = p(k)$.

(iii) $x^t(k) = 1_R$ if $k = t$ and 0_R otherwise.

(iv) Conclude that

$$p = \sum_{i=0}^{d} p(i)x^i.$$

22.3. **A polynomial map.** Suppose $\varphi : R \to S$ is a ring homomorphism. Show that the mapping $\Phi : R[x] \to S[x]$ given by $\Phi(a)(k) = \varphi(a(k))$ for all $k \in \mathbb{N}$ is the unique ring homomorphism which makes the following diagram commute.

$$\begin{array}{ccc} R & \xrightarrow{\varphi} & S \\ {\scriptstyle\iota}\downarrow & & \downarrow{\scriptstyle\iota} \\ R[x] & \xrightarrow[\Phi]{} & S[x] \end{array}$$

Show that if φ is injective, then Φ is injective, and that if φ is surjective, then Φ is surjective.

Def'n 22.6 (Derivative)**.** Let R be a ring. We define a mapping $\mathsf{D} : R[x] \to R[x]$, called the *derivative*, as follows. Given $a(x) \in R[x]$,

$$\mathsf{D}(a)(k) = (k+1)a(k+1)$$

for all $k \in \mathbb{N}$. For instance, if $a(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, then

$$\mathsf{D}(a(x)) = a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^{n-1}.$$

Note that the derivative of a polynomial is defined here purely formally; there are no limits involved.

22.4. Show that if R is a ring and $a, b \in R[x]$, then $\mathsf{D}(a+b) = \mathsf{D}(a) + \mathsf{D}(b)$.

22.5. **Leibniz Rule.** Show that if R is a commutative (@@@) ring and $a, b \in R[x]$, then $\mathsf{D}(ab) = a\mathsf{D}(b) + \mathsf{D}(b)a$.

23 Long division and roots

As we will see, there are some fundamental similarities between polynomials and integers. The first important result in this direction is the division algorithm for polynomials.

Prop'n 23.1. Let R be a commutative unital ring, and let $a(x), b(x) \in R[x]$ be polynomials such that $b(x) \neq 0$ and the leading coefficient of b is a unit in R. Then there exist polynomials $q(x), r(x) \in R[x]$ such that $a(x) = q(x)b(x) + r(x)$ and either $r(x) = 0$ or $\deg r < \deg b$.

Proof. If $a(x) = 0$, set $q(x) = r(x) = 0$. Suppose now that $a(x) \neq 0$; we proceed by strong induction on $\deg a$.

- **Base case.** If $\deg a = 0$, then $a(x) = a_0$ is a constant. If $\deg b = 0$, then $b(x) = b_0$ is also a constant, and in this case b_0 is the leading coefficient of b, hence a unit. Let $q(x) = a_0 b_0^{-1}$ and $r(x) = 0$. If $\deg b > 0$, let $q(x) = 0$ and $r(x) = a_0$. Then $a(x) = q(x)b(x) + r(x)$ and we have $\deg b \geq 1 > 0 = \deg r$.
- **Inductive Step.** Suppose the result holds for all polynomials $\overline{a}(x)$ of degree strictly less than n, where $n > 0$, and suppose that $a(x)$ has degree n. If $\deg a < \deg b$, let $q(x) = 0$ and $r(x) = a(x)$. Now suppose instead that $\deg a \geq \deg b$. Let $m = \deg b$ and let a_n be the leading coefficient of $a(x)$ and b_m the leading coefficient of $b(x)$ (which is a unit). Define $\overline{a}(x) = a(x) - a_n b_m^{-1} x^{n-m} b(x)$. Note that $\deg \overline{a} < \deg a$. By the inductive hypothesis, we have $\overline{q}(x), r(x) \in R[x]$ such that $\overline{a}(x) = \overline{q}(x)b(x) + r(x)$ and either $r(x) = 0$ or $\deg r < \deg b$. Define $q(x) = \overline{q}(x) + a_n b_m^{-1} x^{n-m}$. Now

$$\begin{aligned} a(x) - q(x)b(x) &= a(x) - \overline{q}(x)b(x) - a_n b_m^{-1} x^{n-m} b(x) \\ &= \overline{a}(x) - \overline{q}(x)b(x) \\ &= r(x) \end{aligned}$$

  as needed.

By induction, the result holds for all n. □

Note that this result is very general; the coefficient ring R is only required to be commutative and unital. In fact, with slight adjustments to the punchline, even the C and U conditions on R can be dropped (see the exercises). For us, for now, the most important situation will be when R is a field, because in this case $R[x]$ is a Euclidean domain.

Cor. 23.2. Suppose F is a field.

(i) $F[x]$ is a Euclidean domain with norm $N(a) = 2^{\deg a}$. In particular,

$F[x]$ is also a UFD and a GCD domain.

(ii) $p(x) \in F[x]$ is irreducible iff $p(x)$ cannot be factored as a product of nonconstants.

(iii) If $p(x)$ has degree 1, then $p(x)$ is irreducible in $R[x]$.

In particular, if R is a domain, then $R[x]$ is a subring of a Euclidean domain: namely, $F[x]$, where F is the field of fractions of R. This means that $R[x]$ immediately has a much stronger structure than the average integral domain. For example: carrying out long division over F and then clearing denominators gives the following result.

Cor. 23.3. If R is a domain and $a, b \in R[x]$ with $b(x) \neq 0$, then there exist $q, r \in R[x]$ and $k \in R$ such that $ka(x) = q(x)b(x) + r(x)$ and either $r = 0$ or $r \neq 0$ and $\deg r < \deg b$.

So far, we've been thinking of polynomials as objects in their own right. But we can also treat them like functions in the usual sense by "plugging in" ring elements for the variable.

Def'n 23.4. Given a polynomial $p(x) = \sum_{i=0}^{n} a_n x^n$ in $R[x]$, R a commutative unital ring, we define the *evaluation map* $\varepsilon_p : R \to R$ by $\varepsilon_p(r) = \sum_{i=0}^{n} a_i r^i$. We say that an element $r \in R$ is a *root* of $p(x)$ if $\varepsilon_p(r) = 0_R$.

There is an important distinction between the polynomial p and its corresponding evaluation map ε_p, although in practice this distinction is easy to blur.

Prop'n 23.5. Let R be a commutative unital ring. Then $\varepsilon_{p+q}(r) = \varepsilon_p(r) + \varepsilon_q(r)$ and $\varepsilon_{pq}(r) = \varepsilon_p(r)\varepsilon_q(r)$.

Prop'n 23.6 (Factor Theorem)**.** Let R be a commutative unital ring, with $p(x) \in R[x]$ and $a \in R$. Then a is a root of $p(x)$ if and only if $x - a$ divides $p(x)$ in $R[x]$.

Proof. Certainly if $x-a$ divides $p(x)$ then a is a root of p. Conversely, suppose a is a root of $p(x)$. Now $b(x) = x - a$ is monic, so by the polynomial long division algorithm we have $q(x), r(x) \in R[x]$ such that $p(x) = q(x)(x - a) + r(x)$ and either $r(x) = 0$ or $\deg r < 1$. If $r(x) \neq 0$, then $r(x) = r_0$ is a constant. Evaluating at a we have $p(a) = r_0$, a contradiction. So $r(x) = 0$ and $x - a$ divides $p(x)$. □

Cor. 23.7. Let R be a domain with $p(x) \in R[x]$ a nonzero polynomial of degree d. Then $p(x)$ has at most d roots in R, counting multiplicity.

Cor. 23.8. Let R be a domain with $p(x), q(x) \in R[x]$ polynomials of degree at most d. If there exist $d+1$ distinct elements $a_i \in R$ such that $p(a_i) = q(a_i)$, then $p(x) = q(x)$.

Proof. Let $s(x) = p(x) - q(x)$; note that s has degree at most d. Now the a_i are $d+1$ distinct roots of s. If $s(x) \neq 0$, this contradicts the previous corollary. □

Prop'n 23.9. Let R be a domain and $p(x) \in R[x]$ a polynomial of degree 2 or 3. Then $p(x)$ cannot be written as a product of nonconstants in $R[x]$ if and only if $p(x)$ does not have a root in R.

Proof. Note that if $p(x) = a(x)b(x)$, then $\deg a + \deg b$ is either 2 or 3. Thus $p(x)$ is a product of nonconstants iff it has a factor of degree 1. But $p(x)$ has a factor of degree 1 iff it has a root in R. □

Examples

1. The degree 2 polynomial $p(x) = x^2 + 1$ is irreducible over $\mathbb{Z}/(3)$ since it has no roots.

Prop'n 23.10. Let R be a domain and $q(x) = \sum_{i=0}^{n} \in R[x]$. Suppose $p \in R$ is prime such that $p \not| a_n$, $p | a_i$ for each $0 \leq i < n$, and $p^2 \not| a_0$. Then $q(x)$ cannot be factored as a product of nonconstants.

Proof. Suppose we have

$$q(x) = b(x)c(x) = \left(\sum_i b_i x^i\right)\left(\sum_j c_j x^j\right) = \sum_k \left(\sum_{i+j=k} b_i c_j\right) x^k.$$

Note that $q_0 = b_0 c_0$. Since $p | b_0 c_0$ and $p^2 \not| b_0 c_0$, p divides exactly one of b_0 and c_0; suppose WLOG that $p | b_0$, so $p \not| c_0$. Letting $n = \deg q$, $h = \deg b$, and $k = \deg c$, we have $q_n = b_h c_k$, and since $p \not| q_n$, $p \not| b_h$. Let i be minimal such that $p \not| b_i$. (Note that $0 < i \leq \deg b \leq n$.)

We now have

$$q_i = b_i c_0 + b_{i-1} c_1 + \cdots + b_{i-t} c_t$$

for some t. If $i < n$, then $p | q_i$, and by construction, $p | b_j$ for $j < i$. Thus $p | b_i c_0$, and since $p \not| c_0$ we have $p | b_i$ – a contradiction. So $i = n$ and thus $\deg b = n = \deg q$. But $\deg q = \deg b + \deg c$, so that $\deg c = 0$; hence c is a constant. □

* * Exercises * *

23.1. **Hensel's Lemma** (@@@)

24 Content of a polynomial

One of our big questions is to what extent the structure of R is reflected in the structure of $R[x]$; if R has more "technology" available, perhaps this can be used to say interesting things about the polynomials over R. In fact, thanks to the polynomial long division algorithm, if R is a domain then $R[x]$ is already sitting inside a Euclidean domain – namely $F[x]$ where F is the field of fractions of R. So it doesn't take much to get extra technology in $R[x]$.

Def'n 24.1 (Content of a polynomial)**.** Let R be a GCD domain and let $p(x) \in R[x]$ be a polynomial with coefficients a_i. We define the *content* of $p(x)$ to be

$$\mathsf{content}(p) = \begin{cases} 0 & \text{if } p(x) = 0 \\ \mathsf{gcd}\,(a_0, a_1, \ldots, a_d) & \text{if } p(x) \neq 0, \text{where } d = \deg p. \end{cases}$$

If $\mathsf{content}(p) = 1$, we say that $p(x)$ is *primitive*.

Since the content of a polynomial is defined as a greatest common divisor, it may be difficult to actually compute $\mathsf{content}(p)$. But if our ring of coefficients is a Euclidean domain with an effective division algorithm, computing $\mathsf{content}(p)$ is not too bad with the Euclidean algorithm. Our primary example is the GCD domain $\mathbb{Z}$. For example, over this ring we have

$$\mathsf{content}(2x^3 + 4x - 6) = \mathsf{gcd}\,(2, 4, -6) = 2.$$

Every monic polynomial is primitive, and (less interestingly) every nonzero polynomial over a field is primitive.

Prop'n 24.2. Let R be a GCD domain.

(i) Every polynomial $a(x) \in R[x]$ can be written as $a(x) = \mathsf{content}(a)\overline{a}(x)$, where $\overline{a}(x) \in R[x]$ is primitive.

(ii) Let $d \in R$ and $a(x) \in R[x]$. Then the constant polynomial d divides $a(x)$ in $R[x]$ if and only if d divides $\mathsf{content}(a)$ in R.

(iii) Let $d \in R$ and $a(x), b(x) \in R[x]$. If $d|\mathsf{content}(a+b)$ and $d|\mathsf{content}(a)$, then $d|\mathsf{content}(b)$.

(iv) If $d \in R$ and $a(x) \in R[x]$, then $\mathsf{content}(da) = d\mathsf{content}(a)$.

(v) Let F be the field of fractions of R and let $q(x) \in F[x]$. Then there is a fraction $\frac{u}{v} \in F$ such that $p(x) = \frac{u}{v}q(x)$ is in $R[x]$ and is primitive there.

(vi) $\mathsf{content}(x^n a(x)) = \mathsf{content}(a(x))$.

Proof. (i) If $a(x) = 0$, set $\overline{a}(x) = 1$. Suppose $a(x) \neq 0$. Now $\mathsf{content}(a) = \gcd(a_0, a_1, \ldots, a_n)$, and in particular for each i we have $a_i = \mathsf{content}(a)\overline{a}_i$ for some $\overline{a}_i$, and $\gcd(\overline{a}_0, \overline{a}_1, \ldots, \overline{a}_n) = 1$. Let $\overline{a}(x) = \sum_{i=0}^{n} \overline{a}_i x^i$.

(ii) (write these)

□

Prop'n 24.3 (Gauss' Lemma – Part I)**.** Let R be a GCD Domain with $a(x), b(x) \in R[x]$. If $a(x)$ and $b(x)$ are primitive, then $a(x)b(x)$ is primitive.

Proof. We proceed by induction on the number k of nonzero terms of a and b together.

Base Case ($k = 0$): If a and b together have no nonzero terms, then $a(x) = b(x) = 0$; neither is primitive.

Base Case ($k = 1$): If a and b together have exactly one nonzero term, then either $a(x) = 0$ or $b(x) = 0$; one is not primitive.

Base Case ($k = 2$): If $a(x)$ and $b(x)$ together have exactly two nonzero terms, then each must have exactly one. (Otherwise one is zero and thus not primitive.) Say $a(x) = a_n x^n$ and $b(x) = b_m x^m$. If both $a(x)$ and $b(x)$ are primitive, then $a_n = \mathsf{content}(a)$ and $b_m = \mathsf{content}(b)$ are units, so that $\mathsf{content}(ab) = a_n b_m$ is a unit; hence $a(x)b(x)$ is primitive.

Inductive Step: Suppose the result holds for all pairs of primitive polynomials having less than $n > 2$ nonzero terms together, and suppose that $a(x)$ and $b(x)$ are primitive with exactly n nonzero terms together. Say $\deg a = n$ and $\deg b = m$, so that the leading coefficients of a, b, and ab are a_n, b_m, and $a_n b_m$, respectively. Now let $c = \mathsf{content}(ab)$, and suppose BWOC that c is *not* a unit. Note that $c | a_n b_m$. Now $\gcd(c, a_n)$ and $\gcd(c, b_m)$ cannot both be units in R. (If $\gcd(c, a_n) = 1$, then by Euclid's lemma we have $c | \gcd(c, b_m)$.) So suppose WLOG that $\gcd(c, a_n) = d$ is not a unit.

Now $d | \mathsf{content}(ab)$ in R, so that $d | a(x)b(x)$ in $R[x]$. Since $d | a_n$, we also have $d | a_n x^n$ in $R[x]$. Thus $d | b(x)(a(x) - a_n x^n)$ in $R[x]$, and thus

$$d | \mathsf{content}(b(x)(a(x) - a_n x^n)) = \mathsf{content}(a(x) - a_n x^n)\mathsf{content}(b(x)p(x)),$$

where $p(x) \in R[x]$ is primitive such that $a(x) - a_n x^n = \mathsf{content}(a(x) - a_n x^n)p(x)$. In particular, note that $p(x)$ and $a(x) - a_n x^n$ have the same number of nonzero terms which is one fewer than the number of nonzero terms of $a(x)$. Thus b and p have fewer than n nonzero terms. Since b and p are both primitive, by the inductive hypothesis, $\mathsf{content}(bp) = 1$. Thus we have $d | \mathsf{content}(a(x) - a_n x^n)$. Since $d | \mathsf{content}(a_n x^n)$, by the lemma we have $d | \mathsf{content}(a)$. But a is primitive, so that d is a unit, a contradiction. So $a(x)b(x)$ must be primitive. □

Cor. 24.4. With a, b, and R as in 24.3 we have the following.

(i) $\mathsf{content}(ab) = \mathsf{content}(a)\mathsf{content}(b)$.

(ii) The factors of a primitive polynomial are primitive.

(iii) If $a(x)|b(x)$ in $R[x]$, then $\mathsf{content}(a)|\mathsf{content}(b)$ in R.

* * EXERCISES * *

24.1. Let R be a GCD domain. Show that if $p(x) = a(x)b(x)$ and $\mathsf{content}(p)$ is irreducible, then either $a(x)$ or $b(x)$ must be primitive.

25 Over a GCD Domain

Lemma 25.1. Let R be a GCD domain with field of fractions F.

1. If $p(x) \in R[x]$ is primitive, $r \in R$, and $a(x) \in R$ such that $p(x)|a(x)$ and $r|a(x)$ in $R[x]$, then $rp(x)|a(x)$ in $R[x]$.
2. If $q(x) \in F[x]$ and $p(x) \in R[x]$ such that $p(x)$ is primitive and $p(x)q(x) \in R[x]$, then in fact $q(x) \in R[x]$.

Proof.

1. Write $a(x) = p(x)b(x)$ with $b(x) \in R[x]$. Since $r|a(x)$, we have $r|\mathsf{content}(a) = \mathsf{content}(p)\mathsf{content}(b) = \mathsf{content}(b)$, since p is primitive. So $r|b(x)$ in $R[x]$. Say $b(x) = rc(x)$; then $a(x) = rp(x)c(x)$ as needed.
2. We have $\frac{u}{v} \in F$ (in lowest terms) such that $\frac{u}{v}q(x) \in R[x]$ is primitive; say $\frac{u}{v}q(x) = s(x)$. Now $uq(x) = vs(x)$, and moreover $up(x)q(x) = vp(x)s(x) \in R[x]$. Now

$$\begin{aligned} u \cdot \mathsf{content}(pq) &= \mathsf{content}(up(x)q(x)) \\ &= \mathsf{content}(vp(x)s(x)) \\ &= v \cdot \mathsf{content}(ps) = v, \end{aligned}$$

since p and s are primitive in $R[x]$. In particular, $u|v$. Since $\frac{u}{v}$ is in lowest terms, without loss of generality, $u = 1$, so that $\frac{1}{v}q(x) = s(x)$. Thus $q(x) = vs(x) \in R[x]$ as needed.

□

Prop'n 25.2 (Gilmer-Parker)**.** If R is a GCD Domain, then $R[x]$ is a GCD Domain.

Proof. Let $a(x), b(x) \in R[x]$. Let $k = \mathsf{gcd}(\mathsf{content}(a), \mathsf{content}(b))$ (remember that R is a GCD domain). Let F be the field of fractions of R. Now $F[x]$ is a Euclidean domain, in particular a GCD domain, so that $a(x)$ and $b(x)$ have a greatest common divisor in $F[x]$. By the lemma, we can take an associate (in $F[x]$) of this gcd which is in $R[x]$ and primitive; say $t(x)$. We claim that $kt(x)$ is a gcd of a and b in $R[x]$.

First note that $k|\mathsf{content}(a)$, so that $k|ax$. Now $t(x)|a(x)$ in $F[x]$, where t and a are in $R[x]$ and $t(x)$ is primitive. By the lemma, $t(x)|a(x)$ in $R[x]$, and again using the lemma, $kt(x)|a(x)$ in $R[x]$. Similarly, $kt(x)|b(x)$ in $R[x]$. So $kt(x)$ is a common divisor of $a(x)$ and $b(x)$ in $R[x]$.

Now suppose that $e(x) \in R[x]$ is a common divisor of $a(x)$ and $b(x)$ over R. If $e(x)$ is constant, then $e(x) = e_0|\mathsf{gcd}(\mathsf{content}(a), \mathsf{content}(b), =)k$. Suppose instead that $e(x)$ has positive degree. Now $e(x)$ divides $a(x)$ and $b(x)$ in $F[x]$, which is a GCD domain, and thus $e(x)$ divides $t(x)$ in $F[x]$. Say $e(x)f(x) =$

$t(x)$ where $f(x) \in F[x]$. By the lemma, we may write $f(x) = \frac{u}{v}g(x)$ where $g(x) \in R[x]$ is primitive and $\mathsf{gcd}(u, v) = 1$. We have $ue(x)g(x) = vf(x) \in R[x]$. Now $\mathsf{content}(ue(x)g(x)) = \mathsf{content}(vt(x))$, and since g and t are primitive over R, $u\mathsf{content}(e) = v$. By Euclid's lemma, $v|\mathsf{content}(e)$, so that $v|\mathsf{content}(a)$ and $v|\mathsf{content}(b)$, and thus $v|k$. In particular, we have $kf(x) = k\frac{u}{v}g(x) \in R[x]$, and thus $e(x) \cdot kf(x) = kt(x)$, so that $e(x)|kt(x)$ in $R[x]$.

Thus $kt(x)$ is a greatest common divisor of $a(x)$ and $b(x)$ in $R[x]$. □

* * Exercises * *

25.1. Let R be a GCD domain with $p(x), q(x) \in R[x]$ so that q is irreducible (hence prime), and let k be a natural number. Show that q^{k+1} divides p in $R[x]$ if and only if $q|p$ and $q^k|p'$ in $R[x]$. In particular, show that p is squarefree if and only if $\mathsf{gcd}(p, p') = 1$.

25.2. Let R be a GCD domain, with $p, q \in R[x]$ nonzero. Show that p and q have a common factor of positive degree in $R[x]$ if and only if there exist $a, b \in R[x]$, not zero, such that $\deg a < \deg q$, $\deg b < \deg p$, and $pa - qb = 0$. (@@@ Looking forward to univariate resultant.)

26 The Rational Root Theorem

In this section we establish some important results about irreducibility and factorization for polynomials over a GCD domain.

Prop'n 26.1 (Gauss' Lemma – Part II)**.** Let R be a GCD domain with field of fractions F, and let $p(x) \in R[x]$ have positive degree. Then $p(x)$ is irreducible in $R[x]$ if and only if $p(x)$ is irreducible in $F[x]$ and primitive in $R[x]$.

Proof. (type this) □

Combined with Eisenstein's criterion, Gauss' lemma provides an easy-to-apply irreducibility criterion.

Cor. 26.2. If $p(x) \in R[x]$ (R a GCD domain) is Eisenstein and primitive, then $p(x)$ is irreducible in $R[x]$.

Proof. Suppose $p(x) = a(x)b(x)$ with $a, b \in R[x]$. Since p is Eisenstein, WLOG $a(x)$ is a constant; say $a(x) = a_0$. Now $a_0|p$ in $R[x]$, so that $a_0|\mathsf{content}(p)$ in R. Since $p(x)$ is primitive, a is a unit in R, hence a unit in $R[x]$. So $p(x)$ is irreducible in $R[x]$. □

This criterion can be used to quickly verify that a given polynomial is irreducible – when it applies. Unfortunately there are plenty of irreducible polynomials to which this criterion does not apply. For example, $p(x) = x^2 + 1$ is primitive in $\mathbb{Z}[x]$, and in fact is irreducible. But it is not Eisenstein at any prime.

Prop'n 26.3 (Rational Root Theorem)**.** Let R be a GCD domain with fraction field F. Suppose $p(x) \in R[x]$. Let $\frac{u}{v} \in F$ be a fraction in lowest terms; that is, $\mathsf{gcd}(u, v) = 1$ in R. If $\frac{u}{v}$ is a root of $p(x)$, then u divides the constant coefficient of p, and v divides the leading coefficient of p.

The Rational Root Theorem allows us to restrict the possible "rational roots" (that is, those in F, or equivalently factors over R of the form $ax - b$) to a finite list of possibilities. For example, applying this theorem to $p(x) = x^2 + 1$ we see that the only possible rational roots of $p(x)$ are ± 1, and it is easily seen that neither of these is a root. So by (???) this p is irreducible in $\mathbb{Z}[x]$.

* * Exercises * *

26.1. Let R be a GCD domain with $p(x), q(x) \in R[x]$ so that q is irreducible (hence prime), and let k be a natural number. Show that q^{k+1} divides p in $R[x]$ iff $q|p$ and $q^k|p'$ in $R[x]$. In particular, show that p is squarefree iff $\mathsf{gcd}(p, p') = 1$.

27 Over a UFD

Over a GCD domain, we have unique factorization of *primitive* polynomials.

> **Prop'n 27.1.** Let R be a GCD domain and $p(x) \in R[x]$ primitive. Then $p(x)$ can be written as a product of irreducibles in $R[x]$ in essentially one way (up to a rearrangement and multiplication by units).

Proof. (type this) □

The only obstacle to unique factorization of all polynomials over a GCD domain is the possibility that the content cannot be uniquely factored.

> **Cor. 27.2.** If R is a UFD, then $R[x]$ is a UFD.

Proof. (type this) □

* * Exercises * *

1. **Factorization if R is finite.** Note that if R is a finite ring of order m, then for any given degree d there are only finitely many polynomials in $R[x]$ of degree d; each such polynomial has $d+1$ coefficients, which each take one of m values. So the number of degree d polynomials over R is m^{d+1}.

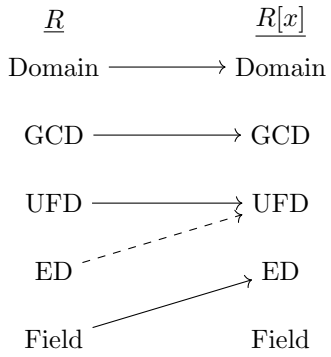
   > **Prop'n 27.3.** Let R be a domain. If $p(x) \in R[x]$ is reducible of degree d, then p has a divisor of degree $1 \leq k \leq \lfloor d/2 \rfloor$.

   If R is a finite UFD (of which examples do exist, such as $\mathbb{Z}/(p)$ with p a prime) we can use this fact to construct a naive factorization algorithm for $R[x]$. Let $p(x) \in R[x]$ have degree d. Choose some $1 \leq k < d$. There are m^{k+1} degree k polynomials over R, which can easily be enumerated. (If R is a field, we can consider only the monic polynomials, of which there are m^k.) Using polynomial long division we can determine whether any are divisors of $p(x)$; if so, recurse on the two factors, and if not, $p(x)$ has no factors of degree k. Repeat for each k in $[1, \lfloor d/2 \rfloor]$; if no divisor of p is found, then p is irreducible.

28 Irreducibility criteria

Summary of Chapter III

Polynomials provide a natural way to construct new rings out of old ones. Rings of polynomials inherit some nice structure from their coefficient rings, particularly if those rings are domains.



One might expect that detecting irreducibles and finding irreducible factorizations would be more difficult in $\mathbb{Z}[x]$, say, than in $\mathbb{Z}$, but in fact the opposite is true. This is ultimately due to the existence of the derivative on $R[x]$ and to some other aspects of the structure of $R[x]$ which will be explored later.

(@@@) We've seen here that if R is a UFD, then $R[x]$ is also a UFD, but have not seen any kind of algorithm which computes factorizations in $R[x]$. In the exercises of this and the next section we construct algorithms which work in some important special cases, including $\mathbb{Z}$ and $\mathbb{Z}/(p)$.

— IV —

Ideals and Quotients

29 Congruences

Recall the Construction Problem for rings: given a ring, we'd like to construct new rings out of its "parts". So far we've seen several examples of how this can be done including by generating subrings, taking direct sums, constructing rings of fractions, and using polynomials. This chapter is devoted to yet another method for building new rings out of old ones – by constructing *quotients*. This idea is very frequently a stumbling block for beginning students of mathematics, so we will spend some time on it. But our effort will be rewarded. Quotient rings are an extremely powerful and natural idea, and more generally the concept of "quotient" used here is pervasive in other branches of math. In a very specific sense quotient rings are the "opposites" of subrings.

Here is the basic idea: given a ring R and an equivalence relation Φ on R, we partition R into equivalence classes as R/Φ. Now we attempt to define an arithmetic on *equivalence classes* as follows.

1. Given two Φ-*classes*, say X and Y in R/Φ, first **choose** some representatives $x \in X$ and $y \in Y$. Now compute the sum $x + y$ in R. This element is in some other Φ-class Z in R/Φ. We define Z to be the sum of X and Y.

2. Likewise, to multiply classes, we **choose** representatives, multiply in R, and determine the class of the product.

This is a fine idea, and in fact this is precisely how arithmetic works in $\mathbb{Z}/(n)$. But unfortunately there is a problem: the sum of two classes may depend on our **choice** of representatives. Specifically, choosing one pair of representatives v_1 and w_1 may end up in one class, while another pair v_2 w_2 may end up in another class; see Figure 29.1 for an illustration. More concretely we can

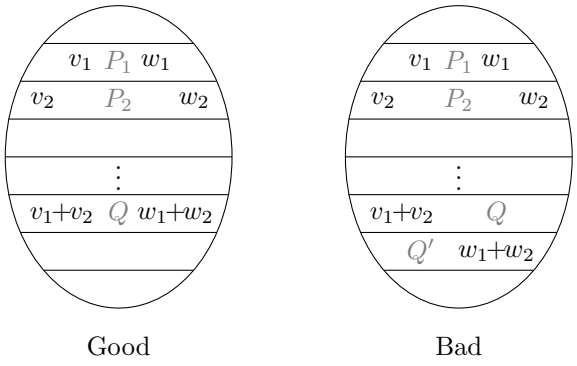


Figure 29.1: What can go wrong when making partitions into rings.

try this strategy by partitioning the set of integers into two sets; the primes

P and the nonprimes N. We might try to compute the sum $P + P$ using representatives 2 and 3; in this case

$$P + P = [2] + [3] = [2 + 3] = [5] = P.$$

But if we use representatives 3 and 7, we have

$$P + P = [3] + [7] = [3 + 7] = [10] = N.$$

This is a problem – it means that this + operation is not well-defined. To fix the problem, we just need to make sure our partition is chosen so that this bad thing never happens.

Def'n 29.1 (Ring Congruence)**.** Let R be a ring. An equivalence relation Φ on R is called a *congruence* if the following hold.

(i) If $x_1 \;\Phi\; x_2$ and $y_1 \;\Phi\; y_2$, then $(x_1 + y_1) \;\Phi\; (x_2 + y_2)$.

(ii) If $x_1 \;\Phi\; x_2$ and $y_1 \;\Phi\; y_2$, then $(x_1 y_1) \;\Phi\; (x_2 y_2)$.

So a congruence is a particular kind of equivalence relation. The congruence condition may at first glance seem to be very strong – so strong that it is not immediately clear that interesting congruences should even exist. (We will see that they do, and are abundant).

Examples

1. Let R be a ring. Then the *diagonal* relation $\Delta = \{(r, r) \mid r \in R\}$ and the *universal* relation $\nabla = \{(r, s) \mid r, s \in R\}$ are both congruences. These are called the *trivial* congruences on R.

2. The relation Φ on $\mathbb{Z}$ defined by $r \;\Phi\; s \Leftrightarrow n|(s - r)$ is a congruence. In fact $\mathbb{Z}$ here can be replaced by any commutative ring R and n by any fixed element of R.

Prop'n 29.2. Let R be a ring and Φ an equivalence on R.

(i) The operations + and $\cdot$ on R/Φ given by

$$[a] + [b] = [a + b] \quad \text{and} \quad [a] \cdot [b] = [a \cdot b]$$

are well-defined if and only if Φ is a congruence.

(ii) In this case, $(R/\Phi, +, \cdot)$ is a ring, called the *quotient* of R by Φ. If R is commutative, then R/Φ is commutative, and if R is unital, the R/Φ is unital.

Proof. (i) If we want to be pedantic (and we do!), the precise operations on R/Φ are

$$\mathsf{plus} = \{(([a],[b]),[a+b]) \mid a,b \in R\}$$

and

$$\mathsf{times} = \{(([a],[b]),[ab]) \mid a,b \in R\}\,.$$

First, suppose Φ is a congruence, and suppose we have elements $(([a_1],[b_1]),[a_1+b_1])$ and $(([a_2],[b_2]),[a_2+b_2])$ in plus such that $[a_1]=[a_2]$ and $[b_1]=[b_2]$. Then (by definition) $a_1 \;\Phi\; a_2$ and $b_1 \;\Phi\; b_2$. Since Φ is a congruence, $a_1+b_1 \;\Phi\; a_2+b_2$, and thus $[a_1+b_1]=[a_2+b_2]$; so plus is well-defined. A similar argument shows that times is well-defined. Conversely, suppose plus and times are well-defined. If $a_1 \;\Phi\; a_2$ and $b_1 \;\Phi\; b_2$, then $[a_1]=[a_2]$ and $[b_1]=[b_2]$, so that $[a_1+b_1]=[a_1]+[b_1]=[a_2]+[b_2]=[a_2+b_2]$, and thus $a_1+b_1 \;\Phi\; a_2+b_2$. Similarly we can show that $a_1b_1 \;\Phi\; a_2b_2$ so that Φ is a congruence.

(ii) Showing that all 6 of the ring axioms is somewhat tedious; for instance, to show A1 note that if $a,b,c \in R$, we have

$$([a]+[b])+[c] = [a+b]+[c] = [(a+b)+c] = [a+(b+c)] = [a]+[b+c] = [a]+([b]+[c]).$$

Similarly if R is commutative then $[a][b] = [ab] = [ba] = [b][a]$ for all $a,b \in R$. Finally, if R is unital, then evidently $[1_R]$ is a one in R/Φ.

□

Quotients are thus – potentially – a rich new source of examples of rings. There are some potential problems, however. First, it is not immediately clear how we can find interesting congruences. Second, once we have a congruence, it is not clear how we can effectively detect when two elements of R are in the same Φ-class. This means that even detecting when two elements of R/Φ are equal may be difficult – which has important computational consequences. Third, since quotient rings are sets of sets, the most natural way to define a mapping on R/Φ, in terms of class representatives, is fraught with danger. We expect that finding well-defined homomorphisms from a quotient ring will be difficult. Fortunately we will see that all three of these problems have nice resolutions.

As an example, what can we say about the ring $\mathbb{Q}[x]/(x^2+1)$? Using the division algorithm, every polynomial in $\mathbb{Q}[x]$ can be written as $(x^2+1)q(x)+r(x)$, where r is either 0, a nonzero constant, or a linear polynomial (@@@).

* * Exercises * *

29.1. Let R be a ring, and let Δ and ∇ denote the trivial and universal congruences on R. Show that $R \cong R/\Delta$ and $0 \cong R/\nabla$.

29.2. (One-sided ideals)

30 Ideals

We've seen that special equivalence relations called *congruences* can be used to build new rings out of old ones via the quotient construction. Now we'd like to understand the congruences in more depth. In this section we will see that the congruences on a ring correspond in a useful way to certain subrings. First, we show that the congruence class of zero plays a very special role.

Prop'n 30.1. Let R be a ring and Φ a congruence on R, and let I be the Φ-class of 0. Then we have the following.

(i) I is a subring of R.

(ii) I absorbs R under multiplication from either side. That is, if $a \in I$ and $r \in R$, then $ar \in I$ and $ra \in I$.

(iii) Every Φ-class is of the form $r + I = \{r + a \mid a \in I\}$ for some $r \in R$. In this case, $r + I$ is called the *r-coset* of I.

Proof. (i) Using the Subring Criterion, certainly $0_R \in [0_R]$, and if $x, y \in [0_R]$, then

$$x - y \in [0_R] - [0_R] = [0_R] \quad \text{and} \quad xy \in [0_R][0_R] = [0_R].$$

(ii) If $a \in [0_R]$ and $r \in R$, then

$$ar \in [0_R][r] = [0_R] \quad \text{and} \quad ra \in [r][0_R] = [0_R].$$

(iii) Let $[r]$ be a Φ-class; we claim that $[r] = r + I$. To see this, note that if $s \in [r]$, then $[s] = [r]$, so that $[s - r] = [0_R]$. That is, $s - r \in [0_R]$. Say $s - r = z$. Then $s = r + z \in r + I$. Conversely, if $s = r + z \in r + I$ then $s - r \in [0_R]$, so that $[s] = [r]$ and thus $s \in [r]$.

□

That is, if Φ is a congruence then every class of Φ is a coset of the class of zero. This alone is an interesting observation, but it has a practical consequence as well. Arithmetic in R/Φ can be carried out in terms of coset representatives.

Prop'n 30.2. Let R be a ring, Φ a congruence on R, and I the Φ-class of zero. Then for all $r, s \in R$, we have the following.

(i) $(r + I) + (s + I) = (r + s) + I$.

(ii) $(r + I)(s + I) = (rs) + I$.

Proof. It's important to realize that the $+$ symbol is used in three different senses here: the $+$ in $r + s$ is the plus in R, the $+$ between $r + I$ and $s + I$

is the plus in R/Φ, and the $+$ in $r + I$ is a structural symbol denoting cosets. With this in mind, we have

$$(r + I) + (s + I) = [r] + [s] = [r + s] = (r + s) + I$$

and

$$(r + I)(s + I) = [r][s] = [rs] = (rs) + I$$

as claimed. □

Apparently the congruence class of zero is special, so we extract its properties in the following definition.

Def'n 30.3 (Ideal)**.** Let R be a ring. A subring $I \subseteq R$ is called an *ideal* if I absorbs R under multiplication from both sides. That is, if $a \in I$ and $r \in R$, then $ar \in I$ and $ra \in I$.

Prop'n 30.4. Let R be a ring and $I \subseteq R$ a subset. Then the following are equivalent.

(i) I is an ideal of R.

(ii) The relation $\Phi = \{(r, s) \mid s - r \in I\}$ is a congruence on R with $[0_R] = I$. We say Φ is *induced* by I.

(iii) There is a surjective homomorphism $\varphi : R \to S$ with $\ker(\varphi) = I$. (If R is unital, this φ can be chosen to be unital.)

Proof. ((*i*) ⇒ (*ii*)) (@@@)

((*ii*) ⇒ (*iii*))

((*iii*) ⇒ (*i*))

□

Examples

1. The diagonal relation Δ is a congruence on any ring R, and the class of zero is $\{0_R\}$. So the cosets which comprise R/Δ are of the form

$$r + [0_R] = r + \{0_R\} = \{r + 0_R\} = \{r\};$$

that is, Δ is induced by the zero ideal and the elements of R/Δ are singletons.

2. The universal relation ∇ is a congruence on any ring R, and the class of zero is all of R. So there is only one coset in R/∇: $[0_R] = R$.

3. Let R be a commutative ring and $a \in R$ a fixed element. If Φ is the congruence given by $r \;\Phi\; s$ iff $a|(s-r)$, then the class of zero is the set

$$[0_R] = \{r \in R \mid a|r\}$$

of all elements of R which are divisible by a.

Congruences and ideals are equivalent: every ideal induces a congruence, and every congruence is induced by an ideal. For this reason from now on we will consider ideals alone and keep all congruences implicit. Typically we abuse the notation by referring to a quotient ring R/I (with I an ideal) when we really mean R/Φ, where Φ is the congruence induced by I.

Prop'n 30.5 (Ideal Criterion)**.** Let R be a ring. A subset $I \subseteq R$ is an ideal if and only if the following hold.

(i) I is not empty.

(ii) I is closed under subtraction.

(iii) I absorbs R under multiplication from either side.

Prop'n 30.6 (Unital Ideal Criterion)**.** Let R be a unital ring. A subset $I \subseteq R$ is an ideal if and only if I is not empty and $a - rbs \in I$ whenever $a, b \in I$ and $r, s \in R$.

Proof. (@@@) □

* * Exercises * *

30.1. Let R be a ring such that $R/(a)$ is finite for any nonzero $a \in R$. Define $N : R \to \mathbb{N}$ by $N(a) = |R/(a)|$. Show that N is a Euclidean norm on R. (@@@) is this true?

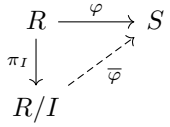
30.2. Every ideal is a union of associate classes.

31 The Isomorphism Theorems

Defining a mapping on a quotient set is generally difficult, for the same reason that defining operations on a quotient set is difficult. The most natural thing to try is to define our mapping in terms of representatives, but this is generally not well-defined.

The first important result about quotient rings gives us a standard way to construct homomorphisms on a quotient ring that bypasses this difficulty. This result, known as the First Isomorphism Theorem, is due to Emmy Noether, and is an important tool in ring theory.

Prop'n 31.1 (First Isomorphism Theorem for Rings)**.** Let R be a ring and I an ideal of R and suppose $\varphi : R \to S$ is a ring homomorphism. If $I \subseteq \ker(\varphi)$, then there is a ring homomorphism $\overline{\varphi} : R/I \to S$ such that $\overline{\varphi}(x + I) = \varphi(x)$. That is, $\varphi = \overline{\varphi} \circ \pi_I$, so that the following diagram commutes.



Proof. (@@@) □

Cor. 31.2. If $\varphi : R \to S$ is a homomorphism then the induced map $\overline{\varphi} : R/\ker(\varphi) \to S$ is injective. If φ is surjective then $\overline{\varphi}$ is an isomorphism.

Another way to state FIT is as follows: If φ is a ring homomorphism and I an ideal contained in the kernel of φ, then φ factors through the projection induced by I. This terminology is inspired by the commutative diagram associated to FIT, in which we literally have the "factorization" $\varphi = \overline{\varphi} \circ \pi$.

Two other nice results relate the ideals of a ring to ideals of its subrings and of its quotients. In a nutshell, every ideal in R/I corresponds nicely to an ideal in R. Ideals in a subring $S \subseteq R$ correspond (kind of) to ideals in R, but the correspondence is not nearly as nice. So while subrings and quotient rings are opposite concepts in a concrete sense, there is a fundamental asymmetry. This is not too surprising; for instance every domain (which tend to have lots of nontrivial ideals) can be embedded in a field (which we will see have none).

Prop'n 31.3 (Second Isomorphism Theorem for Rings)**.** Let R be a ring with $S \subseteq R$ a subring and $I \subseteq R$ an ideal. Then we have the following.

(i) $S \cap I$ is an ideal in S,

(ii) I is an ideal in $S + I$, and

(iii) $S/(S \cap I)$ is isomorphic to $(S + I)/I$.

Proof. (@@@) □

Prop'n 31.4 (Third Isomorphism Theorem)**.** Let R be a ring with $I \subseteq R$ an ideal.

(i) If $J \subseteq R$ is an ideal such that $I \subseteq J$, then I is an ideal in J.

(ii) If $K \subseteq R/I$ is an ideal, then there is an ideal $J \subseteq R$ such that $I \subseteq J$ and $K = J/I$.

(iii) If $J \subseteq R$ is an ideal and $I \subseteq J \subseteq R$, then $(R/I)/(J/I)$ is isomorphic to R/J.

Proof. (@@@) □

32 Ideal Arithmetic

The ideals of a ring enjoy an arithmetic of their own.

Def'n 32.1. If R is a ring, $\mathcal{Q}(R)$ denotes the set of all ideals of R.

Prop'n 32.2. Let R be a ring and $I, J \subseteq R$ ideals. We can construct two new subsets $I + J$ and IJ of R as follows.

$$\begin{aligned} I + J &= \{a + b \mid a \in I, b \in J\}, \text{ and} \\ IJ &= \left\{ \sum_{i=0}^{n} a_i b_i \mid n \in \mathbb{N}, a_i \in I, b_i \in J \right\}. \end{aligned}$$

Moreover, $I + J$ and IJ are both ideals of R, called an *ideal sum* and *ideal product*, respectively.

Prop'n 32.3. Let R be a ring and $I, J, K \subseteq R$ ideals. Then we have the following.

(i) $IJ \subseteq I \cap J$ and $I, J \subseteq I + J$.

(ii) $I + (J + K) = (I + J) + K$.

(iii) $I + 0 = 0 + I = I$.

(iv) $I + R = R + I = R$.

(v) $I(JK) = (IJ)K$.

(vi) $I0 = 0I = 0$.

(vii) $IR = RI = I$.

(viii) $I(J + K) = IJ + IK$ and $(I + J)K = IK + JK$.

* * Exercises * *

33 Maximal Ideals

34 Generating Sets

Prop'n 34.1. Let R be a ring, and let $\mathcal{I}$ be a collection of ideals of R. Then $\bigcap \mathcal{I}$ is an ideal of R.

Not every subset of R is an ideal; in fact most aren't. However, *every subset of R is contained in a unique smallest ideal.*

Prop'n 34.2 (Generated Ideals)**.** Let R be a ring and $A \subseteq R$ a subset. We define the set (A) by

$$(A) = \bigcap \{I \mid I \subseteq R \text{ is an ideal and } A \subseteq I\}.$$

Then the following hold.

(i) (A) is an ideal of R.

(ii) If A is an ideal of R, then $(A) = A$.

(iii) $A \subseteq (A)$.

(iv) If I is an ideal of R and $A \subseteq R$, then $(A) \subseteq I$.

We call (A) the ideal of R *generated by* A. If I is an ideal and $I = (A)$, we say that A is a *generating set* for I.

Prop'n 34.3. Let R be a ring and A and B be subsets of R. Then the following hold.

(i) $(A) + (B) = (A \cup B)$

(ii) $(A)(B) = (ab \mid a \in A, b \in B)$

Def'n 34.4. Let I be an ideal of a ring.

1. We say I is *finitely generated* if there is a finite set A such that $I = (A)$.
2. We say a subset A is a *minimal generating set* of I if $I = (A)$ and whenever $B \subsetneq A$ is a proper subset, $(B) \subsetneq (A)$ is also a proper subset.

Important note: "minimal" here does not mean *smallest size*, but rather *contains no redundant elements*. In general an ideal will have many minimal generating sets, and these may have very different cardinalities. For example, the ideal $(3) \subseteq \mathbb{Z}$ is minimally generated by the set $\{3\}$, but also by the set $\{6, 15\}$.

Prop'n 34.5. Let R be a commutative unital ring and A a subset of R. Then

$$(A) = \left\{ \sum_{i=0}^{n} r_i a_i \mid n \in \mathbb{N}, r_i \in R, a_i \in A \right\}.$$

That is, in a commutative unital ring, the ideal generated by A consists of all finite R-linear combinations of elements of A.

Proof. (@@@ type this) □

* * Exercises * *

34.1. Show that for any natural number $k \geq 1$, the ideal (3 in $\mathbb{Z}$ has a minimal generating set containing k elements.

35 Prime Ideals

36 The Chinese Remainder Theorem

Summary of Chapter IV

— V —

Fields

37 Subfields and Extensions

— A —

Rings with Absolute Value

In these notes we've been carefully avoiding the use of the real numbers, $\mathbb{R}$. Even though $\mathbb{R}$ is in some ways the most familiar example of a ring – most of your mathematical training so far has taken place inside $\mathbb{R}$ – our intuitive familiarity with $\mathbb{R}$ is deceptive. The real numbers are **very strange**, and it is far too easy to coast through calculus without appreciating how strange they are.

Nevertheless, the real numbers (and the closely related complex numbers) are extremely useful, and we lose something important by ignoring them completely. In this appendix we will fill this gap by constructing the real numbers from scratch. It turns out that if we approach this construction with the right amount of abstraction, we can build not just the real numbers $\mathbb{R}$ but a whole family of interesting rings.

38 Just Enough Metric Topology

Recall that if R is a ring and A any nonempty set, then the set R^A of all mappings $A \to R$ is a ring with pointwise arithmetic. As a special case, an element α of the ring $R^{\mathbb{N}}$ is called a *sequence* of R. We will denote the image of a sequence α at a particular natural number index i by α_i rather than the usual $\alpha(i)$. Given $r \in R$, we denote by $\widetilde{r}$ the constant sequence with value r; that is, $\widetilde{r}_i = r$ for all i.

It's about to start looking a lot like calculus in here. Do not be alarmed; we will only do as much analysis as is absolutely necessary. This section proceeds at a much faster pace.

Def'n 38.1 (Absolute Value)**.** Let R be a ring. A mapping $v : R \to \mathbb{Q}$ is called a (rational) *absolute value* if the following properties are satisfied.

(i) $v(x) \geq 0$ for all $x \in R$.

(ii) $v(x) = 0$ if and only if $x = 0_R$.

(iii) $v(xy) = v(x)v(y)$ for all $x, y \in R$.

(iv) The Triangle Inequality: $v(x + y) \leq v(x) + v(y)$ for all $x, y \in R$.

The basic example of an absolute value on a ring is the usual absolute value on $\mathbb{Q}$. For now, in the interest of brevity, we will save any more examples for the exercises or the later sections of this chapter.

Prop'n 38.2. If R is a ring with absolute value v then we have the following.

(i) If $x, y \in R$ and $xy = 0_R$, then either $x = 0_R$ or $y = 0_R$.

(ii) $v(x - y) \geq v(x) - v(y)$ for all $x, y \in R$.

(iii) $v(-x) = v(x)$ for all $x \in R$.

(iv) If R is unital, then $v(1_R) = 1$.

(v) If R is unital and $u \in R$ a unit, then $v(u^{-1}) = 1/v(u)$.

Proof. (i) If $xy = 0_R$, then $v(x)v(y) = v(xy) = 0$. Since $\mathbb{Q}$ is a field, either $v(x) = 0$ (so that $x = 0_R$) or $v(y) = 0$ (so that $y = 0_R$).

(ii) We have

$$v(x) = v(x - y + y) \leq v(x - y) + v(y)$$

so that $v(x) - v(y) \leq v(x - y)$.

(iii) Note that $v(-x)^2 = v((-x)^2) = v(x^2) = v(x)^2$, so that $v(-x)^2 - v(x)^2 = 0$ in $\mathbb{Q}$. Thus $(v(-x)+v(x))(v(-x)-v(x)) = 0$. Since $\mathbb{Q}$ is a field, one of these factors must be zero. But note that $v(-x) + v(x) = 0$ can only be true if $v(x) = v(-x) = 0$ since the value of v is nonnegative. In this case we have $x = 0_R = -x$, and so $v(x) = v(-x)$. If $x \neq 0_R$, then $v(x)+v(-x)$ is nonzero, so that $v(x) - v(-x) = 0$, and thus $v(-x) = v(x)$.

(iv) Note that $v(1_R) = v(1_R^2) = v(1_R)^2$, so that $v(1_R)(1 - v(1_R)) = 0$. Since $1_R \neq 0_R$, we have $v(1_R) \neq 0$, and thus $1 - v(1_R) = 0$. So $v(1_R) = 1$.

(v) We have $1 = v(1_R) = v(uu^{-1}) = v(u)v(u^{-1})$ as needed.

□

Def'n 38.3 (Bounded Sequence)**.** Let R be a ring with absolute value v. A sequence $\alpha : \mathbb{N} \to R$ is called *bounded* if there is a positive rational number B such that $v(\alpha_i) \leq B$ for all $i \in \mathbb{N}$. In this case we say that B is a *bound* of α.

For instance the constant sequence $r_i = r$ is bounded since $v(\widetilde{r_i}) \leq v(r) < v(r) + 1$.

Prop'n 38.4. Let R be a ring and v an absolute value on R. Then the set $\mathcal{B}$ of all bounded sequences of R is a subring of the ring of all sequences of R. If R is unital, then $\mathcal{B}$ is a unital subring.

Proof. Following the Subring Criterion (11.3), it suffices to show that $\mathcal{B}$ is nonempty and closed under multiplication and subtraction.

(i) Every constant sequence, such as $\widetilde{0_R}$, is bounded. So $\mathcal{B}$ is nonempty.

(ii) Suppose α and β are bounded sequences, with $v(\alpha_i) \leq B_\alpha$ and $v(\beta_i) \leq B_\beta$ for all natural numbers i. Then for all i we have

$$v((\alpha\beta)_i) = v(\alpha_i\beta_i) = v(\alpha_i)v(\beta_i) \leq B_\alpha B_\beta,$$

using the fact that $v(x)$ is nonnegative for all x. So $B_\alpha B_\beta$ is a bound of $\alpha\beta$.

(iii) Suppose again that α and β are bounded by B_α and B_β, respectively, and let $i \in \mathbb{N}$. Then we have

$$v((\alpha - \beta)_i) = v(\alpha_i - \beta_i) \leq v(\alpha_i) + v(-\beta_i) = v(\alpha_i) + v(\beta_i) \leq B_\alpha + B_\beta.$$

So $B_\alpha + B_\beta$ is a bound of $\alpha - \beta$.

Finally, if R is unital, then $\mathcal{B}$ contains the constant sequence $\widetilde{1_R}$, which is the one in $R^{\mathbb{N}}$. □

Def'n 38.5 (Convergent Sequence)**.** Let R be a ring with absolute value v. A sequence $\alpha : \mathbb{N} \to R$ is called *convergent* if there is an $\ell \in R$ such that for every rational number $\varepsilon > 0$, there exists a natural number N such that $v(\alpha_i - \ell) < \varepsilon$ whenever $i \geq N$. In this case we say ℓ is a *limit* of the sequence α.

Prop'n 38.6. The limit of a convergent sequence is unique. If α is a convergent sequence then we denote its (unique) limit by $\lim(\alpha)$ or by $\lim_n(\alpha_n)$.

Proof. Suppose α is a convergent sequence with limits ℓ_1 and ℓ_2, and suppose further that $\ell_1 \neq \ell_2$. In particular we have $v(\ell_1 - \ell_2) > 0$. Choose a rational number ε such that $0 < \varepsilon < v(\ell_1 - \ell_2)/2$. Since α is convergent with limits ℓ_1 and ℓ_2, there exists a natural number k sufficiently large that $v(\alpha_k - \ell_1) < \varepsilon$ and $v(\alpha_k - \ell_2) < \varepsilon$. Then we have

$$v(\ell_1 - \ell_2) = v(\ell_1 - \alpha_k + \alpha_k - \ell_2) \leq v(\ell_1 - \alpha_k) + v(\alpha_k - \ell_2) < 2\varepsilon < v(\ell_1 - \ell_2),$$

a contradiction. □

Prop'n 38.7. Let R be a ring with absolute value v. Then among the sequences of R we have the following.

(i) Every constant sequence is convergent.

(ii) Every convergent sequence is bounded.

Proof. (i) Let $r \in R$. Given a rational number $\epsilon > 0$, set $N = 0$. Now for all natural numbers $i \geq N$ we have

$$v(\widetilde{r}_i - r) = v(r - r) = v(0_R) = 0 < \varepsilon,$$

so $\widetilde{r}$ is convergent with limit r.

(ii) Suppose α is a convergent sequence with limit ℓ. Since $1 > 0$, there is a natural number N with the property that $v(\alpha_i - \ell) < 1$ for all $i \geq N$. Using 38.2(ii), we have $v(\alpha_i) - v(\ell) < 1$, and thus $v(\alpha_i) < v(\ell) + 1$ for all $i \geq N$. Now set

$$B = \max(v(\alpha_0), v(\alpha_1), \ldots, v(\alpha_{N-1}), v(\ell) + 1).$$

By construction we have $v(\alpha_i) \leq B$ for all $i \in \mathbb{N}$; thus B is a bound of α. □

Prop'n 38.8. Let R be a ring with absolute value v. Then the set $\mathcal{V}$ of all convergent sequences of R is a subring of the ring of all sequences of R. If R is unital, then $\mathcal{V}$ is a unital subring.

Proof. Again following the Subring Criterion (11.3) it suffices to show that $\mathcal{V}$ is nonempty and closed under subtraction and multiplication.

(i) Every constant sequence, such as $\widetilde{0_R}$, is convergent, so that $\mathcal{V}$ is not empty.

(ii) Suppose α and β are convergent sequences with $\lim(\alpha) = \ell_\alpha$ and $\lim(\beta) = \ell_\beta$, and let $\varepsilon > 0$ be an arbitrary rational number. Note that $\varepsilon/2$ is positive, so that (since α and β are convergent) there exist natural numbers N_α and N_β such that

$$v(\alpha_i - \ell_\alpha) < \varepsilon/2 \quad \text{when } i \geq N_\alpha$$

and

$$v(\beta_i - \ell_\beta) < \varepsilon/2 \quad \text{when } i \geq N_\beta.$$

Set $N = \max(N_\alpha, N_\beta)$. Now if $i \geq N$, we have the following.

$$\begin{aligned} v\left((\alpha - \beta)_i - (\ell_\alpha - \ell_\beta)\right) &= v(\alpha_i - \beta_i - \ell_\alpha + \ell_\beta) \\ &= v\left((\alpha_i - \ell_\alpha) + (\ell_\beta - \beta_i)\right) \\ &\leq v(\alpha_i - \ell_\alpha) + v(\ell_\beta - \beta_i) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

Thus $\alpha - \beta$ is convergent.

(iii) Again suppose α and β are convergent sequences with $\lim(\alpha) = \ell_\alpha$ and $\lim(\beta) = \ell_\beta$, and let $\varepsilon > 0$. By 38.7(ii), α is also bounded; say $v(\alpha_i) \leq B_\alpha$ for all i. Now let $U = \max(B_\alpha, v(\ell_\beta))$. Since $U > 0$, we have $\varepsilon/(2U) > 0$. Since α and β are convergent, there exist natural numbers N_α and N_β such that

$$v(\alpha_i - \ell_\alpha) < \frac{\varepsilon}{2U} \quad \text{when } i \geq N_\alpha$$

and

$$v(\beta_i - \ell_\beta) < \frac{\varepsilon}{2U} \quad \text{when } i \geq N_\beta.$$

Let $N = \max(N_\alpha, N_\beta)$. Then if $i \geq N$ we have the following.

$$\begin{aligned} v((\alpha\beta)_i - \ell_\alpha\ell_\beta) &= v(\alpha_i\beta_i - \ell_\alpha\ell_\beta) \\ &= v(\alpha_i\beta_i - \alpha_i\ell_\beta + \alpha_i\ell_\beta - \ell_\alpha\ell_\beta) \\ &= v\left(\alpha_i(\beta_i - \ell_\beta) + (\alpha_i - \ell_\alpha)\ell_\beta\right) \\ &\leq v(\alpha_i)v(\beta_i - \ell_\beta) + v(\alpha_i - \ell_\alpha)v(\ell_\beta) \\ &< U \cdot \frac{\varepsilon}{2U} + \frac{\varepsilon}{2U} \cdot U \\ &= \varepsilon. \end{aligned}$$

(Note that R is not necessarily commutative!) So $\alpha\beta$ is convergent.

Finally, if R is unital, then $\mathcal{V}$ contains the constant sequence $\widetilde{1_R}$, which is the one in $R^{\mathbb{N}}$. □

Cor. 38.9. If α and β are convergent sequences, then so are $\alpha - \beta$, $\alpha\beta$, and $\alpha + \beta$, and in fact we have $\lim(\alpha - \beta) = \lim(\alpha) - \lim(\beta)$, $\lim(\alpha\beta) = \lim(\alpha)\lim(\beta)$, and $\lim(\alpha + \beta) = \lim(\alpha) + \lim(\beta)$. In particular, letting $\mathcal{V}$ denote the set of convergent sequences of R, the mapping $\lim : \mathcal{V} \to R$ is a ring homomorphism.

Def'n 38.10 (Cauchy Sequence)**.** Let R be a ring with absolute value v. A sequence $\alpha : \mathbb{N} \to R$ is called *cauchy* if for every rational number $\varepsilon > 0$ there exists a natural number M such that $v(\alpha_j - \alpha_i) < \varepsilon$ whenever $i, j \geq M$.

Prop'n 38.11. Let R be a ring with absolute value v. Then among the sequences of R we have the following.

(i) Every convergent sequence is cauchy.

(ii) Every cauchy sequence is bounded.

Proof. (i) Let α be a convergent sequence with limit ℓ. Let $\varepsilon > 0$. Now $\varepsilon/2 > 0$, and since α is convergent there exists a natural number N such that $v(\alpha_i - \ell) < \varepsilon/2$ whenever $i, j \geq N$. Setting $M = N$, whenever $i \geq M$, we have

$$v(\alpha_j - \alpha_i) = v(\alpha_j - \ell + \ell - \alpha_i) \leq v(\alpha_j - \ell) + v(\ell - \alpha_i) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus α is cauchy.

(ii) Suppose α is cauchy. Since $1 > 0$, there is a natural number M such that $v(\alpha_j - \alpha_i) < 1$ whenever $i, j \geq M$. In particular, we have $v(\alpha_j - \alpha_M) < 1$ whenever $j \geq M$. Using 38.2(ii) we have $v(\alpha_j) - v(\alpha_M) < 1$, so that $v(\alpha_j) < v(\alpha_M) + 1$ whenever $j \geq M$. Now set

$$B = \max(v(\alpha_0), v(\alpha_1), \ldots, v(\alpha_{M-1}, v(\alpha_M) + 1).$$

We then have $v(\alpha_i) < B$ for all $i \in \mathbb{N}$, so that B is a bound for α.

□

Prop'n 38.12. Let R be a ring with absolute value v. Then the set $\mathcal{C}$ of all cauchy sequences of R is a subring of the ring of all sequences of R. If R is unital, then $\mathcal{C}$ is a unital subring.

Proof. Again we follow the Subring Criterion (11.3), showing that $\mathcal{C}$ is not empty and closed under subtraction and multiplication.

(i) Every constant sequence, such as $\widetilde{0_R}$, is convergent, hence cauchy. So $\mathcal{C}$ is not empty.

(ii) Suppose α and β are cauchy, and let $\varepsilon > 0$. Now $\varepsilon/2 > 0$, and so there exist natural numbers M_α and M_β such that

$$v(\alpha_j - \alpha_i) < \varepsilon/2 \quad \text{when } i, j \geq M_\alpha$$

and

$$v(\beta_j - \beta_i) < \varepsilon/2 \quad \text{when } i, j \geq M_\beta.$$

Let $M = \max(M_\alpha, M_\beta)$. Now if $i, j \geq M$, we have the following.

$$\begin{aligned} v\left((\alpha - \beta)_j - (\alpha - \beta)_i\right) &= v(\alpha_j - \beta_j - \alpha_i + \beta_j) \\ &= v\left((\alpha_j - \alpha_i) + (\beta_i - \beta_j)\right) \\ &\leq v(\alpha_j - \alpha_i) + v(\beta_j - \beta_i) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

Thus $\alpha - \beta$ is cauchy as needed.

(iii) Suppose α and β are cauchy, and choose a rational number $\varepsilon > 0$. Now by 38.11(ii), both α and β are bounded; say B_α and B_β are rational numbers such that $v(\alpha_i) \leq B_\alpha$ and $v(\beta_i) \leq B_\beta$ for all i. Now $\varepsilon/(2B_\beta) > 0$, and since α is cauchy, there is a natural number M_α such that $v(\alpha_j - \alpha_i) < \varepsilon/(2B_\alpha)$ whenever $i, j \geq M_\alpha$. Similarly, there is a natural number M_β such that $v(\beta_j - \beta_i) < \varepsilon/(2B_\alpha)$ whenever $i, j \geq M_\beta$. Now let $M = \max(M_\alpha, M_\beta)$. Then if $i, j \geq M$, we have the following.

$$\begin{aligned} v\left((\alpha\beta)_j - (\alpha\beta)_i\right) &= v(\alpha_j\beta_j - \alpha_i\beta_i) \\ &= v(\alpha_j\beta_j - \alpha_j\beta_i + \alpha_j\beta_i - \alpha_i\beta_i) \\ &= v\left(\alpha_j(\beta_j - \beta_i) + (\alpha_j - \alpha_i)\beta_i\right) \\ &\leq v(\alpha_j)v(\beta_j - \beta_i) + v(\alpha_j - \alpha_i)v(\beta_i) \\ &< B_\alpha \frac{\varepsilon}{2B_\alpha} + \frac{\varepsilon}{2B_\beta} B_\beta \\ &= \varepsilon. \end{aligned}$$

Thus $\alpha\beta$ is cauchy as needed.

Finally, if R is unital then the constant sequence $\widetilde{1_R}$ is cauchy. □

Prop'n 38.13. If α is a cauchy sequence which *does not* converge to 0_R, then there is a rational number $\delta > 0$ and a natural number T such that $v(\alpha_i) \geq \delta$ for all $i \geq T$. That is, if a cauchy sequence does not converge to zero, the it is eventually bounded away from zero.

Proof. Since α does not converge to zero, there exists a rational number ε such that for every natural number N, there exists an index $i \geq N$ such that $v(\alpha_i) \geq \varepsilon$. (This is the negation of the statement "α converges to zero".) Now since α is cauchy, and since $\varepsilon/2 > 0$, there is a natural number M such that $v(\alpha_j - \alpha_i) < \varepsilon/2$ whenever $i, j \geq M$. There exists an index $k \geq M$ such that $v(\alpha_k) \geq \varepsilon$. Now $v(\alpha_k - \alpha_i) < \varepsilon/2$ for all $i \geq M$, and so for all i we have the following.

$$\begin{aligned} \varepsilon &\leq v(\alpha_k) \\ &= v(\alpha_k - \alpha_i + \alpha_i) \\ &\leq v(\alpha_k - \alpha_i) + v(\alpha_i) \\ &< \varepsilon/2 + v(\alpha_i). \end{aligned}$$

So $v(\alpha_i) > \varepsilon/2$. The result then holds with $\delta = \varepsilon/2$ and $T = M$. □

So we have a chain of inclusions among the sequences:

$$\text{convergent} \subseteq \text{cauchy} \subseteq \text{bounded} \subseteq \text{sequences}.$$

Moreover this is not just a chain of sub*sets*, but of sub*rings*, meaning that we get some additional ring-theoretic results for free. We say that a sequence is *null* if it converges to 0_R. Then we can show that the set of null sequences is an *ideal* in the ring of bounded sequences.

Prop'n 38.14. Let R be a ring with absolute value v. Then the set $\mathcal{N}$ of all null sequences of R is a two-sided ideal in the ring $\mathcal{B}$ of all bounded sequences of R. In particular, $\mathcal{N}$ is also a two-sided ideal in the ring $\mathcal{C}$ of all cauchy sequences and the ring $\mathcal{V}$ of all convergent sequences.

Proof. Certainly we have $\mathcal{N} \subseteq \mathcal{V} \subseteq \mathcal{C} \subseteq \mathcal{B}$. Now we establish that $\mathcal{N}$ is a subring of $R^{\mathbb{N}}$. Using the Subring Criterion, we certainly have $\widetilde{0_R} \in \mathcal{N}$, and if $\alpha, \beta \in \mathcal{N}$ then $\alpha - \beta$ and $\alpha\beta$ are convergent by 38.8 and we have

$$\lim(\alpha - \beta) = \lim(\alpha) - \lim(\beta) = 0_R - 0_R = 0_R$$

and

$$\lim(\alpha\beta) = \lim(\alpha)\lim(\beta) = 0_R \cdot 0_R = 0_R$$

as needed. It remains to be seen that $\mathcal{N}$ absorbs $\mathcal{B}$ under multiplication from either side. To this end, suppose α is a bounded sequence of R with $v(\alpha_i) \leq B$ for all i, and that ζ is a null sequence. Let $\varepsilon > 0$. Now $\varepsilon/B > 0$, and since

ζ converges to 0_R, there exists a natural number N such that $v(\alpha_i) < \varepsilon/B$ whenever $i \geq N$. Now for all $i \geq N$ we have

$$v((\alpha\zeta)_i) = v(\alpha_i\zeta_i) = v(\alpha_i)v(\zeta_i) \leq \frac{\varepsilon}{B}B = \varepsilon$$

and similarly $v((\zeta\alpha)_i) < \varepsilon$. So $\alpha\zeta$ and $\zeta\alpha$ are null sequences as needed. □

Prop'n 38.15. The set of null sequences is a prime ideal in the set of cauchy sequences. That is, if α and β are cauchy sequences and $\alpha\beta$ converges to 0_R, then either α or β must converge to 0_R.

Proof. Suppose α and β are cauchy and that $\alpha\beta$ converges to 0_R. Suppose further without loss of generality that α does not converge to 0_R. By 38.13, there exists a rational number $\delta > 0$ and a natural number T such that $v(\alpha_i) \geq \delta$ whenever $i \geq T$. Now let $\varepsilon > 0$; then also $\varepsilon\delta > 0$. Since $\alpha\beta$ converges to 0_R, there is a natural number N such that $v(\alpha_i\beta_i) < \varepsilon\delta$ whenever $i \geq N$. If we set $N_1 = \max(N, T)$, then whenever $i \geq N_1$ in fact we have

$$\delta v(\beta_i) \leq v(\alpha_i)v(\beta_i) = v(\alpha_i\beta_i) < \delta\varepsilon;$$

since $\delta > 0$, we have $v(\beta_i) < \varepsilon$ for all $i \geq N_1$. Thus β converges to 0_R. □

We are finally prepared for the punchline of this section.

Def'n 38.16 (Cauchy Completion)**.** Let R be a ring with absolute value v. Letting $\mathcal{C}$ denote the set of cauchy sequences on R and $\mathcal{N}$ the set of null sequences on R, we define $\overline{R}_v = \mathcal{C}/\mathcal{N}$ to be the *cauchy completion* of R with respect to v.

Prop'n 38.17. (i) If R is a commutative unital ring then $\overline{R}_v$ is an integral domain.

(ii) Let R be a ring with absolute value v. Then the mapping $\iota : R \to \overline{R}_v$ given by $\iota(r) = \widetilde{r} + N$ is an injective ring homomorphism.

Proof. (@@@) □

* * Exercises * *

38.1. Let k be a field and let R be the field of rational functions in one variable over k. Show that $v(p/q) = 2^{degp-degq}$ is an absolute value on R. (@@@) is this true?

39 The Real Numbers

We are now prepared to define the set of real numbers; (most of) the heavy lifting is already done.

Def'n 39.1 (Real Numbers)**.** The usual absolute value $|\cdot|$ on the set $\mathbb{Q}$ of rational numbers is an absolute value in the sense of 38.1, and so the constructions of Section 38 are immediately available. We define the set of *real numbers* $\mathbb{R}$ to be the cauchy completion $\overline{\mathbb{Q}}_{|\cdot|}$ of $\mathbb{Q}$ with respect to $|\cdot|$. Since $\mathbb{Q}$ is a field, $\mathbb{R}$ is also a field.

Doing arithmetic in $\mathbb{R}$ is a bit awkward because elements of $\mathbb{R}$ are equivalence classes of cauchy sequences. Even detecting whether two specific real numbers are equal to each other is difficult.

Def'n 39.2. Given two cauchy sequences of rational numbers, α and β, we say that $\alpha \leq \beta$ if either $\beta - \alpha$ converges to 0 or there exists a natural number W such that $\alpha_i < \beta_i$ for all $i \geq W$.

Prop'n 39.3. If α, $\overline{\alpha}$, β, and $\overline{\beta}$ are cauchy sequences of rational numbers such that $\alpha - \overline{\alpha}$ and $\beta - \overline{\beta}$ converge to zero and $\alpha \leq \beta$, then $\overline{\alpha} \leq \overline{\beta}$.

Proof. If $\beta - \alpha$ converges to zero, we have the following.

$$\begin{aligned} 0 &= \lim(\beta - \alpha) - \lim\left(\beta - \overline{\beta}\right) + \lim(\alpha - \overline{\alpha}) \\ &= \lim\left(\beta - \alpha - \beta + \overline{\beta} + \alpha - \overline{\alpha}\right) \\ &= \lim\left(\overline{\beta} - \overline{\alpha}\right). \end{aligned}$$

Thus $\overline{\alpha} \leq \overline{\beta}$. Suppose instead that $\beta - \alpha$ does not converge to zero. Since $\alpha \leq \beta$, there is a natural number W such that $\alpha_i < \beta_i$ whenever $i \geq W$. Also, by 38.13, there is a rational number δ and a natural number T such that $|\beta_i - \alpha_i| \geq \delta$ whenever $i \geq T$. Since $\delta/2 > 0$ and $\alpha - \overline{\alpha}$ converges to 0, there is a natural number N_α such that $|\alpha_i - \overline{\alpha}_i| < \delta/2$ when $i \geq N_\alpha$. Similarly there is an N_β such that $|\beta_i - \overline{\beta}_i| < \delta/2$ when $i \geq N_\beta$.

Let $W_1 = \max(W, T, N_\alpha, N_\beta)$. Then for all $i \geq W_1$, we have $\alpha_i < \beta_i$, and thus $\beta_i - \alpha_i = |\beta_i - \alpha_i| \geq \delta$. Now we have the following.

$$\begin{aligned} \delta - (\overline{\beta}_i - \overline{\alpha}_i) &\leq (\beta_i - \alpha_i) - (\overline{\beta}_i - \overline{\alpha}_i) \\ &= (\beta_i - \overline{\beta}_i) + (\overline{\alpha}_i - \alpha_i) \\ &\leq |\beta_i - \overline{\beta}_i| + |\overline{\alpha}_i - \alpha_i| \\ &< \delta/2 + \delta/2 \\ &= \delta. \end{aligned}$$

So we have $0 < \overline{\beta}_i - \overline{\alpha}_i$, and thus $\overline{\alpha}_i < \overline{\beta}_i$ for all $i \geq W_1$. Thus $\overline{\alpha} \leq \overline{\beta}$. □

Cor. 39.4. We have a relation $\leq$ on $\mathbb{R}$ defined as follows: given $[\alpha]$ and $[\beta]$ in $\mathbb{R}$, we say that $[\alpha] \leq [\beta]$ if either $\beta - \alpha$ converges to 0 or there is a natural number W such that $\alpha_i < \beta_i$ whenever $i \geq W$. We say $[\alpha] < [\beta]$ if $[\alpha] \leq [\beta]$ and $[\alpha] \neq [\beta]$.

Prop'n 39.5. The relation $\leq$ on $\mathbb{R}$ has the following properties.

(i) $[\alpha] \leq [\alpha]$ for all $[\alpha] \in \mathbb{R}$.

(ii) If $[\alpha] \leq [\beta]$ and $[\beta] \leq [\alpha]$, then $[\alpha] = [\beta]$.

(iii) If $[\alpha] \leq [\beta]$ and $[\beta] \leq [\gamma]$, then $[\alpha] \leq [\gamma]$.

(iv) If $[\alpha]$ is a real number, then either $[\alpha] > [\widetilde{0}]$, $[\alpha] = [\widetilde{0}]$, or $[\alpha] < [\widetilde{0}]$.

Proof.

(i) Since $\alpha - \alpha$ converges to 0, we have $[\alpha] \leq [\alpha]$.

(ii) Suppose $[\alpha] \leq [\beta]$ and $[\beta] \leq [\alpha]$. Now if either $\beta - \alpha$ or $\alpha - \beta$ converge to zero, we have $[\alpha] = [\beta]$. So we assume that this does not happen. In this case we have natural numbers W_1 and W_2 such that $\alpha_i < \beta_i$ whenever $i \geq W_1$ and $\beta_i < \alpha_i$ whenever $i \geq W_2$. But if $i \geq \max(W_1, W_2)$, then $\alpha_i < \beta_i < \alpha_i$, a contradiction. So we have $[\alpha] = [\beta]$ as needed.

(iii) If either $\beta - \alpha$ or $\gamma - \beta$ converge to zero, we have $[\alpha] = [\beta]$ or $[\beta] = [\gamma]$; in either case, $[\alpha] \leq [\gamma]$. Now suppose that neither $\beta - \alpha$ nor $\gamma - \beta$ converges to zero. Then there exist natural numbers W_1 and W_2 such that $\alpha_i < \beta_i$ when $i \geq W_1$ and $\beta_i < \gamma_i$ when $i \geq W_2$. Letting $W = \max(W_1, W_2)$, whenever $i \geq W$ we have

$$\gamma_i - \alpha_i = \gamma_i - \beta_i + \beta_i - \alpha_i > 0$$

so that $\alpha_i < \gamma_i$. Thus $[\alpha] \leq [\gamma]$ as needed.

(iv) Let $[\alpha]$ be a real number and suppose $[\alpha] \neq [\widetilde{0}]$. That is, α does not converge to zero. By 38.13, there is a rational number δ and a natural number T such that $|\alpha_i| \geq \delta$ when $i \geq T$. Also, since α is cauchy, there exists a natural number M such that $|\alpha_j - \alpha_i| < \delta$ when $i, j \geq M$. Set $W = \max(T, M)$. We will show that if $i, j \geq W$, then α_i and α_j must have the same sign in $\mathbb{Q}$ (either both positive or both negative). To this end, suppose $\alpha_j > 0$ and $\alpha_i < 0$. On one hand we have $|\alpha_j - \alpha_i| < \delta$. But on the other hand we have $\alpha_j \geq \delta$ and $-\alpha_i \geq \delta$. So

$$2\delta \leq \alpha_j - \alpha_i \leq |\alpha_j - \alpha_i| < \delta,$$

a contradiction. Likewise if $\alpha_j < 0$ and $\alpha_i > 0$ then

$$2\delta \leq \alpha_i - \alpha_j \leq |\alpha_i - \alpha_j| < \delta.$$

That is, for all $i \geq W$, the α_i have the same sign, either positive or negative. In the first case we have $[\widetilde{0}] \leq [\alpha]$, and in the second case $[\alpha] \leq [\widetilde{0}]$. But since $[\alpha] \neq [\widetilde{0}]$, in fact either $[\alpha] < [\widetilde{0}]$ or $[\alpha] > [\widetilde{0}]$. □

Prop'n 39.6. The following properties hold.

(i) If $[\alpha] \leq [\beta]$ then $[\alpha] + [\gamma] \leq [\beta] + [\gamma]$.

(ii) If $[\widetilde{0}] \leq [\alpha]$ and $[\widetilde{0}] \leq [\beta]$ then $[\widetilde{0}] \leq [\alpha][\beta]$.

(iii) $[\widetilde{0}] < [\widetilde{1}]$.

Proof.

(i) If $\beta - \alpha$ converges to 0, then $[\alpha] = [\beta]$ and the result holds. If $\beta - \alpha$ does not converge to 0, then there is an index W such that $\alpha_i < \beta_i$ for all $i \geq W$. For all such i, we have $\alpha_i + \gamma_i < \beta_i + \gamma_i$, so that $[\alpha] + [\gamma] \leq [\beta] + [\gamma]$ as claimed.

(ii) If either $[\alpha]$ or $[\beta]$ converge to zero, then $[\alpha][\beta] = [\widetilde{0}]$. Suppose instead that neither converges to zero. Then there exist natural numbers W_α and W_β such that $\alpha_i > 0$ when $i \geq W_\alpha$ and $\beta_i > 0$ when $i \geq W_\beta$. Let $W = \max(W_\alpha, W_\beta)$. Then for all $i \geq W$, we have $\alpha_i \beta_i > 0$, so that $[\widetilde{0}] \leq [\alpha][\beta]$ as claimed.

(iii) Certainly $0 < 1$, so that $[\widetilde{0}] \leq [\widetilde{1}]$. Since $1 \neq 0$, $[\widetilde{0}] \neq [\widetilde{1}]$.

□

Prop'n 39.7. (@@@) every nonempty set with an upper bound has a least upper bound.

We can summarize (@@@) in the following proposition.

Prop'n 39.8. There is a field $\mathbb{R}$, whose elements are called *real numbers*, which comes with a linear order relation $\leq$ having the following properties.

(i) $0 < 1$.

(ii) If $\alpha, \beta, \gamma \in \mathbb{R}$ such that $\alpha \leq \beta$, then $\alpha + \gamma \leq \beta + \gamma$.

(iii) If $\alpha, \beta \in \mathbb{R}$ such that $0 \leq \alpha$ and $0 \leq \beta$, then $0 \leq \alpha\beta$.

(iv) If $S \subseteq \mathbb{R}$ is a nonempty set which has an upper bound, then S has a *least* upper bound.

Working directly with equivalence classes of cauchy sequences - that is, real numbers - can get awkward very fast. Fortunately, there is a better way! It turns out that 39.8 *characterizes* the real numbers, in the sense that any other field F with a linear order relation $\preceq$ that *also* satisfies the properties in 39.8 must be isomorphic to $\mathbb{R}$. This means that any theorem which can be proved about $\mathbb{R}$ can be proved using the properties of 39.8, together with the definition of field and linear order, as a list of "axioms", rather than the complicated construction using cauchy sequences. In fact this is the approach typically taken by high school textbooks, if you pay close attention. With this in mind, why should we bother explicitly constructing $\mathbb{R}$ in the first place? The (potential) problem is that we can make lists of axioms all day and derive theories based on them, but without a concrete *model* of those axioms we cannot assume that the axioms are consistent with each other. This is a great example of the abstract and concrete points of view each bringing something useful to the table: the axioms are nice to work with, and the model tells us the axioms describe something interesting.

Prop'n 39.9 (Bolzano's Theorem for Polynomials)**.** Let $p(x) \in \mathbb{R}[x]$. If a and b are real numbers such that $a < b$ and $p(a)$ and $p(b)$ have opposite signs, then there exists a real number c such that $p(c) = 0$.

Proof. Suppose, without loss of generality, that $f(a) < 0$ and $f(b) > 0$. Now define the set $S = \{x \in [a, b] \mid f(x) \leq 0\}$. Certainly S is not empty, since $f(a) < 0$. Moreover, S has an upper bound, since $x < b$ for all $x \in S$. Thus S has a least upper bound, say $c \in S$. (@@@) □

Cor. 39.10.

(i) If $p(x) \in \mathbb{R}[x]$ has odd degree, then $p(x)$ has a root in $\mathbb{R}$.

(ii) If $\alpha \in \mathbb{R}$ is positive, then $q(x) = x^2 - \alpha$ has a unique positive root in $\mathbb{R}$, denoted $\sqrt{\alpha}$.

Proof. (@@@) □

* * Exercises * *

39.1. (@@@) Let F be an ordered field with the least upper bound property.

(i) Note that $\mathbb{Q}$ embeds uniquely into F and $\mathbb{R}$.

(ii) (@@@) map LUBs to LUBs yadda yadda

(iii) Show that this map is an isomorphism of ordered fields.

40 The p-adic Numbers

— B —

Posets

41 Posets and Zorn's Lemma

Def'n 41.1 (Poset)**.** A *partially ordered set* or *poset* is a set P equipped with a binary relation $\preceq$ having the following properties.

PO1. $x \preceq x$ for all $x \in P$.

PO2. If $x \preceq y$ and $y \preceq x$, then $x = y$.

PO3. If $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

Def'n 41.2. Let P be a poset.

(i) A *chain* is a mapping $x : \mathbb{N} \to P$ which is order-preserving; that is, such that $x_i \leq x_j$ whenever $i \leq j$.

(ii) Let $S \subseteq P$ be a subset. An element $b \in P$ is called an *upper bound* of S if $s \preceq b$ for all $s \in S$.

(iii) An element $m \in P$ is called *maximal* if whenever $x \in P$ such that $m \preceq x$, in fact $x = m$. In other words, the inequality $m \preceq x$ has only the trivial solution.

We take the following statement as an axiom, which for historical reasons is called a lemma.

Axiom 41.3 (Zorn's Lemma)**.** If P is a nonempty poset in which every chain has an upper bound, then P has at least one maximal element.

Note that Zorn's Lemma tells us nothing about how such maximal elements are to be *found*, only that they exist. In this sense it is a nonconstructive statement very similar to the Well-Ordering Property of natural numbers; proofs which use Zorn may be slick, but cannot generally be turned into algorithms.

In practice we occasionally need to show that some "extreme" doo-dad having certain properties must exist, and there are two basic ways to do this. The first and usually preferable way is to explicitly construct an example of such a doodad. But this is not always possible, or may not be possible at the desired level of generality; in this case we use the second way: Zorn's Lemma. This requires us to capture our "extreme" doodads as the maximal elements of an appropriate poset. For example, in many cases it is easy to show that a *particular* ring such as $\mathbb{Z}$ or $\mathbb{Q}[x]$ has maximal ideals. But to show that *every* ring has maximal ideals, by far the easiest strategy is to use Zorn's Lemma in the poset of ideals.

Here is a concrete example. Suppose we want to prove that every nonempty set A has maximal proper subsets; that is, subsets $B \subsetneq A$ such that the inequality $B \subseteq X \subsetneq A$ has no nontrivial solutions. It is straightforward enough

to construct an example – we can show that $B = A \setminus \{x\}$, where $x \in A$, is maximal. For the sake of illustration let's suppose that constructing this example is very difficult, so we must instead use Zorn's Lemma. That proof proceeds as follows.

(i) Verify that the set $P = \{B \in \mathcal{P}(A) \mid B \neq A\}$ of proper subsets of A is partially ordered by set containment. This follows from the basic properties of $\subseteq$.

(ii) Verify that P is not empty. Since $A \neq \varnothing$ and $\varnothing \subseteq A$, we have $\varnothing \in P$.

(iii) Verify that every chain in P has an upper bound. Suppose X_i is a chain of proper subsets of A; we need to show that $\{X_i \mid i \in \mathbb{N}\}$ has an upper bound. Usually this step involves some kind of union; in this case we use $B = \bigcup X_i$. Clearly $X_i \subseteq B$ for all $i \in \mathbb{N}$. But to show that B is an upper bound of X_i we have to make sure that B is actually in P. (This is an important but easy-to-forget step!) Certainly $B \subseteq A$; suppose this containment is not proper, so that $B = A$. So every element of A is in one of the X_i. (@@@) This example is broken!

It is possible that a poset has maximal elements, but there exist chains with no upper bound – that is, the converse of Zorn's Lemma is not true.

* * Exercises * *

41.1. (@@@) The set $\mathbb{N}$ of natural numbers is a poset under the usual $\leq$ relation. However, this poset has chains with no upper bound, such as $x_n = 2n$. Thus Zorn's Lemma does not apply.

42 The Axiom of Choice

Index