Part II

Rings

Chapter 7:
Rings

While monoids are defined by one operation, groups are arguably defined by two: addition and subtraction, for example, or multiplication and division. The second operation is so closely tied to the first that we consider groups to have only one operation, for which (unlike monoids) every element has an inverse.

Of course, a set can be closed under more than one operation; for example, $\mathbb{Z}$ is closed under both addition and multiplication. As with subtraction, it is possible to define the multiplication of integers in terms of addition, just as we did with groups. However, this is not possible for all sets where an addition and a multiplication are both defined. Think of the multiplication of polynomials; how would you define $(x+1)(x-1)$ as repeated addition of $x-1$, a total of $x+1$ times? Does that even make sense? This motivates the study of a structure that incorporates common properties of two operations, which are related as loosely as possible.

Section 7.1 of this chapter introduces us to this structure, called a *ring*. A ring has two operations, "addition" and "multiplication". As you should expect from your experience with groups, what we call "addition" and "multiplication" may look nothing at all like the usual addition and multiplication of numbers. In fact, while the multiplication of integers has a natural definition from addition, multiplication in a ring may have absolutely nothing to do with addition, with one exception: the distributive property must still hold.

The rest of the chapter examines special kinds of rings. In Section 7.2 we introduce special kinds of rings that model useful properties of $\mathbb{Z}$ and $\mathbb{Q}$. In Section 7.3 we introduce rings of polynomials. The Euclidean algorithm, which proved so important in chapter 6, serves as the model for a special kind of ring described in Section 7.4.

A concept related to monoids is useful for definitions related to rings.

Definition 7.1. Let S be a set, and $\circ$ an operation. We say that $(S, \circ)$ is a **semigroup** if its operation is closed and associative, although it might not have an identity element.

Notice that

- a monoid is a semigroup,
- a semigroup is almost a monoid, but lacks an identity, and
- the "absorbing subsets" of Section 1.4 are "subsemigroups" of monoids.

A "semigroup" is "half a group", in that it satisfies half of the properties of a group. We will take this up further in Chapter 8.

7.1: A structure for addition and multiplication

What sort of properties do we associate with both addition and multiplication? We typically associate the properties of addition with an abelian group, and the properties of multiplication with a monoid, although it really depends on the set. The most basic properties of multiplication are encapsulated by the notion of a semigroup, so we'll start from there, and add more as needed.

Definition 7.2. Let R be a set *with at least one element*, and $+$ and $\times$ two binary operations on that set. We say that $(R,+,\times)$ is a **ring** if it satisfies the following properties:

(R1) $(R,+)$ is an abelian group.

(R2) $(R,\times)$ is a semigroup.

(R4) R satisfies the distributive property of addition over multiplication: that is,
for all $a,b,c\in R$, $a(b+c)=ab+ac$ and $(a+b)c=ac+bc$.

Notation 7.3. As with groups, we usually refer simply to the ring as R, rather than $(R,+,\times)$. Since $(R,+)$ is an abelian group, the ring has an additive identity, 0. We sometimes write 0_R to emphasize that it is the additive identity of R.

Notice the following:

- While addition is commutative on account of (R1), multiplication need not be.
- There is no requirement that a multiplicative identity exists.
- There is no requirement that multiplicative inverses exist.
- There is no guarantee (yet) that the additive identity interacts with multiplication according to properties you have seen before. In particular, there is *no guarantee* that
  - the zero-product rule holds; or even that
  - $0_R\cdot a=0_R$ for any $a\in R$.

Example 7.4. Let $R=\mathbb{R}^{m\times m}$ for some positive integer m. It turns out that R is a ring under the usual addition and multiplication of matrices. After all, Example 1.8 shows that the matrices satisfy the properties of a monoid under multiplication, and Example 2.4 shows that they are a group under addition, though most of the work was done in Section 0.3. The only part missing is distribution, and while that isn't hard, it is somewhat tedious, so we defer to your background in linear algebra.

However, we do want to point out something that should make you at least a *little* uncomfortable. Let

$$A=\begin{pmatrix}1&0\\0&0\end{pmatrix}\quad\text{and}\quad B=\begin{pmatrix}0&0\\0&1\end{pmatrix}.$$

Routine computation shows that

$$AB=\begin{pmatrix}0&0\\0&0\end{pmatrix},$$

or in other words, $AB=0$. This is true even though $A,B\neq 0$! Hence

Not every ring R satisfies the **zero product property**

$$\forall a,b\in R\qquad ab=0\quad\Longrightarrow\quad a=0\text{ or }b=0.$$

Example 7.4 shouldn't surprise you that much; first, you've seen it in linear algebra, and second, you met zero divisors in Section 6.4. In fact, we will shortly generalize that idea into zero divisors for rings.

Likewise, the sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, with which you are long familiar, are also rings. We omit the details, but you should think about them a little bit, and ask your instructor if some part of it isn't clear. You will study other example rings in the exercises. For now, we prove a familiar property of the additive identity.

Proposition 7.5. For all $r \in R$,

$$r \cdot 0_R = 0_R \cdot r = 0_R.$$

If you see that and ask, "Isn't that obvious?" then you *really* need to read the proof. While you read it, ask yourself, "What properties of a ring make this statement true?" The answer to that question will indicate your hidden assumptions. Try to prove the proposition without those properties, and you will see why it is *not* in fact obvious.

Proof. Let $r \in R$. Since $(R,+)$ is an abelian group, we know that $0_R + 0_R = 0_R$. By substitution, $r\,(0_R + 0_R) = r \cdot 0_R$. By distribution, $r \cdot 0_R + r \cdot 0_R = r \cdot 0_R$. Since $(R,+)$ is an abelian group, $r \cdot 0_R$ has an additive inverse; call it s. Applying the properties of a ring, we have

$$\begin{aligned} s + (r \cdot 0_R + r \cdot 0_R) &= s + r \cdot 0_R && \text{(substitution)} \\ (s + r \cdot 0_R) + r \cdot 0_R &= s + r \cdot 0_R && \text{(associative)} \\ 0_R + r \cdot 0_R &= 0_R && \text{(additive inverse)} \\ r \cdot 0_R &= 0_R. && \text{(additive identity)} \end{aligned}$$

A similar argument shows that $0_R \cdot r = 0_R$. □

We now turn our attention to two properties that, while pleasant, are not necessary for a ring.

Definition 7.6. Let R be a ring. If R has a multiplicative identity 1_R such that

$$r \cdot 1_R = 1_R \cdot r = r \qquad \forall r \in R,$$

we say that R is a **ring with unity**. (Another name for the multiplicative identity is **unity**.)

If R is a ring and the multiplicative operation is commutative, so that

$$rs = sr \quad \forall r \in R,$$

then we say that R is a **commutative ring**.

A ring with unity is

- an abelian group under multiplication, and
- a (possibly commutative) monoid under addition.

Example 7.7. The set of matrices $\mathbb{R}^{m\times m}$ is a ring with unity, where I_m is the multiplicative identity. However, it is not a commutative ring.

You will show in Exercise 7.13 that $2\mathbb{Z}$ is a ring. It is a commutative ring, but not a ring with unity.

For a commutative ring with unity, consider $\mathbb{Z}$.

Remark 7.8. While non-commutative rings are interesting,

Unless we state otherwise,
all rings in these notes are commutative.

As with groups, we can characterize all rings with only two elements.

Example 7.9. Let R be a ring with only two elements. There are two possible structures for R.

Why? Since $(R,+)$ is an abelian group, by Example 2.9 on page 60 the addition table of R has the form

$+$	0_R	a
0_R	0_R	a
a	a	0_R

.

By Proposition 7.5, we know that the multiplication table *must* have the form

$\times$	0_R	a
0_R	0_R	0_R
a	0_R	?

where $a \cdot a$ is undetermined. Nothing in the properties of a ring tell us whether $a \cdot a = 0_R$ or $a \cdot a = a$; in fact, rings exist with both properties:

- if $R = \mathbb{Z}_2$ (see Exercise 7.14 to see that this is a ring), then $a = [1]$ and $a \cdot a = a$; but
- if

$$R = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, a = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right\} \subsetneq (\mathbb{Z}_2)^{2\times 2},$$

  then $a \cdot a = 0 \neq a$.

Just as groups have subgroups, rings have subrings:

Definition 7.10. Let R be a ring, and S a nonempty subset of R. If S is also a ring under the same operations as R, then S is a subring of R.

Example 7.11. Recall from Exercise 7.13 that $2\mathbb{Z}$ is a ring; since $2\mathbb{Z} \subsetneq \mathbb{Z}$, it is a subring of $\mathbb{Z}$.

To show that a subset of a ring is a subring, do we have to show all four ring properties? No: as with subgroups, we can simplify the characterization to two properties:

Theorem 7.12 (The Subring Theorem)**.** Let R be a ring and S be a nonempty subset of R. The following are equivalent:

(A) S is a subring of R.

(B) S is closed under subtraction and multiplication: for all $a, b \in S$
  (S1) $a - b \in S$, and
  (S2) $ab \in S$.

Proof. That (A) implies (B) is clear, so assume (B). From (B) we know that for any $a, b \in S$ we have (S1) and (S2). As (S1) is essentially the Subgroup Theorem, S is an additive subgroup of the additive group R. On the other hand, (S2) only tells us that S satisfies property $(R2)$ of a ring, but any elements of S are elements of R, so the associative and distributive properties follow from inheritance. Thus S is a ring in its own right, which makes it a subring of R. □

Exercises

Exercise 7.13.

(a) Show that $2\mathbb{Z}$ is a ring under the usual addition and multiplication of integers.

(b) Show that for any $n \in \mathbb{Z}$, $n\mathbb{Z}$ is a ring under the usual addition and multiplication of integers.

Exercise 7.14. Recall the definition of multiplication for $\mathbb{Z}_n$ from Section 6.4: for $[a], [b] \in \mathbb{Z}_n$, $[a][b] = [ab]$.

(a) Show that $\mathbb{Z}_2$ is a ring under the addition and multiplication of cosets defined in Section 3.5.

(b) Show that for any $n \in \mathbb{N}^+$ where $n > 1$, $\mathbb{Z}_n$ is a ring under the addition and multiplication of cosets defined in Section 3.5.

(c) Show that there exist a, b, n such that $[a]_n [b_n] = [0]_n$ but $[a]_n, [b]_n \neq [0]_n$.

Exercise 7.15. Let R be a ring.

(a) Show that for all $r, s \in R$, $(-r)s = r(-s) = -(rs)$.

(b) Suppose that R has unity. Show that $-r = -1_R \cdot r$ for all $r \in R$.

Exercise 7.16. Let R be a ring with unity. Show that $1_R = 0_R$ if and only if R has only one element.

Exercise 7.17. Consider the two possible ring structures from Example 7.9. Show that if a ring R has only two elements, one of which is unity, then it can have only one of the structures.

Exercise 7.18. Let $R = \{T, F\}$ with the additive operation $\oplus$ (Boolean xor) and a multiplicative operation $\wedge$ (Boolean and where

$$\begin{aligned} F \oplus F &= F & F \wedge F &= F \\ F \oplus T &= T & F \wedge T &= F \\ T \oplus F &= T & T \wedge F &= F \\ T \oplus T &= F & T \wedge T &= T. \end{aligned}$$

(See also Exercises 2.20 and 2.21 on page 64.) Is $(R, \oplus, \wedge)$ a ring? If it is a ring, then

(a) what is the zero element?

(b) does it have a unity element? if so, what is it?

(c) is it commutative?

Exercise 7.19. Let R and S be rings, with $R \subseteq S$ and $\alpha \in S$. The **extension of R by α** is

$$R[\alpha] = \{r_n \alpha^n + \cdots + r_1 \alpha + r_0 : n \in \mathbb{N},\ r_0, r_1, \ldots, r_n \in R\}.$$

(a) Show that $R[\alpha]$ is also a ring.

(b) Suppose $R = \mathbb{Z}$, $S = \mathbb{C}$, and $\alpha = \sqrt{-5}$.

(i) Explain why every element of $R[\alpha]$ can be written in the form $a + b\alpha$.

(ii) Show that 6 can be factored two distinct ways in $R[\alpha]$: one is the ordinary factorization in $R = \mathbb{Z}$, while the other exploits the difference of squares with $\alpha = \sqrt{-5}$.

Exercise 7.20. In Exercise 7.14, you showed that $\mathbb{Z}_n$ is a ring. A nonzero element r of a ring R is **nilpotent** if we can find $n \in \mathbb{N}^+$ such that $r^n = 0_R$.

(a) Identify the nilpotent elements, *if any*, of $\mathbb{Z}_n$ for $n = 2, 3, 4, 5, 6$. If not, state that.

(b) Do you think there is a relationship between n and the nilpotents of $\mathbb{Z}_n$? If so, state it.

7.2: Integral Domains and Fields

In this section, R is always a commutative ring with unity.

Example 7.4 illustrates an important point: not all rings satisfy properties that we might like to take for granted. Not only does it show that not all rings possess the zero product property, it also demonstrates that multiplicative inverses do not necessarily exist in all rings. Both multiplicative inverses and the zero product property are very useful; we use them routinely to solve equations! Rings with these properties deserve special attention.

Two convenient kinds of rings

We first classify rings that satisfy the zero product property.

Definition 7.21. If the elements of R satisfy the zero product property, then we call R an **integral domain**.

We use the word "integral" here because R is like the ring of "integ"ers, $\mathbb{Z}$. We do *not* mean that you can compute the integrals of calculus.

Whenever R is not an integral domain, we can find two elements of R that *do not* satisfy the zero product property; that is, we can find nonzero $a, b \in R$ such that $ab = 0_R$. Recall that we used a special term for this phenomenon in the group $\mathbb{Z}_n^*$, **zero divisors** (Section 6.4). The ideas are identical, so the term is appropriate, and we will call a and b **zero divisors** in a ring, as well.

Example 7.22. As you might expect, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are integral domains.

In Exercise 7.14, you showed that $\mathbb{Z}_n$ was a ring under clockwork addition and multiplication. However, it need not be an integral domain. For example, in $\mathbb{Z}_6$ we have $[2] \cdot [3] = [6] = [0]$, making $[2]$ and $[3]$ zero divisors. On the other hand, it isn't hard to see that $\mathbb{Z}_2$, $\mathbb{Z}_3$, and $\mathbb{Z}_5$ are integral domains, if only via an exhaustive check. What about $\mathbb{Z}_4$? We leave that, and all of $\mathbb{Z}_n$ to the exercises.

Next, we turn to multiplicative inverses.

Definition 7.23. If every non-zero element of R has a multiplicative inverse, then we call R a **field**.

Example 7.24. The rings $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields, while $\mathbb{Z}$ is not.

What about $\mathbb{Z}_n$ and $\mathbb{Z}_n^*$? Again, we leave those to the exercises. For now, we need to notice an important relationship between fields and integral domains.

The examples show that some integral domains are not fields, but all the fields we've listed are also integral domains. It would be great if this turned out to be true in general: that is, if every field is an integral domain. Determining the relationships between different classes of rings, and remembering which class you're working with, is a crucial point of ring theory.

Theorem 7.25. Every field is an integral domain.

Proof. Let $\mathbb{F}$ be a field. We claim that $\mathbb{F}$ is an integral domain: that is, the elements of $\mathbb{F}$ satisfy the zero product property. Let $a, b \in \mathbb{F}$ and assume that $ab = 0$. We need to show that $a = 0$ or

$b = 0$. If $a = 0$, we're done, so assume that $a \neq 0$. Since $\mathbb{F}$ is a field, a has a multiplicative inverse. Apply Proposition 7.5 to obtain

$$b = 1 \cdot b = \left(a^{-1}a\right) b = a^{-1}(ab) = a^{-1} \cdot 0 = 0.$$

Hence $b = 0$.

We had assumed that $ab = 0$ and $a \neq 0$. By concluding that $b = 0$, the fact that a and b are arbitrary show that $\mathbb{F}$ is an integral domain. Since $\mathbb{F}$ is an arbitrary field, every field is an integral domain. □

Not every integral domain is a field, however. The most straightforward example is $\mathbb{Z}$.

The field of fractions

Speaking of $\mathbb{Q}$, it happens to be the smallest field that contains $\mathbb{Z}$, an integral domain. So there's another interesting question: can we form a field from any ring R, simply by adding fractions?

No, of course not — we just saw that a field must be an integral domain, and some rings are not integral domains. Even if you add fractions, the zero divisors remain, so you cannot have a field. So, then, can we form a field from any integral domain in the same way that we form $\mathbb{Q}$ from $\mathbb{Z}$? We need some precision in this discussion, which requires a definition.

Definition 7.26. Let R be an arbitrary ring. **The set of fractions over** a ring R is

$$\text{Frac}(R) := \left\{ \frac{p}{q} : \ p, q \in R \text{ and } q \neq 0 \right\},$$

with addition and multiplication defined in the usual way for "fractions", and equality defined by

$$\frac{a}{b} = \frac{p}{q} \iff aq = bp.$$

The answer to our question turns out to be yes!

Theorem 7.27. If R is an integral domain, then $\text{Frac}(R)$ is a ring.

To prove Theorem 7.27, we need two useful properties of fractions that you should be able to prove yourself.

Proposition 7.28. Let R be a ring, $a, b, r \in R$. If $br \neq 0$, then in $\text{Frac}(R)$

- $\dfrac{a}{b} = \dfrac{ar}{br}$, and
- $\dfrac{0_R}{a} = \dfrac{0_R}{b}$.

Proof. You do it! See Exercise 7.33. □

Watch for these properties in what follows.

Proof of Theorem 7.27. Assume that R is an integral domain. First we show that $\mathrm{Frac}(R)$ is an additive group. Let $f, g, h \in R$; choose $a, b, p, q, r, s \in \mathrm{Frac}(R)$ such that $f = a/b$, $g = p/q$, and $h = r/s$. First we show that $\mathrm{Frac}(R)$ is an abelian group.

closure: This is fairly routine, using common denominators. Since R is a domain and $b, q \neq 0$, we know that $bq \neq 0$. Thus,

$$\begin{aligned} f + g &= \frac{a}{b} + \frac{p}{q} && \text{(substitution)} \\ &= \frac{aq}{bq} + \frac{bp}{bq} && \text{(Proposition 7.28)} \\ &= \frac{aq + bp}{bq} && \text{(definition of addition in } \mathrm{Frac}(R)\text{)} \\ &\in \mathrm{Frac}(R). \end{aligned}$$

Why did we need R do be an integral domain? If not, then it is possible that $bq = 0$, and if so, $f + g \notin \mathrm{Frac}(R)$!

associative: This is the hardest one; watch for Proposition 7.28 to show up in many places. As before, since R is a domain and $b, q, s \neq 0$, we know that bq, $(bq)s$, $b(qs)$, and qs are all non-zero. Thus,

$$\begin{aligned} (f + g) + h &= \frac{aq + bp}{bq} + \frac{r}{s} \\ &= \frac{(aq + bp)s}{(bq)s} + \frac{(bq)r}{(bq)s} \\ &= \frac{((aq)s + (bp)s) + (bq)r}{(bq)s} \\ &= \frac{a(qs) + (b(ps) + b(qr))}{b(qs)} \\ &= \frac{a(qs)}{b(qs)} + \frac{b(ps) + b(qr)}{b(qs)} \\ &= \frac{a}{b} + \frac{ps + qr}{qs} \\ &= \frac{a}{b} + \left(\frac{p}{q} + \frac{r}{s}\right) \\ &= f + (g + h.) \end{aligned}$$

identity: We claim that the additive identity of $\mathrm{Frac}(R)$ is $0_R/1_R$. This is easy to see, since

$$f + \frac{0_R}{1_R} = \frac{a}{b} + \frac{0_R \cdot b}{1_R \cdot b} = \frac{a}{b} + \frac{0_R}{b} = \frac{a}{b} = f.$$

additive inverse: For each $f = p/q$, we claim that $(-p)/q$ is the additive inverse. This is easy

to see, but a little tedious. It is straightforward enough that,

$$f + \frac{-p}{q} = \frac{p}{q} + \frac{-p}{q} = \frac{(p + (-p))}{q} = \frac{0_R}{q}.$$

Don't conclude too quickly that we are done! We have to show that $f + (-p)/q = 0_{\mathrm{Frac}(R)}$, which is $0_R/1_R$. By Proposition 7.28, $0_R/1_R = 0_R/q_R$, so we did in fact compute the identity.

commutative: Using the fact that R is commutative, we have

$$\begin{aligned} f + g &= \frac{a}{b} + \frac{c}{d} = \frac{ad}{bd} + \frac{bc}{bd} \\ &= \frac{ad + bc}{bd} = \frac{cb + da}{db} \\ &= \frac{cb}{db} + \frac{da}{db} = \frac{c}{d} + \frac{a}{b} \\ &= g + f. \end{aligned}$$

Next we have to show that $\mathrm{Frac}(R)$ satisfies the requirements of a ring.

closure: Using closure in R and the fact that R is an integral domain, this is straightforward: $fg = (ap)/(bq) \in \mathrm{Frac}(R)$.

associative: Using the associative property of R, this is straightforward:

$$\begin{aligned} (fg)h &= \left(\frac{ap}{bq}\right)\frac{r}{s} = \frac{(ap)r}{(bq)s} = \frac{a(pr)}{b(qs)} \\ &= \frac{a}{b}\frac{(pr)}{qs} = f(gh). \end{aligned}$$

distributive: We rely on the distributive property of R:

$$\begin{aligned} f(g+h) &= \frac{a}{b}\left(\frac{p}{q} + \frac{r}{s}\right) = \frac{a}{b}\left(\frac{ps + qr}{qs}\right) \\ &= \frac{a(ps + qr)}{b(qs)} = \frac{a(ps) + a(qr)}{b(qs)} \\ &= \frac{a(ps)}{b(qs)} + \frac{a(qr)}{b(qs)} = \frac{ap}{bq} + \frac{ar}{bs} \\ &= fg + fh. \end{aligned}$$

Finally, we show that $\mathrm{Frac}(R)$ is a field. We have to show that it is commutative, that it has a multiplicative identity, and that every non-zero element has a multiplicative inverse.

commutative: We claim that the multiplication of $\mathrm{Frac}(R)$ is commutative. This follows from the fact that R, as an integral domain, has a commutative multiplication, so

$$\begin{aligned} fg &= \frac{a}{b} \cdot \frac{p}{q} = \frac{ap}{bq} = \frac{pa}{qb} \\ &= \frac{p}{q} \cdot \frac{a}{b} = gf. \end{aligned}$$

multiplicative identity: We claim that $\frac{1_R}{1_R}$ is a multiplicative identity for $\mathrm{Frac}(R)$. In fact,

$$f \cdot \frac{1_R}{1_R} = \frac{a}{b} \cdot \frac{1_R}{1_R} = \frac{a \cdot 1_R}{b \cdot 1_R} = \frac{a}{b} = f.$$

multiplicative inverse: Let $f \in \mathrm{Frac}(R)$ be a non-zero element. Let $a, b \in R$ such that $f = a/b$ and $a \neq 0$. Let $g = b/a$; then

$$fg = \frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ab}.$$

By Proposition 7.28

$$\frac{ab}{ab} = \frac{1_R}{1_R},$$

which we just showed to be the identity of $\mathrm{Frac}(R)$.

□

Definition 7.29. For any integral domain R, we call $\mathrm{Frac}(R)$ the **field of fractions of** R.

Exercises.

Exercise 7.30. Explain why $n\mathbb{Z}$ is not always an integral domain. For what values of n is it an integral domain?

Exercise 7.31. Show that $\mathbb{Z}_n$ is an integral domain if and only if n is irreducible. Is it also a field in these cases?

Exercise 7.32. You might think from Exercise 7.31 that we can turn $\mathbb{Z}_n$ into a field, or at least an integral domain, in the same way that we turned $\mathbb{Z}_n$ into a multiplicative group: that is, working with $\mathbb{Z}_n^*$. Explain that this doesn't work in general, because $\mathbb{Z}_n^*$ isn't even a ring.

Exercise 7.33. Show that if R is an integral domain, then the set of fractions has the following properties for any nonzero $a, b, c \in R$:

$$\frac{ac}{bc} = \frac{ca}{cb} = \frac{a}{b}, \qquad \frac{0_R}{a} = \frac{0_R}{1} = 0_{\mathrm{Frac}(R)},$$
$$\text{and} \qquad \frac{a}{a} = \frac{1_R}{1_R} = 1_{\mathrm{Frac}(R)}.$$

Exercise 7.34. To see concretely why $\mathrm{Frac}(R)$ is not a field if R is not a domain, consider $R = \mathbb{Z}_4$. Find nonzero $b, q \in R$ such that $bq = 0$, using them to find $f, g \in \mathrm{Frac}(R)$ such that $fg \notin \mathrm{Frac}(R)$.

7.3: Polynomial rings

When the average man on the street thinks of "algebra", he typically thinks not of "monoids", "groups", or "rings", but of "polynomials". Polynomials are certainly the focus of high school algebra, and they are also a major focus of higher algebra. The last few chapters of these notes are dedicated to the classical applications of the structural theory to important problems about polynomials.

While one can talk of a monoid or group of polynomials under addition, it is more natural to talk about a ring of polynomials under addition and multiplication. Polynomials helped motivate the distinction between the "two operations" of groups, which we decided was really two sides of one coin, and the "two operations" of rings, which really can be quite different operations. Polynomials provide great examples for the remaining topics. It is time to give them a good, hard look.

Some of the following may seem pedantic and needlessly detailed, and there's some truth to that, but it is important to fix these terms now to avoid confusion later. The difference between a "monomial" and a "term" is of special note; some authors reverse the notions. Similarly, pay attention to the notion of the support $\mathcal{T}_f$ of a polynomial f.

As usual, R is a ring.

Fundamental notions

Definition 7.35. An **indeterminate over** R is a symbol that represents an unknown value of R. A **constant of** R is a symbol that represents a fixed value of R. An **variable over** R is an indeterminate whose value is *not* fixed.

Notice that a constant can be indeterminate, as in the usual use of letters like a, b, and c, or quite explicitly determined, as in 1_R, 0_R, and so forth. Variables are always indeterminate. The main difference is that a constant is *fixed*, while a variable is not.

Definition 7.36. A **monomial over** R is a finite product of variables over R.

The use of "monomial" here is meant to be both consistent with its definition in Section 1.1, and with our needs for future work. Typically, though, we refer simply to "a monomial" rather than "a monomial over R".

By referring to "variables", the definition of a monomial explicitly excludes constants. Even though a^2 looks like a monomial, if a is a constant, we do not consider it a monomial; from our point of view, it is a constant.

Definition 7.37. The **total degree** of a monomial is the number of factors in the product. We say that two monomials are **like monomials** if they have the same variables, and corresponding variables have the same exponents.

A **term of** R is a constant, *or* the product of a monomial over R and a constant of R. The constant in a term is called the **coefficient** of the term. Two terms are **like terms** if their monomials are like monomials.

Now we define *polynomials*.

Definition 7.38. A **polynomial over** R is a finite sum of terms of R. We can write a generic polynomial f as $f = a_1t_1 + a_2t_2 + \cdots + a_mt_m$ where each $a_i \in R$ and each t_i is a monomial.

We call the set of monomials of f with non-zero coefficient its **support**. If we denote the support of f by $\mathcal{T}_f$, then we can write f as

$$f \quad = \quad \sum_{i=1,\ldots,\#\mathcal{T}_f} a_i t_i \quad = \quad \sum_{t\in\mathcal{T}_f} a_t t.$$

We call R the **ground ring** of each polynomial.

We say that two polynomials f and g are equal if $\mathcal{T}_f = \mathcal{T}_g$ and the coefficients of corresponding monomials are equal.

Notation 7.39. We adopt a convention that $\mathcal{T}_f$ is the support of a polynomial f.

Definition 7.40. $R[x]$ is the set of **univariate** polynomials in the variable x over R. That is, $f \in R[x]$ if and only if there exist $m \in \mathbb{N}$ and $a_m, a_{m-1}, \ldots, a_1 \in R$ such that

$$f(x) = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0.$$

The set $R[x,y]$ is the set of **bivariate** polynomials in the variables x and y whose coefficients are in R.

For $n \geq 2$, the set $R[x_1, x_2, \ldots, x_n]$ is the set of **multivariate** polynomials in the variables $x_1, x_2, \ldots, x_n$ whose coefficients are in R.

The **degree** of a univariate polynomial f, written $\deg f$, is the largest of the total degrees of the monomials of f. We write $\mathrm{lm}(f)$ for the monomial of f with that degree, and $\mathrm{lc}(f)$ for its coefficient. Unless we say otherwise, the degree of a multivariate polynomial is undefined.

Example 7.41. Definition 7.40 tells us that $\mathbb{Z}_6[x,y]$ is the set of bivariate polynomials in x and y whose coefficients are in $\mathbb{Z}_6$. For example,

$$f(x,y) = 5x^3 + 2x \in \mathbb{Z}_6[x,y]$$

and

$$g(x,y) = x^2y^2 - 2x^3 + 4 \in \mathbb{Z}_6[x,y].$$

The ground ring for both f and g is $\mathbb{Z}_6$. Observe that f can be considered a univariate polynomial, in which case $\deg f = 3$.

We also consider constants to be polynomials of degree 0; thus $4 \in \mathbb{Z}_6[x,y]$ and even $0 \in \mathbb{Z}_6[x,y]$.

It is natural to think of a constant as a polynomial. This leads to some unexpected, but interesting and important consequences.

Definition 7.42. Let $f \in R[x_1,\ldots,x_n]$.

We say that f is a **constant polynomial** if $\mathcal{T}_f = \{1\}$ or $\mathcal{T}_f = \emptyset$; in other words, all the non-constant terms have coefficient zero.

We say that f is a **vanishing polynomial** if for all $r_1,\ldots,r_n \in R$, $f(r_1,\ldots,r_n) = 0$. We will see that this can happen even if $f \neq 0_R$.

The definition of vanishing and constant polynomials implies that 0_R satisfies both. However, the definition of equality means that vanishing polynomials need not be zero polynomials!

Example 7.43. Let $f(x) = x^2 + x \in \mathbb{Z}_2[x]$. Since $\mathcal{T}_f \neq \emptyset$, $f \neq 0_R$. However,

$$\begin{aligned} f(0) &= 0^2 + 0 && \text{and} \\ f(1) &= 1^2 + 1 = 0 && (\text{in } \mathbb{Z}_2!). \end{aligned}$$

Here f is a vanishing polynomial *even though it is not zero.*

Properties of polynomials

We can now turn our attention to the properties of $R[x]$ and $R[x_1,\ldots,x_n]$. First up is a question raised by Example 7.43: when must a vanishing polynomial be the constant polynomial 0?

Proposition 7.44. If R is a non-zero integral domain, then the following are equivalent.

(A) 0 is the only vanishing polynomial in $R[x_1,\ldots,x_n]$.

(B) R has infinitely many elements.

As is often the case, we can't answer that question immediately. Before proving Proposition 7.44, we need the following, extraordinary theorem.

Theorem 7.45 (The Factor Theorem)**.** If R is a non-zero integral domain, $f \in R[x]$, and $a \in R$, then $f(a) = 0$ if and only if $x - a$ divides $f(x)$.

To prove Theorem 7.45, we need to make precise our notions of addition and multiplication of polynomials.

Definition 7.46. To **add** two polynomials $f, g \in R[x_1, \ldots, x_n]$, let $\mathcal{T} = \mathcal{T}_f \cup \mathcal{T}_g$. Choose $a_t, b_t \in R$ such that

$$f = \sum_{t \in \mathcal{T}} a_t t \quad \text{and} \quad g = \sum_{t \in \mathcal{T}} b_t t.$$

We add the polynomials by adding like terms; that is,

$$f + g = \sum_{t \in \mathcal{T}} (a_t + b_t)\, t.$$

To **multiply** f and g, compute the sum of all products of terms in the first polynomial with terms in the second; that is,

$$fg = \sum_{t \in \mathcal{T}} \sum_{u \in \mathcal{T}} (a_t b_u)\, (tu).$$

We use u in the second summand to distinguishes the terms of g from those of f. Notice that fg is really the distribution of g to the terms of f, followed by the distribution of each term of f to the terms of g.

Proof of the Factor Theorem. If $x - a$ divides $f(x)$, then there exists $q \in R[x]$ such that $f(x) = (x-a) \cdot q(x)$. By substitution, $f(a) = (a-a) \cdot q(a) = 0_R \cdot q(a) = 0_R$.

Conversely, assume $f(a) = 0$. You will show in Exercise 7.49 that we can write $f(x) = q(x) \cdot (x-a) + r$ for some $r \in R$. Thus

$$0 = f(a) = q(a) \cdot (a-a) + r = r,$$

and substitution yields $f(x) = q(x) \cdot (x-a)$. In other words, $x - a$ divides $f(x)$, as claimed. □

We now turn our attention to proving Proposition 7.44.

Proof of Lemma 7.44. Assume that R is a non-zero integral domain.

(A) ⇒ (B): We proceed by the contrapositive. Assume that R has finitely many elements. We can enumerate them all as $r_1, r_2, \ldots, r_m$. Let

$$f(x_1, \ldots, x_n) = (x_1 - r_1)(x_1 - r_2) \cdots (x_1 - r_m).$$

Let $b_1, \ldots, b_n \in R$. By assumption, R is finite, so $b_1 = r_i$ for some $i \in \{1, 2, \ldots, m\}$. Notice that f is not only multivariate, it is also univariate: $f \in R[x_i]$. By the Factor Theorem, $f = 0$. We have shown that ¬(B) implies ¬(A); thus, (A) implies (B).

(A) ⇐ (B): Assume that R has infinitely many elements. Let f be any vanishing polynomial. We proceed by induction on n, the number of variables in $R[x_1, \ldots, x_n]$.

Inductive base: Suppose $n = 1$. By the Factor Theorem, $x - a$ divides f for every $a \in R$. By definition of polynomial multiplication, each distinct factor of f adds 1 to the degree of f; for

example, if $f = (x-0)(x-1)$, then $\deg f = 2$. However, the definition of a polynomial implies that f has finite degree. Hence, if $f \neq 0$, then it can be factored as only finitely many polynomials of the form $x-a$. If so, then choose $a_1, a_2, \ldots, a_n$ such that

$$f = (x-a_1)(x-a_2)\cdots(x-a_n).$$

Since R has infinitely many elements, we can find $b \in R$ such that $b \neq a_1, \ldots, a_n$. That means $b - a_i \neq 0$ for each $i = 1, \ldots, n$. As R is an integral domain,

$$f(b) = (b-a_1)(b-a_2)\cdots(b-a_n) \neq 0.$$

This contradicts the choice of f as a vanishing polynomial. Hence, $f = 0$.

Inductive hypothesis: Assume for all i satisfying $1 \leq i < n$, if $f \in R[x_1, \ldots, x_i]$ is a zero polynomial, then f is the constant polynomial 0.

Inductive step: Let $n > 1$, and $f \in R[x_1, \ldots, x_n]$ be a vanishing polynomial. Let $a_n \in R$, and substitute $x_n = a_n$ into f. Denote the resulting polynomial as g. The substitution means that $x_n \notin \mathcal{T}_g$. Hence, $g \in R\left[x_1, \ldots, x_{n-1}\right]$.

It turns out that g is also a vanishing polynomial in $R\left[x_1, \ldots, x_{n-1}\right]$. *Why?* By way of contradiction, assume that it is not. Then there exist $a_1, \ldots, a_{n-1} \in R$ such that $f\left(a_1, \ldots, a_{n-1}\right) \neq 0$. However, the definition of g implies that

$$f(a_1, \ldots, a_n) = g\left(a_1, \ldots, a_{n-1}\right) \neq 0.$$

This contradicts the choice of f as a vanishing polynomial. The assumption was wrong; g must be a vanishing polynomial in $R\left[x_1, \ldots, x_{n-1}\right]$, after all. We can now apply the inductive hypothesis, and infer that g is the constant polynomial 0.

We chose a_n arbitrarily, so this argument holds for any $a_n \in R$. Thus, any of the terms of f containing any of the variables $x_1, \ldots, x_{n-1}$ has a coefficient of zero. The only non-zero terms are those whose only variables are x_n, so $f \in R[x_n]$. This time, the inductive base implies that f is zero. □

We come to the main purpose of this section.

Theorem 7.47. The univariate and multivariate polynomial rings over a ring R are themselves rings.

Proof. Let $n \in \mathbb{N}^+$ and R a ring. We claim that $R[x_1, \ldots, x_n]$ is a ring. To consider the requirements of a ring, let f, g, $h \in R[x_1, \ldots, x_n]$, and let $\mathcal{T} = \mathcal{T}_f \cup \mathcal{T}_g \cup \mathcal{T}_h$. For each $t \in \mathcal{T}$, choose $a_t, b_t, c_t \in R$ such that

$$f = \sum a_t t, \quad g = \sum b_t t, \quad h = \sum c_t t.$$

(Naturally, if $t \in \mathcal{T} \backslash \mathcal{T}_f$, then $a_t = 0$; if $t \in \mathcal{T} \backslash \mathcal{T}_g$, then $b_t = 0$, and if $t \in \mathcal{T} \backslash \mathcal{T}_h$, then $c_t = 0$.) Although we do not write it, all the sums below are indexed over $t \in \mathcal{T}$.

(R1) First we show that $R[x_1, \ldots, x_n]$ is an abelian group.

(closure) By the definition of polynomial addition,

$$(f+g)(x) = \sum (a_t + b_t)\, t.$$

Since R is closed under addition, we conclude that $f+g \in R[x_1,\ldots,x_n]$.

(associative) We rely on the associativity of R:

$$\begin{aligned} f+(g+h) &= \sum a_t t + \left(\sum b_t t + \sum c_t t\right) \\ &= \sum a_t t + \sum (b_t + c_t)\, t \\ &= \sum [a_t + (b_t + c_t)]\, t \\ &= \sum [(a_t + b_t) + c_t]\, t \\ &= \sum (a_t + b_t)\, t + \sum_{t \in T} c_t t \\ &= \left(\sum a_t t + \sum b_t t\right) + \sum c_t t \\ &= (f+g)+h. \end{aligned}$$

(identity) We claim that the constant polynomial 0 is the identity. Recall that 0 is a polynomial whose coefficients are all 0. We have

$$\begin{aligned} f+0 &= \sum a_t t + 0 \\ &= \sum a_t t + \sum 0 \cdot t \\ &= \sum (a_t + 0)\, t \\ &= f. \end{aligned}$$

(inverse) Let $p = \sum (-a_t)\, t$. We claim that p is the additive inverse of f. In fact,

$$\begin{aligned} p+f &= \sum (-a_t)\, t + \sum a_t t \\ &= \sum (-a_t + a_t)\, t \\ &= \sum 0 \cdot t \\ &= 0. \end{aligned}$$

(commutative) By the definition of polynomial addition, $g+f = \sum (b_t + a_t)\, t$. Since R is commutative under addition, addition of coefficients is commutative, so

$$\begin{aligned} f+g &= \sum a_t t + \sum b_t t \\ &= \sum (a_t + b_t)\, t \\ &= \sum (b_t + a_t)\, t \\ &= \sum b_t t + \sum a_t t \\ &= g+f. \end{aligned}$$

(R2) Next, we show that $R[x_1,\ldots,x_n]$ is a semigroup.

(closed) Applying the definition of polynomial multiplication, we have

$$f\,g = \sum_{t\in T}\left[\sum_{u\in T}\left(a_t b_u\right)\left(t\,u\right)\right].$$

Since R is closed under multiplication, each $\left(a_t b_u\right)\left(t\,u\right)$ is a term. Thus $f\,g$ is a sum of sums of terms, or a sum of terms. In other words, $f\,g \in R\left[x_1,\ldots,x_n\right]$.

(associative) We start by applying the product $f\,g$, then multiplying the result to h:

$$\begin{aligned}\left(f\,g\right)h &= \left[\sum_{t\in T}\left[\sum_{u\in T}\left(a_t b_u\right)\left(t\,u\right)\right]\right]\cdot\sum_{v\in T}c_v\,v\\ &= \sum_{t\in T}\left[\sum_{u\in T}\left[\sum_{v\in T}\left[\left(a_t b_u\right)c_v\right]\left[\left(t\,u\right)v\right]\right]\right].\end{aligned}$$

Now apply the associative property of multiplication in R:

$$\left(f\,g\right)h = \sum_{t\in T}\left[\sum_{u\in T}\left[\sum_{v\in T}\left[a_t\left(b_u c_v\right)\right]\left[t\left(u\,v\right)\right]\right]\right].$$

(Notice the associative property of R applies to terms over R, as well, inasmuch as those terms represent undetermined elements of R.) Now unapply the product:

$$\begin{aligned}\left(f\,g\right)h &= \sum_{t\in T}\left[\sum_{u\in T}\left[\sum_{v\in T}\left[a_t\left(b_u c_v\right)\right]\left[t\left(u\,v\right)\right]\right]\right]\\ &= \sum_{t\in T}a_t\,t\cdot\left[\sum_{u\in T}\left[\sum_{v\in T}\left(b_u c_v\right)\left(u\,v\right)\right]\right]\\ &= f\left(g\,h\right).\end{aligned}$$

(R3) To show the distributive property, first apply addition, then multiplication:

$$\begin{aligned}f\left(g+h\right) &= \sum_{t\in T}a_t\,t\cdot\left(\sum_{u\in T}b_u u+\sum_{u\in T}c_u u\right)\\ &= \sum_{t\in T}a_t\,t\cdot\sum_{u\in T}\left(b_u+c_u\right)u\\ &= \sum_{t\in T}\left[\sum_{u\in T}\left[a_t\left(b_u+c_u\right)\right]\left(t\,u\right)\right].\end{aligned}$$

Now apply the distributive property in the ring, and unapply the addition and multipli-

cation:

$$\begin{aligned} f(g+h) &= \sum_{t\in T}\left[\sum_{u\in T}(a_t b_u + a_t c_u)(tu)\right] \\ &= \sum_{t\in T}\left[\sum_{u\in T}[(a_t b_u)(tu) + (a_t c_u)(tu)]\right] \\ &= \sum_{t\in T}\left[\sum_{u\in T}(a_t b_u)(tu) + \sum_{u\in T}(a_t c_u)(tu)\right] \\ &= \sum_{t\in T}\left[\sum_{u\in T}(a_t b_u)(tu)\right] + \sum_{t\in T}\left[\sum_{u\in T}(a_t c_u)(tu)\right] \\ &= fg + fh. \end{aligned}$$

(commutative) Since we are working in commutative rings, we must also show that that $R[x_1,\ldots,x_n]$ is commutative. This follows from the commutativity of R:

$$\begin{aligned} fg &= \left(\sum_{t\in T} a_t t\right)\left(\sum_{u\in T} b_u u\right) \\ &= \sum_{t\in T}\sum_{u\in T}(a_t b_u)(tu) \\ &= \sum_{u\in T}\sum_{t\in T}(b_u a_t)(ut) \\ &= gf. \end{aligned}$$

(We can swap the sums because of the commutative and associative properties of addition.)

□

Exercises.

Exercise 7.48. Let $f(x) = x$ and $g(x) = x+1$ in $\mathbb{Z}_2[x]$.

(a) Show that f and g are not vanishing polynomials.
(b) Compute the polynomial $p = fg$.
(c) Show that $p(x)$ is a vanishing polynomial.
(d) Explain why this does *not* contradict Proposition 7.44.

Exercise 7.49. Fill in each blank of Figure 7.1 with the justification.

Exercise 7.50. Pick at random a degree 5 polynomial f in $\mathbb{Z}[x]$. Then pick at random some $a \in \mathbb{Z}$.

(a) Find $q \in \mathbb{Z}[x]$ and $r \in \mathbb{Z}$ such that $f(x) = q(x)\cdot(x-a) + r$.
(b) Explain why you *cannot* pick a nonzero integer b at random and expect willy-nilly to find $q \in \mathbb{Z}[x]$ and $r \in \mathbb{Z}$ such that $f(x) = q(x)\cdot(bx-a) + r$.

Let R be an integral domain, $f \in R[x]$, and $a \in R$.
Claim: There exist $q \in R[x]$ and $r \in R$ such that $f(x) = q(x) \cdot (x-a) + r$.
Proof:

1. Without loss of generality, we may assume that $\deg f = n$.
2. By _____, choose $a_1, \ldots, a_n$ such that $f = \sum_{k=1}^{n} a_k x^k$. We proceed by induction on n.
3. For the *inductive base,* assume that $n = 0$. Then $q(x) =$_____ and $r =$_____.
4. For the *inductive hypothesis,* assume that for all $i \in \mathbb{N}$ satisfying $0 \leq i < n$, there exist $q \in R[x]$ and $r \in R$ such that $f(x) = q(x) \cdot (x-a) + r$.
5. For the *inductive step,*
   (a) Let $p(x) = a_n x^{n-1}$, and $g(x) = f(x) - p(x) \cdot (x-a)$.
   (b) Notice that $\deg g <$_____.
   (c) By _____, there exist $p' \in R[x]$ and $r \in R$ such that $g(x) = p'(x) \cdot (x-a) + r$.
   (d) Let $q = p + p'$. By _____, $q \in R[x]$.
   (e) By _____ and _____, $f(x) = q(x) \cdot (x-a) + r$.
6. We have shown that, for arbitrary n, we can find $q \in R[x]$ and $r \in R$ such that $f(x) = q(x) \cdot (x-a) + r$. The claim holds.

Figure 7.1. Material for Exercise 7.49

(c) Explain why you *can* pick a nonzero integer b at random and expect willy-nilly to find $q \in \mathbb{Z}[x]$ and $r, s \in \mathbb{Z}$ such that $s \cdot f(x) = q(x) \cdot (bx-a) + r$. (Neat, huh?)
(d) If the requirements of (b) were changed to finding $q \in \mathbb{Q}[x]$ and $r \in \mathbb{Q}$, would you then be able to carry out (b)? Why or why not?

Exercise 7.51. Let $R = \mathbb{Z}_3[x]$ and $f(x) = x^3 + 2x + 1 \in R$.
(a) Explain how we can infer that f does not factor in R without performing a brute force search of factorizations.
(b) If we divide $g \in R$ by f, how many possible remainders can we obtain?

Exercise 7.52. Let R be an integral domain.
(a) Show that $R[x]$ is also an integral domain.
(b) How does this not contradict Exercise 7.48? After all, $\mathbb{Z}_2$ is a field, and thus an integral domain!

Exercise 7.53. Let R be a ring, and $f, g \in R[x]$. Show that $\deg(f+g) \leq \max(\deg f, \deg g)$.

Exercise 7.54. Let R be a ring and define

$$R(x) = \mathrm{Frac}(R[x]);$$

for example,

$$\mathbb{Z}(x) = \mathrm{Frac}(\mathbb{Z}[x]) = \left\{ \frac{p}{q} : \ p, q \in \mathbb{Z}[x] \right\}.$$

Is $R(x)$ a ring? is it a field?

Exercise 7.55. Let $R = \mathbb{Q}\left[\sqrt{2}\right]$, an extension of $\mathbb{Q}$ by $\sqrt{2}$. (See Exercise 7.19.)

(a) Find $g \in \mathbb{Q}[x]$ such that g factors with coefficients in R, but not with coefficients in $\mathbb{Q}$.
(b) Let $S = \mathbb{Q}\left[\sqrt{2}+\sqrt{3}\right]$ and $T = R\left[\sqrt{3}\right]$. Show that $S = T$.
(c) Is $\mathbb{Z}\left[\sqrt{2}+\sqrt{3}\right] = \mathbb{Z}\left[\sqrt{2}\right]\left[\sqrt{3}\right]$?

Exercise 7.56. Let $p \in \mathbb{Z}$ be irreducible, and $R = \mathbb{Z}_p[x]$. Show that $\varphi : R \to R$ by $\varphi(f) = f^p$ is a group automorphism. This is called the **Frobenius automorphism**.

7.4: Euclidean domains

In this section we consider an important similarity between the ring of integers and the ring of polynomials. This similarity will motivate us to define a new kind of ring. We will then show that all rings of this type allow us to perform important operations that we find both useful and necessary. What is the similarity? The ability to *divide with remainder*.

Division of polynomials

We start with polynomials, but we will take this a step higher in a moment.

Theorem 7.57 (The Division Theorem for polynomials)**.** Let $\mathbb{F}$ be a field, and consider the polynomial ring $\mathbb{F}[x]$. Let $f, g \in \mathbb{F}[x]$ with $f \neq 0$. There exist unique $q, r \in \mathbb{F}[x]$ satisfying (D1) and (D2) where
(D1) $g = qf + r$;
(D2) $r = 0$ or $\deg r < \deg f$.
We call g the **dividend**, f the **divisor**, q the **quotient**, and r the **remainder**.

Proof. The proof is essentially the procedure of long division of polynomials.

If $g = 0$, let $r = q = 0$. Then $g = qf + r$ and $r = 0$.

Now assume $g \neq 0$. If $\deg g < \deg f$, let $r = g$ and $q = 0$. Then $g = qf + r$ and $\deg r < \deg f$.

Otherwise, $\deg g \geq \deg f$. Let $m = \deg f$ and $n = \deg g - \deg f$. We proceed by induction on n.

For the *inductive base* $n = 0$, we have $\deg g = \deg f = m$. Let $a_m, \ldots, a_1, b_m, \ldots, b_1 \in R$ such that

$$g = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0$$
$$f = b_m x_m + b_{m-1} x^{m-1} + \cdots + b_1 x + b_0.$$

Let $q = \frac{a_m}{b_m}$ and $r = g - qf$. Since $\mathbb{F}$ is a field and $b_m \neq 0$, we can safely conclude that q is a constant polynomial. Arithmetic shows that $g = qf + r$, but can we guarantee that $r = 0$ or $\deg r < \deg f$? Apply substitution, distribution, and polynomial addition to obtain

$$\begin{aligned} r &= g - qf \\ &= \left(a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0\right) \\ &\quad - \frac{a_m}{b_m}\left(b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0\right) \\ &= \left(a_m - \frac{a_m}{b_m}\cdot b_m\right)x^m + \left(a_{m-1} - \frac{a_m}{b_m}\cdot b_{m-1}\right)x^{m-1} + \cdots + \left(a_0 - \frac{a_m}{b_m}\cdot b_0\right) \\ &= 0x^m + \left(a_{m-1} - \frac{a_m}{b_m}\cdot b_{m-1}\right)x^{m-1} + \cdots + \left(a_0 - \frac{a_m}{b_m}\cdot b_0\right). \end{aligned}$$

Since the coefficient of x^m is zero, we see that if $r \neq 0$, then $\deg r < \deg f$.

For the *inductive hypothesis*, assume that for all $i < n$ there exist $q, r \in R[x]$ such that $g = qf + r$ and $r = 0$ or $\deg r < \deg f$.

For the *inductive step*, let $\ell = \deg g$. Let $a_m, \ldots, a_0, b_\ell, \ldots, b_0 \in R$ such that

$$\begin{aligned} f &= a_m x^m + \cdots + a_0 \\ g &= b_\ell x^\ell + \cdots + b_0. \end{aligned}$$

Let $p = \frac{b_\ell}{a_m}\cdot x^n$ and $r = g - pf$. Once again, since $\mathbb{F}$ is a field and $a_m \neq 0$, we can safely conclude that $p \in \mathbb{F}[x]$. Apply substitution and distribution to obtain

$$\begin{aligned} g' &= g - pf \\ &= g - \frac{b_\ell}{a_m}\cdot x^n\left(a_m x^m + \cdots + a_0\right) \\ &= g - \left(b_\ell x^{m+n} + \frac{b_\ell a_{m-1}}{a_m}\cdot x^{m-1+n} + \cdots + \frac{b_\ell a_0}{a_m}\cdot x^n\right). \end{aligned}$$

Recall that $n = \deg g - \deg f = \ell - m$, so $\ell = m + n$. Apply substitution and polynomial addition to obtain

$$\begin{aligned} g' = g - pf &= \left(b_\ell x^\ell + \cdots + b_0\right) \\ &\quad - \left(b_\ell x^\ell + \frac{b_\ell a_{m-1}}{a_m}\cdot x^{\ell-1} + \cdots + \frac{b_\ell a_0}{a_m}\cdot x^n\right) \\ &= 0x^\ell + \left(b_{\ell-1} - \frac{b_\ell a_{m-1}}{a_m}\right)x^{\ell-1} \\ &\quad + \cdots + \left(b_n - \frac{b_\ell a_0}{a_m}\right)x^n + b_{n-1}x^{n-1}\cdots + b_0. \end{aligned}$$

Since $\mathbb{F}$ is a field and $a_m \neq 0$, we can safely conclude that $g' \in \mathbb{F}[x]$. Observe that $\deg g' < \ell = \deg g$, so $\deg g' - \deg f < n$. Apply the inductive hypothesis to find $p', r \in R[x]$ such that

$g' = p'f + r$ and $r = 0$ or $\deg r < \deg f$. Then

$$\begin{aligned} g &= pf + g' = pf + (p'f + r) \\ &= (p + p')\, f + r. \end{aligned}$$

Let $q = p + p'$. By closure, $q \in R[x]$, and we have shown the existence of a quotient and remainder.

For uniqueness, assume that there exist $q_1, q_2, r_1, r_2 \in R[x]$ such that $g = q_1 f + r_1 = q_2 f + r_2$ and $\deg r_1, \deg r_2 < \deg f$. Then

$$\begin{aligned} q_1 f + r_1 &= q_2 f + r_2 \\ 0 &= (q_2 - q_1)\, f + (r_2 - r_1). \end{aligned} \tag{26}$$

If $q_2 - q_1 \neq 0$, then no term of $(q_2 - q_1)\,\mathrm{lm}(f)$ has degree smaller than $\deg f$. Since every term of $r_2 - r_1$ has degree smaller than $\deg f$, there are no like terms between the two. Thus, there can be no cancellation between $(q_2 - q_1)\,\mathrm{lm}(f)$ and $r_2 - r_1$, and for similar reasons there can be no cancellation between $(q_2 - q_1)\,\mathrm{lm}(f)$ and lower-degree terms of $(q_2 - q_1) f$. However, the left hand side of equation 26 is the zero polynomials, so coefficients of $(q_2 - q_1)\,\mathrm{lm}(f)$ are all 0 on the left hand side. They must likewise be all zero on the right hand side. That implies $(q_2 - q_1)\,\mathrm{lm}(f)$ is equal to the constant polynomial 0. We are working in an integral domain (Exercise 7.52), and $\mathrm{lm}(f) \neq 0$, so it mus tbe that $q_2 - q_1 = 0$. In other words, $q_1 = q_2$.

Once we have $q_2 - q_1 = 0$, substitution into (26) implies that $0 = r_2 - r_1$. Immediately we have $r_1 = r_2$. We have shown that q and r are unique. □

Notice that the theorem does *not* apply if $R = \mathbb{Z}$, and Exercise 7.50 explains why. That's a shame.

Euclidean domains

Recall from Section 6.1 that the Euclidean algorithm for integers is basically repeated division. You can infer, more or less correctly, that a similar algorithm works for polynomials.

Why stop there? We have a notion of divisibility in rings, and we just found that the Division Theorem for integers can be generalized to any polynomial ring whose ground ring is a field. Can we generalize the Division Theorem beyond a ring of polynomials over a field? We can, but it requires us to generalize the notion of a remainder, as well.

Definition 7.58. Let R be an integral domain and v a function mapping the nonzero elements of R to $\mathbb{N}^+$. We say that R is a **Euclidean Domain** with respect to the **valuation function** v if it satisfies (E1) and (E2) where

(E1) $v(r) \leq v(rs)$ for all nonzero $r, s \in R$.

(E2) For all nonzero $f \in R$ and for all $g \in R$, there exist $q, r \in R$ such that
- $g = qf + r$, and
- $r = 0$ or $v(r) < v(f)$.

Example 7.59. By the Division Theorem, $\mathbb{Z}$ is a Euclidean domain with the valuation function $v(r) = |r|$.

Theorem 7.60. Let $\mathbb{F}$ be a field. Then $\mathbb{F}[x]$ is a Euclidean domain with the valuation function $v(r) = \deg r$.

Proof. You do it! See Exercise 7.70. □

Example 7.61. On the other hand, $\mathbb{Z}[x]$ is *not* a Euclidean domain if the valuation function is $v(r) = \deg r$. If $f = 2$ and $g = x$, we cannot find $q, r \in \mathbb{Z}[x]$ such that $g = qf + r$ and $\deg r < \deg f$. The best we can do is $x = 0 \cdot 2 + x$, but $\deg x > \deg 2$.

If you think back to the Euclidean algorithm, you might remember that it requires only *the ability to perform a division with a unique remainder that was smaller than the divisor*. This means that we can perform the Euclidean algorithm in a Euclidean ring! — But will the result have the same properties as when we perform it in the ring of integers?

Yes and no. We *do* get an object whose properties resemble those of the greatest common divisor of two integers. However, the result *might not be unique!* On the other hand, if we relax our expectation of uniqueness, we can get a greatest common divisor that is… *sort of* unique.

Definition 7.62. Let R be a ring. If $a, b, r \in R$ satisfy $ar = b$ or $ra = b$, then a **divides** b, a is a **divisor of** b, and b is **divisible** by a.

Now suppose that R is a Euclidean domain with respect to v, and let $a, b \in R$. If there exists $d \in R$ such that $d \mid a$ and $d \mid b$, then we call d a **common divisor** of a and b. If in addition all other common divisors d' of a and b divide d, then d is a **greatest common divisor** of a and b.

Two subtle differences with the definition for the integers have profound consequences.

- The definition refers to "a" greatest common divisor, not "the" greatest common divisor. *There can be many great"est" common divisors!*
- Euclidean domains measure "greatness" using divisibility (or multiplication) rather than order (or subtraction). As a consequence, the Euclidean domain R need not have a well ordering, or even a linear ordering — it needs only a valuation function! This is *why* there can be many great"est" common divisors.

Example 7.63. Consider $x^2 - 1, x^2 + 2x + 1 \in \mathbb{Q}[x]$. By Theorem 7.60, $\mathbb{Q}[x]$ is a Euclidean domain with respect to the valuation function $v(p) = \deg p$. Both of the given polynomials factor:

$$x^2 - 1 = (x + 1)(x - 1) \quad \text{and} \quad x^2 + 2x + 1 = (x + 1)^2,$$

so we see that $x + 1$ is a divisor of both. In fact, it is a greatest common divisor, since no polynomial of degree two divides both $x^2 - 1$ and $x^2 + 2x + 1$.

However, $x + 1$ is not the *only* greatest common divisor. Another greatest common divisor is $2x + 2$. It may not be obvious that $2x + 2$ divides both $x^2 - 1$ and $x^2 + 2x + 1$, but it does:

$$x^2 - 1 = (2x + 2)\left(\frac{x}{2} - \frac{1}{2}\right)$$

and

$$x^2 + 2x + 1 = (2x + 2)\left(\frac{x}{2} + \frac{1}{2}\right).$$

Notice that $2x+2$ divides $x+1$ and vice-versa; also that $\deg(2x+2) = \deg(x+1)$.

Likewise, $\frac{x+1}{3}$ is also a greatest common divisor of x^2-1 and x^2+2x+1.

This new definition will allow more than one greatest common divisor even in $\mathbb{Z}$! For example, for $a = 8$ and $b = 12$, both 4 *and* -4 are greatest common divisors! This happens because each divides the other, emphasizing that in a generic Euclidean domain, the notion of a "greatest" common divisor is relative to divisibility, not to other orderings. However, when speaking of greatest common divisors in the integers, we typically use the ordering, not divisibility.

That said, all greatest common divisors have something in common.

Proposition 7.64. Let R be a Euclidean domain with respect to v, and $a, b \in R$. Suppose that d is a greatest common divisor of a and b. If d' is a common divisor of a and b, then $v(d') \leq v(d)$. If d' is another greatest common divisor of a and b, then $v(d) = v(d')$.

Proof. Since d is a greatest common divisor of a and b, and d' is a common divisor, the definition of a greatest common divisor tells us that d divides d'. Thus there exists $q \in R$ such that $qd' = d$. From property (E1) of a Euclidean domain,

$$v(d') \leq v(qd') = v(d).$$

On the other hand, if d' is also a greatest common divisor of a and b, an argument similar to the one above shows that

$$v(d) \leq v(d') \leq v(d).$$

Hence $v(d) = v(d')$. □

Finally we come to the point of a Euclidean domain: we can use the Euclidean algorithm to compute a gcd of any two elements! Essentially we transcribe the Euclidean Algorithm for integers (Theorem 6.4 on page 179 of Section 6.1).

Theorem 7.65 (The Euclidean Algorithm for Euclidean domains)**.** Let R be a Euclidean domain with valuation v and $m, n \in R \setminus \{0\}$. One can compute a greatest common divisor of m, n in the following way:

1. Let $s = m$ and $t = n$.
2. Repeat the following steps until $t = 0$:
   (a) Let q be the quotient and r the remainder after dividing s by t.
   (b) Assign s the current value of t.
   (c) Assign t the current value of r.

The final value of s is a greatest common divisor of m and n.

Proof. You do it! See Exercise 7.71. □

Just as we could adapt the Euclidean Algorithm for integers to the Extended Euclidean Algorithm in order to compute $a, b \in \mathbb{Z}$ such that Bezout's Identity holds,

$$am + bn = \gcd(m, n),$$

we can do the same in Euclidean domains. You will need this for Exercise 7.71.

Exercises.

Exercise 7.66. Try to devise a division algorithm for $\mathbb{Z}_n$? Does the value of n matter?

Exercise 7.67. Let $f = 2x^2 + 1$ and $g = x^3 - 1$.

(a) Show that 1 is a greatest common divisor of f and g in $\mathbb{Q}[x]$, and find $a, b \in \mathbb{Q}[x]$ such that $1 = af + bg$.

(b) Recall that $\mathbb{Z}_5$ is a field. Show that 1 is a greatest common divisor of f and g in $\mathbb{Z}_5[x]$, and find $a, b \in \mathbb{Z}_5[x]$ such that $1 = af + bg$.

(c) Recall that $\mathbb{Z}[x]$ is not a Euclidean domain. Explain why the result of part (a) cannot be used to show that 1 is a greatest common divisor of f and g in $\mathbb{Z}[x]$. What would you get if you used the Euclidean algorithm on f and g in $\mathbb{Z}[x]$?

Exercise 7.68. Let $f = x^4 + 9x^3 + 27x^2 + 31x + 12$ and $g = x^4 + 13x^3 + 62x^2 + 128x + 96$.

(a) Compute a greatest common divisor of f and g in $\mathbb{Q}[x]$.

(b) Recall that $\mathbb{Z}_{31}$ is a field. Compute a greatest common divisor of f and g in $\mathbb{Z}_{31}[x]$.

(c) Recall that $\mathbb{Z}_3$ is a field. Compute a greatest common divisor of f and g in $\mathbb{Z}_3[x]$.

(d) Even though $\mathbb{Z}[x]$ is not a Euclidean domain, it still has greatest common divisors. What's more, we can compute the greatest common divisors using the Euclidean algorithm! How?

(e) You can even compute the greatest common divisors *without* using the Euclidean algorithm, but by examining the answers to parts (b) and (c) slowly. How?

Exercise 7.69. Show that every field is a Euclidean domain.

Exercise 7.70. Prove Theorem 7.60.

Exercise 7.71. Prove Theorem 7.65, the Euclidean Algorithm for Euclidean domains.

Exercise 7.72. A famous Euclidean domain is the ring of *Gaussian integers*

$$\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$$

where $i^2 = -1$. Let $v : \mathbb{Z}[i] \to \mathbb{Z}$ by

$$v(a + bi) = a^2 + b^2.$$

(a) Show that $a + bi$ is "orthogonal" to $i(a + bi)$, in the sense that the slope of the line segment connecting 0 and $a + bi$ in the complex plane is orthogonal to the slope of the line segment connecting 0 and $i(a + bi)$.

(b) Assuming the facts given about v, divide:
   (i) 11 by 3;
   (ii) 11 by $3i$;
   (iii) $2 + 3i$ by $1 + 2i$.

(c) Show that v is, in fact, a valuation function suitable for a Euclidean domain.

(d) Describe a method for dividing Gaussian integers. (Again, it helps to think of them as vectors in the plane. See Exercise 0.52 on page 19.)

Chapter 8:
Ideals

This chapter fills two roles. Some sections describe ring analogs to structures that we introduced in group theory:

- Section 8.1 introduces the *ideal*, an analog to a normal subgroup;
- Section 8.4 provides an analog of quotient groups; and
- Section 8.5 decribes ring homomorphisms.

The remaining sections use these ring structures to introduce new kinds of ring structures:

- Section 8.2 describes an important class of rings;
- Section 8.3 highlights an important class of ideals; and
- Section 10.3 brings us to finite fields, which are important for computation in polynomial rings.

8.1: Ideals

You might think that, just as we moved from subgroups to quotient groups via cosets, we will move from subrings to "quotient rings" via the ring analog of normal subgroups. While this is true, the analog may not have quite the form you expect.

Definition and examples

Definition 8.1. Let A be a subring of R that satisfies the **absorption property**:

$$\forall r \in R \quad \forall a \in A \qquad ra \in A.$$

Then A is an **ideal subring** of R, or simply, an **ideal**, and we write $A \triangleleft R$. An ideal A is **proper** if $\{0\} \neq A \neq R$.

Recall that our rings are assumed to be commutative, so if $ra \in A$ then $ar \in A$, also.

Example 8.2. Recall the subring $2\mathbb{Z}$ of the ring $\mathbb{Z}$. We claim that $2\mathbb{Z} \triangleleft \mathbb{Z}$. Why? Let $r \in \mathbb{Z}$, and $a \in 2\mathbb{Z}$. By definition of $2\mathbb{Z}$, there exists $d \in \mathbb{Z}$ such that $a = 2d$. Substitution gives us

$$ra = r \cdot 2d = 2(rd) \in 2\mathbb{Z},$$

so $2\mathbb{Z}$ "absorbs" multiplication by $\mathbb{Z}$. This makes $2\mathbb{Z}$ an ideal of $\mathbb{Z}$.

Naturally, we can generalize this proof to arbitrary $n \in \mathbb{Z}$: see Exercises 8.15 and 8.16.

Ideals in the ring of integers have a nice property that we will use in future examples.

Lemma 8.3. Let $a, b \in \mathbb{Z}$. The following are equivalent:
(A) $a \mid b$;
(B) $b\mathbb{Z} \subseteq a\mathbb{Z}$.

Proof. You do it! See Exercise 8.17. □