Chapter 16

An Introduction to Rings and Fields

GOALS

In our early elementary school days we began the study of mathematics by learning addition and multiplication on the set of positive integers. We then extended this to operations on the set of all integers. Subtraction and division are defined in terms of addition and multiplication. Later we investigated the set of real numbers under the operations of addition and multiplication. Hence, it is quite natural to investigate those structures on which we can define these two fundamental operations, or operations similar to them. The structures similar to the set of integers are called rings, and those similar to the set of real numbers are called fields.

In coding theory, highly structured codes are needed for speed and accuracy. The theory of finite fields is essential in the development of many structured codes. We will discuss basic facts about finite fields and introduce the reader to polynomial algebra.

16.1 Rings—Basic Definitions and Concepts

As mentioned in our goals, we would like to investigate algebraic systems whose structure imitates that of the integers.

Definition: Ring. *A ring is a set R together with two binary operations, addition and multiplication, denoted by the symbols + and · such that the following axioms are satisfied:*

(1) $[R, +]$ *is an abelian group.*

(2) Multiplication is associative on R.

(3) Multiplication is distributive over addition; that is, for all a, b, c $\in R$*, the left distributive law,* $a(b + c) = ab + ac$*, and the right distributive law,* $(b + c)a - ba + ca$*, hold.*

Comments:

(1) A ring is designated as $[R, +, \cdot]$ or as just plain R if the operations are understood.

(2) The symbols + and · stand for arbitrary operations, not just "regular" addition and multiplication. These symbols are referred to by the usual names. For simplicity, we will write $a\,b$ instead of $a \cdot b$ if it is not ambiguous.

(3) For the abelian group $[R, +]$, we use additive notation. In particular, the group identity is designated by 0 rather than by e and is customarily called the "zero" of the ring. The group inverse is also written in additive notation: $-a$ rather than a^{-1}.

We now look at some examples of rings. Certainly all the additive abelian groups of Chapter 11 are likely candidates for rings.

Example 16.1.1. $[\mathbb{Z}, +, \cdot]$ is a ring, where + and · stand for regular addition and multiplication on $\mathbb{Z}$. From Chapter 11, we already know that $[\mathbb{Z}, +]$ is an abelian group, so we need only check parts 2 and 3 of the definition of a ring. From elementary algebra, we know that the associative law under multiplication and the distributive laws are true for $\mathbb{Z}$. This is our main example of an infinite ring.

Example 16.1.2. $[\mathbb{Z}_n, +_n, \times_n]$ is a ring. The properties of modular arithmetic on $\mathbb{Z}_n$ were described in Section 11.4, and they give us the information we need to convince ourselves that $[\mathbb{Z}_n, +_n, \times_n]$ is a ring. This example is our main example of finite rings of different orders.

Definition: Commutative Ring. *A ring in which the commutative law holds under the operation of multiplication is called a commutative ring.*

It is common practice to use the word abelian when referring to the commutative law under addition and the word commutative when referring to the commutative law under the operation of multiplication.

Definition: Unity. *A ring $[R, +, \cdot]$ that has a multiplicative identity is called a ring with unity. The multiplicative identity itself is called the unity of the ring. More formally, if there exists an element in R, designated by 1, such that for all $x \in R$, $x \cdot 1 = 1 \cdot x = x$, then R is called a ring with unity.*

Example 16.1.3. The rings in Examples 16.1.1 and 16.1.2 are commutative rings with unity, the unity in both cases being the number 1. The ring $[M_{2\times2}(\mathbb{R}), +, \cdot]$ is a noncommutative ring with unity, the unity being the identity matrix $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

DIRECT PRODUCTS OF RINGS

Let $R_1, R_2, \ldots, R_n$ be rings under the operations $+_1, +_2, \ldots, +_n$ and $\cdot_1, \cdot_2, \ldots, \cdot_n$ respectively. Let

$$P = \overset{n}{\underset{i=1}{\times}} R_i$$

and $a = \{a_1, a_2, \ldots, a_n), b = (b_1, b_2, \ldots, b_n) \in P$.

From Chapter 11 we know that P is an abelian group under the operation of componentwise addition:

$$a + b = (a_1 +_1 b_1, a_2 +_2 b_2, \ldots, a_n +_n b_n).$$

We also define multiplication on P componentwise:

$$a \cdot b = (a_1 \cdot_1 b_1, a_2 \cdot_2 b_2, \ldots, a_n \cdot_n b_n).$$

To show that P is a ring under the above operations, we need only show that the (multiplicative) associative law and the distributive laws hold. This is indeed the case, and we leave it as an exercise. If each of the R_i is commutative, then P is commutative, and if each contains a unity, then P is a ring with unity, which is the $n-$ tuple consisting of the unities of each of the R_i's.

Example 16.1.4. Since $[\mathbb{Z}_4, +_4, \times_4]$ and $[\mathbb{Z}_3, +_3, \times_3]$ are rings, then $\mathbb{Z}_4 \times \mathbb{Z}_3$ is a ring, where, for example,

$$(2, 1) + (2, 2) = (2 +_4 2, 1 +_3 2) = (0, 0)$$

and

$$(3, 2) \cdot (2, 2) = (3 \times_4 2, 2 \times_3 2) = (2, 1).$$

To determine the unity, if it exists, in the ring $\mathbb{Z}_4 \times \mathbb{Z}_3$, we look for the element (m, n) such that for all elements $(x, y) \in \mathbb{Z}_4 \times \mathbb{Z}_3$,

$$(x, y) = (x, y) \cdot (m, n) = (m, n) \cdot (x, y),$$

or, equivalently,

$$(x \times_4 m, y \times_3 n) = (m \times_4 x, n \times_3 y) = (x, y).$$

So we want m such that $x \times_4 m = m \times_4 x = x$ in the ring $\mathbb{Z}_4$. The only element m in $\mathbb{Z}_4$ that satisfies this equation is $m = 1$. Similarly, we obtain a value of 1 for n. So the unity of $\mathbb{Z}_4 \times \mathbb{Z}_3$, which is unique by Exercise 15 of this section, is (1, 1). We leave to the reader to verify that this ring is commutative.

Hence, products of rings are analogous to products of groups or products of Boolean algebras. We now consider the extremely important concept of multiplicative inverses. Certainly many basic equations in elementary algebra (e.g., $2x = 3$) are solved with this concept. We introduce the main idea here and develop it more completely in the next section.

Example 16.1.5. The equation $2x = 3$ has a solution in the ring $[\mathbb{R}, +, \cdot]$ but does not have a solution in $[\mathbb{Z}, +, \cdot]$, since, to solve this equation, we multiply both sides of the equation $2x = 3$ by the multiplicative inverse of 2. This number, 2^{-1} exists in $\mathbb{R}$ but does not exist in $\mathbb{Z}$. We formalize this important idea in a definition which by now should be quite familiar to you.

Definition: Multiplicative Inverses. *Let $[R, +, \cdot]$ be a ring with unity, 1. If $u \in R$ and there exists an element $v \in R$ such that $u \cdot v = v \cdot u = 1$, then u is said to have a multiplicative inverse, v. We call a ring element that possesses a multiplicative inverse a unit of the ring. The set of all units of a ring R is denoted by U(R).*

By Theorem 11.3.2, the multiplicative inverse of a ring element is unique, if it exists. For this reason, we can use the notation u^{-1} for the multiplicative inverse of u, if it exists.

Example 16.1.6. In the rings $[\mathbb{R}, +, \cdot]$ and $[\mathbb{Q}, +, \cdot]$ every nonzero element has a multiplicative inverse. The only elements in $\mathbb{Z}$ that have multiplicative inverses are -1 and 1. That is, $U(\mathbb{R}) = \mathbb{R}^*$, $U(\mathbb{Q}) = \mathbb{Q}^*$, and $U(\mathbb{Z}) = \{-1, 1\}$.

Example 16.1.7. Let us find the multiplicative inverses, when they exist, of each element of the ring $[\mathbb{Z}_6, +_6, \times_6]$. If $u = 3$, we want an element v such that $u \times_6 v = 1$. We do not have to check whether $v \times_6 u = 1$ since $\mathbb{Z}_6$ is commutative. If we try each of the six elements, 0, 1, 2, 3, 4, and 5, of $\mathbb{Z}_6$, we find that none of them satisfies the above equation, so 3 does not have a multiplicative inverse in $\mathbb{Z}_6$. However, since $5 \times_6 5 = 1$, 5 does have a multiplicative inverse in $\mathbb{Z}_6$, namely itself: $5^{-1} = 5$. The following table summarizes all results for $\mathbb{Z}_6$.

u	u^{-1}
0	does not exist
1	1
2	does not exist
3	does not exist
4	does not exist
5	5

It shouldn't be a surprise that the zero of a ring is never going to have a multiplicative inverse except in the trivial case of $R = \{0\}$.

Isomorphism is a universal concept that is important in every algebraic structure. Two rings are isomorphic as rings if and only if they have the same cardinality and if they behave exactly the same under corresponding operations. They are essentially the same ring. For this to be true, they must behave the same as groups (under +) and they must behave the same under the operation of multiplication.

Definition: Ring Isomorphism. *Let* $[R, +, \cdot]$ *and* $[R', +', \cdot']$ *be rings. Then* R *is isomorphic to* R' *if and only if there exists a map,* $f : R \to R'$*, called a ring isomorphism, such that*

(1) f *is one-to-one and onto,*

(2) $f(a + b) = f(a) +' f(b)$ *for all* $a, b \in R$*, and*

(3) $f(a \cdot b) = f(a) \cdot' f(b)$ *for all* $a, b \in R$*.*

Conditions 1 and 2 tell us that f is a group isomorphism. Therefore, to show that two rings are isomorphic, we must produce a map, called an isomorphism, that satisfies the definition. Sometimes it is quite difficult to find a map that works. This does not necessarily mean that no such isomorphism exists, but simply that we cannot find it.

This leads us to the problem of how to show that two rings are not isomorphic. This is a universal concept. It is true for any algebraic structure and was discussed in Chapter 11. To show that two rings are not isomorphic, we must demonstrate that they behave differently under one of the operations. We illustrate through several examples.

Example 16.1.8. Consider the rings $[\mathbb{Z}, +, \cdot]$ and $[2\mathbb{Z}, +, \cdot]$. In Chapter 11 we showed that as groups, the two sets $\mathbb{Z}$ and $2\mathbb{Z}$ with addition were isomorphic. The group isomorphism that proved this was the map $f : \mathbb{Z} \to 2\mathbb{Z}$, defined by $f(n) = 2n$. Is f a ring isomorphism? We need only check whether $f(m \cdot n) = f(m) \cdot f(n)$ for all $m, n \in \mathbb{Z}$:

$f(m \cdot n) = 2 \cdot m \cdot n$ and

$f(m) \cdot f(n) = 2m \cdot 2n = 4 \cdot m \cdot n$

Therefore, f is not a ring isomorphism. This does not necessarily mean that the two rings $\mathbb{Z}$ and $2\mathbb{Z}$ are not isomorphic, but simply that the f doesn't satisfy the conditions. We could imagine that some other function does. We could proceed and try to determine another function f to see whether it is a ring isomorphism, or we could try to show that $\mathbb{Z}$ and $2\mathbb{Z}$ are not isomorphic as rings. To do the latter, we must find something different about the ring structure of $\mathbb{Z}$ and $2\mathbb{Z}$.

We already know that they behave identically under addition, so if they are different as rings, it must have something to do with how they behave under the operation of multiplication. Let's begin to develop a checklist of how the two rings could differ:

(1) Do they have the same cardinality? Yes, they are both countable.

(2) Are they both commutative? Yes.

(3) Are they both rings with unity? No.

$\mathbb{Z}$ is a ring with unity, namely the number 1. $2\mathbb{Z}$ is not a ring with unity, $1 \notin 2\mathbb{Z}$. Hence, they are not isomorphic as rings.

Example 16.1.9. Next consider whether $[2\mathbb{Z}, +, \cdot]$ and $[3\mathbb{Z}, +, \cdot]$ are isomorphic. Because of the previous example, we might guess that they are not. However, checklist items 1 through 3 above do not help us. Why? We add another checklist item:

(4) Find an equation that makes sense in both rings, which is solvable in one and not the other.

The equation $x + x = x \cdot x$, or $2x = x^2$, makes sense in both rings. However, this equation has a nonzero solution, $x = 2$, in $2\mathbb{Z}$, but does not have a nonzero solution in $3\mathbb{Z}$. Thus we have an equation solvable in one ring that cannot be solved in the other, so they cannot be isomorphic.

Another universal concept that applies to the theory of rings is that of a subsystem. A subring of a ring $[R, +, \cdot]$ is any nonempty subset S of R that is a ring under the operations of R. First, for S to be a subring of the ring R, S must be a subgroup of the group $[R, +]$. Also, S must be closed under $\cdot$, satisfy the associative law (under $\cdot$), and satisfy the distributive laws. But since R is a ring, the associative and distributive laws are true for every element in R, and, in particular, for all elements in S, since $S \subseteq R$. We have just proven the following theorem:

Theorem 16.1.1. *A subset S of a ring* $[R, +, \cdot]$ *is a subring of R if and only if:*

(1) $[S, +]$ *is a subgroup of the group* $[R, +]$*, which by Theorem 11.5.1, means we must show:*

(a) If $a, b \in S$*, then* $a + b \in S$*,*

(b) $0 \in S$*, and*

(c)) If $a \in S$, then $-a \in S$.

(2) S is closed under multiplication: if $a, b \in S$, then $a \cdot b \in S$.

Example 16.1.10. The set of even integers, $2\mathbb{Z}$, is a subring of the ring $[\mathbb{Z}, +, \cdot]$ since $[2\mathbb{Z}, +]$ is a subgroup of the group $[\mathbb{Z}, +]$ and since it is also closed with respect to multiplication:

$$2m, 2n \in 2\mathbb{Z} \Rightarrow (2m)\cdot(2n) = 2(2\cdot m\cdot n) \in 2\mathbb{Z}.$$

Several of the basic facts that we are familiar with are true for any ring. The following theorem lists a few of the elementary properties of rings.

Theorem 16.1.2. *Let $[R, +, \cdot]$ be a ring, with $a, b \in R$. Then*

(1) $a \cdot 0 = 0 \cdot a = 0$

(2) $a\cdot(-b) = (-a)\cdot b = -(a\cdot b)$

(3) $(-a)\cdot(-b) = a\cdot b$

Proof of Part 1:

$$\begin{aligned} a \cdot 0 &= a \cdot (0 + 0) \\ &= a \cdot 0 + a \cdot 0 \text{ by the left distributive law.} \end{aligned}$$

Hence if we add $-(a \cdot 0)$ to both sides of the above, we obtain $a \cdot 0 = 0$. Similarly, we can prove that $0 \cdot a = 0$.

Proof of Part 2: Before we begin the proof of part 2, recall that the inverse of each element of the group $[R, +]$ is unique. Hence the inverse of the element $a \cdot b$ is unique and it is denoted $-(a \cdot b)$.

Therefore, to prove that $a\cdot(-b) = -(a \cdot b)$, we need only show that $a\cdot(-b)$ inverts $a\cdot b$.

$$\begin{aligned} a\cdot(-b) + a\cdot b &= a\cdot(-b + b) && \text{by the distributive axiom} \\ &= a\cdot 0 && \text{since } -b \text{ inverts } b \\ &= 0 && \text{by part 1 of this theorem} \end{aligned}$$

Similarly, it can be shown that $(-a)\cdot b = -(a \cdot b)$. This completes the proof of part 2.

We leave the proof of part 3 to the reader (see Exercise 16 of this section). ■

Example 16.1.11. We will compute $2\cdot(-2)$ in the ring $[\mathbb{Z}_6, +_6, \times_6]$.

$$2 \times_6 (-2) = -(2 \times_6 2) = -4 = 2,$$

since the additive inverse of 4 (mod 6) is 2. Of course, we could have done the calculation directly as

$$2 \times_6 (-2) = 2 \times_6 4 = 2.$$

As the example above illustrates, Theorem 16.1.2 is a modest beginning in the study of which algebraic manipulations are possible in the solution of problems in rings. A fact in elementary algebra that is used frequently in problem solving is the cancellation law. We know that the cancellation laws are true under addition for any ring (Theorem 11.3.5).

Are the cancellation laws true under multiplication? More specifically, let $[R, +, \cdot]$ be a ring and let $a, b, c \in R$ with $a \neq 0$. When can we cancel the a's in the equation $a \cdot b = a \cdot c$? We can certainly do so if a^{-1} exists, but we cannot assume that a has a multiplicative inverse. The answer to this question is found with the following definition and Theorem 16.1.3.

Definition: Divisors of Zero. *Let $[R, +, \cdot]$ be a ring. If a and b are two nonzero elements of R such that $a \cdot b = 0$, then a and b are called divisors of zero.*

Example 16.1.12 (a) In the ring $[\mathbb{Z}_8, +_8, \times_8]$, the numbers 4 and 2 are divisors of zero since $4 \times_8 2 = 0$. In addition, 6 is a divisor of zero because $6 \times_8 4 = 0$.

(b) In the ring $[M_{2\times2}(\mathbb{R}), +, \cdot]$ the matrices $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ are divisors of zero since $AB = 0$.

Example 16.1.13. $[\mathbb{Z}, +, \cdot]$ has no divisors of zero.

Now here is why divisors of zero are related to cancellation.

Theorem 16.1.3. *The (multiplicative) cancellation law holds in a ring $[R, +, \cdot]$ if and only if R has no divisors of zero.*

We prove the theorem using the left cancellation law, namely that if $a \neq 0$ and $a \cdot b = a \cdot c$, then $b = c$ for all $a, b, c \in R$. The proof is similar using the right cancellation law.

Proof: ($\Rightarrow$) Assume the left cancellation law holds in R and assume that a and b are two elements in R such that $a \cdot b = 0$. We must show that either $a = 0$ or $b = 0$. To do this, assume that $a \neq 0$ and show that b must be 0.

$$\begin{aligned} a\cdot b = 0 \Rightarrow\ & a\cdot b = a\cdot 0 \text{ by Theorem 16.2.1, part 1} \\ \Rightarrow\ & b = 0 \text{ by the cancellation law} \end{aligned}$$

(⇐) Conversely, assume that R has no divisors of 0 and we will prove that the cancellation law must hold. To do this, assume that $a, b, c \in R$, $a \neq 0$, such that $a \cdot b = a \cdot c$ and show that $b = c$.

$$\begin{aligned} a \cdot b = a \cdot c &\Rightarrow a \cdot b - a \cdot c = 0 \quad \text{Why?} \\ &\Rightarrow a \cdot (b - c) = 0 \quad \text{Why?} \\ &\Rightarrow b - c = 0 \quad \text{Why} \\ &\Rightarrow b = c \end{aligned}$$ ■

Hence, the only time that the cancellation laws hold in a ring is when there are no divisors of zero. The commutative rings with unity in which the above is true are given a special name.

Definition: Integral Domain. *A commutative ring with unity containing no divisors of zero is called an integral domain.*

In this chapter, Integral domains will be denoted generically by the letter D.

We state the following two useful facts without proof.

Theorem 16.1.4. *The element m in the ring $\mathbb{Z}_n$ is a divisor of zero if and only if m is not relatively prime to n (i.e., gcd(m, n) ≠ 1).*

Corollary. *If p is a prime, then $\mathbb{Z}_p$ has no divisors of zero.*

Example 16.1.14. $[\mathbb{Z}, +, \cdot]$, $[\mathbb{Z}_p, +_p, \times_p]$ with p a prime, $[\mathbb{Q}, +, \cdot]$, $[\mathbb{R}, +, \cdot]$, and $[\mathbb{C}, +, \cdot]$ are all integral domains. The key example of an infinite integral domain is $[\mathbb{Z}, +, \cdot]$. In fact, it is from $\mathbb{Z}$ that the term integral domain is derived. The main example of a finite integral domain is $[\mathbb{Z}_p, +_p, \times_p]$, when p is prime.

We close this section with the verification of an observation that was made in Chapter 11, namely that the product of two algebraic systems may not be an algebraic system of the same type.

Example 16.1.15. Both $[\mathbb{Z}_2, +_2, \times_2]$ and $[\mathbb{Z}_3, +_3, \times_3]$ are integral domains. Consider the product $\mathbb{Z}_2 \times \mathbb{Z}_3$. It's true that $\mathbb{Z}_2 \times \mathbb{Z}_3$ is a commutative ring with unity (see Exercise 13). However, $(1, 0) \cdot (0, 2) = (0, 0)$, so $\mathbb{Z}_2 \times \mathbb{Z}_3$ has divisors of zero and is therefore not an integral domain.

EXERCISES FOR SECTION 16.1

A Exercises

1. Review the definition of rings to show that the following are rings. The operations involved are the usual operations defined on the sets. Which

of these rings are commutative? Which are rings with unity? For the rings with unity, determine the unity and all units.

(a) $[\mathbb{Z}, +, \cdot]$

(b) $[\mathbb{C}, +, \cdot]$

(c) $[M_{n \times n}(\mathbb{R}), +, \cdot]$

(d) $[\mathbb{Q}, +, \cdot]$

(e) $[M_{2 \times 2}(\mathbb{R}), +, \cdot]$

(f) $[\mathbb{Z}_2, +_2, \times_2]$

2. Follow the instructions for Exercise 1 and the following rings:

(a) $[\mathbb{Z}_6, +_6, \times_6]$

(b) $[\mathbb{Z}_5, +_5, \times_5]$

(c) $[\mathbb{Z}_2{}^3, +, \cdot]$

(d) $[\mathbb{Z}_8, +_8, \times_8]$

(e) $[\mathbb{Z} \times \mathbb{Z}, +, \cdot]$

(f) $[\mathbb{R}^2, +, \cdot]$

3. Show that the following pairs of rings are not isomorphic:

(a) $[\mathbb{Z}, +, \cdot]$ and $[M_{2 \times 2}(\mathbb{Z}), +, \cdot]$

(b) $[3\mathbb{Z}, +, \cdot]$ and $[4\mathbb{Z}, +, \cdot]$.

4. Show that the following pairs of rings are not isomorphic:

(a) $[\mathbb{R}, +, \cdot]$ and $[\mathbb{Q}, +, \cdot]$.

(b) $[\mathbb{Z}_2 \times \mathbb{Z}_2, +, \cdot]$and $[\mathbb{Z}_4, +, \cdot]$.

5. (a) Show that $3\mathbb{Z}$ is a subring of the ring $[\mathbb{Z}, +, \cdot]$

(b) Find all subrings of $\mathbb{Z}_8$.

(c) Find all subrings of $\mathbb{Z}_2 \times \mathbb{Z}_2$.

6. Verify the validity of Theorem 16.1.3 by finding examples of elements a, b, and c ($a \neq 0$) in the following rings, where $a \cdot b = a \cdot c$ and yet $b \neq c$:

(a) $\mathbb{Z}_8$

(b) $M_{2\times2}(\mathbb{R})$

(c) ${\mathbb{Z}_2}^2$

7. (a) Determine all solutions of the equation $x^2 - 5x + 6 = 0$ in $\mathbb{Z}$. Can there be any more than two solutions to this equation (or any quadratic equation) in $\mathbb{Z}$?

(b) Find all solutions of the equation in part a in $\mathbb{Z}_{12}$. Why are there more than two solutions?

8. Solve the equation $x^2 + 4x + 4 = 0$ in the following rings. Interpret 4 as $1 + 1 + 1 + 1$, where 1 is the unity of the ring.

(a) in $\mathbb{Z}_8$

(b) in $M_{2\times2}(\mathbb{R})$

(c) in $\mathbb{Z}$

(d) in $\mathbb{Z}_3$

B Exercises

9. The relation "is isomorphic to" on rings is an equivalence relation. Explain the meaning of this statement.

10. Let $R_1, R_2, \ldots, R_n$ be rings. Prove the multiplicative, associative, and distributive laws for the ring

$$R = \bigtimes_{i=1}^{n} R_i$$

(a) If each of the R_i is commutative, is R commutative?

(b) Under what conditions will R be a ring with unity?

(c) What will the units of R be when it has a unity?

11. (a) Prove that the ring $\mathbb{Z}_2$ x $\mathbb{Z}_3$ is commutative and has unity.

(b) Determine all divisors of zero for the ring $\mathbb{Z}_2$ x $\mathbb{Z}_3$.

(c) Give another example illustrating the fact that the product of two integral domains may not be an integral domain. Is there an example where the product is an integral domain?

12. **Boolean Rings.** Let U be a nonempty set.

(a) Verify that $[\mathcal{P}(U), \oplus, \bigcap]$ is a commutative ring with unity.

(b) What are the units of this ring?

13. (a) For any ring $[R, +, \cdot]$, expand $(a + b)(c + d)$ for $a, b, c, d \in R$.

(b) If R is commutative, prove that $(a + b)^2 = a^2 + 2ab + b^2$ for all $a, b \in R$.

14. (a) Let R be a commutative ring with unity. Prove by induction that for $n \geq 1$,

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

(b) Simplify $(a + b)^5$ in $\mathbb{Z}_5$.

(c) Simplify $(a + b)^{10}$ in $\mathbb{Z}_{10}$.

15. Prove: If R is a ring with unity then this unity is unique.
16. Prove part 3 of Theorem 16.1.2.
17. Prove the Corollary to Theorem 16.1.4.
18. Let U be a finite set. Prove that the Boolean ring $[\mathcal{P}(U), \oplus, \cap]$ is isomorphic to the ring $[\mathbb{Z}_2{}^n, +, \cdot]$. where $n = |U|$

16.2 Fields

Although the algebraic structures of rings and integral domains are widely used and play an important part in the applications of mathematics, we still cannot solve the simple equation $a\,x = b, a \neq 0$ in all rings or in all integral domains. Yet this is one of the first equations we learn to solve in elementary algebra and its solubility is basic to innumerable questions. Certainly, if we wish to solve a wide range of problems in a system we need at least all of the laws true for rings and the cancellation laws together with the ability to solve the equation $a\,x = b, a \neq 0$. We summarize the above in a definition and list several theorems without proof that will place this concept in the context of the previous section.

Definition: Field. A field is a commutative ring with unity such that each nonzero element has a multiplicative inverse.

In this chapter, we denote a field generically by the letter F. The letters k, K and L are also conventionally used for fields.

Example16.2.1. $[\mathbb{Q}, +, \cdot]$, $[\mathbb{R}, +, \cdot]$, and $[\mathbb{C}, +, \cdot]$ are all fields.

Reminder: Since every field is a ring, all facts and concepts that are true for rings are true for any field.

Theorem 16.2.1*. Every field is an integral domain.*

Of course the converse of Theorem 16.2.1 is not true. Consider $[\mathbb{Z}, +, \cdot]$.

Theorem 16.2.2. *Every finite integral domain is a field.*

Theorem 16.2.3. *If p is a prime, then $\mathbb{Z}_p$ is a field.*

Theorem 16.2.3 is immediate from Theorem 16.2.2.

Theorem 16.2.1 reminds us that the cancellation laws must be true for any field. Theorem 16.2.3 gives us a large number of finite fields, but we must be cautious. This theorem does not tell us that all finite fields are of the form $\mathbb{Z}_p$, p a prime. To see this, let's try to construct a field of order 4.

Example 16.2.2: a field of order 4. First the field must contain the additive and multiplicative identities, 0 and 1, so, without loss of generality, we can assume that the field we are looking for is of the form $F = \{0, 1, a, b\}$. Since there are only two nonisomorphic groups of order 4, we have only two choices for the group table for $[F, +]$. If the additive group is isomorphic to $\mathbb{Z}_4$ then two of the nonzero elements of F would not be their own additive inverse (as are 1 and 3 in $\mathbb{Z}_4$). Lets assume $\beta \in F$ is one of those elements and $\beta + \beta = \gamma \neq 0$. An isomorphism between the additive groups F and $\mathbb{Z}_4$ would require that γ in F correspond with 2 in $\mathbb{Z}_4$. We could continue our argument and infer that $\gamma \cdot \gamma = 0$, producing a zero divisor, which we need to avoid if F is to be a field. We leave the remainder of the argument to the reader. We can thus complete the addition table so that $[F, +]$ is isomorphic to ${\mathbb{Z}_2}^2$:

+	0	1	a	b
0	0	1	a	b
1	1	0	b	a
a	a	b	0	1
b	b	a	1	0

Next, by Theorem 16.1.2, Part 1, and since 1 is the unity of F, the table for multiplication must look like:

·	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	–	–
b	0	b	–	–

Hence, to complete the table, we have only four entries to find, and, since F must be commutative, this reduces our task to filling in three entries. Next, each nonzero element of F must have a unique multiplicative inverse. The inverse of a must be either a itself or b. If $a^{-1} = a$, then $b^{-1} = b$. (Why?) But

$a^{-1} = a \Rightarrow a \cdot a = 1$. And if $a \cdot a = 1$, then $a \cdot b$ is equal to a or b. In either case, by the cancellation law, we obtain $a = 1$ or $b = 1$, which is impossible. Therefore we are forced to conclude that $a^{-1} = b$ and $b^{-1} = a$. To determine the final two products of the table, simply note that, $a \cdot a \neq a$ because the equation $x^2 = x$ has only two solutions, 0 and 1 in any field. We also know that $a \cdot a$ cannot be 1 because a doesn't invert itself and cannot be 0 because a can't be a zero divisor. This leaves us with one possible conclusion, that $a \cdot a = b$ and similarly $b \cdot b = a$. Hence, our multiplication table for F is:

·	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	b	1
b	0	b	1	a

The table listing the multiplicative inverse of each nonzero element is:

u	u^{-1}
1	1
a	b
b	a

We leave it to the reader to convince him- or herself, if it is not already clear, that $[F, +, \cdot]$, as described above, is a field. Hence, we have produced a field of order 4 and 4 is not a prime.

This construction would be difficult to repeat for larger fields. In section 16.4 we will introduce a different approach to constructing fields that will be far more efficient.

Even though not all finite fields are isomorphic to $\mathbb{Z}_p$, for some prime p it can be shown that every field F must have either:

(1) a subfield isomorphic to $\mathbb{Z}_p$ for some prime p, or

(2) a subfield isomorphic to $\mathbb{Q}$.

In particular, if F is a finite field, a subfield of F must exist that is isomorphic to $\mathbb{Z}_p$. One can think of all fields as being constructed from either $\mathbb{Z}_p$ or $\mathbb{Q}$.

Example 16.2.3. $[\mathbb{R}, +, \cdot\ ]$ is a field, and it contains a subfield isomorphic to $[\mathbb{Q}, +, \cdot]$, namely $\mathbb{Q}$ itself.

Example 16.2.4. The field F that we constructed in Example 16.2.2 should have a subfield isomorphic to $\mathbb{Z}_p$ for some prime p. From the tables, we note that the subset $\{0, 1\}$ of $\{0, 1, a, b\}$ under the given operations of F behaves exactly like $[\mathbb{Z}_2, +_2, \times_2]$. Hence, the field in Example 16.2.2 has a subfield isomorphic to $\mathbb{Z}_2$. Does it have a subfield isomorphic to a larger field, say $\mathbb{Z}_3$? We claim not and leave this investigation to the reader (see Exercise 3 of this section).

We close this section with a brief discussion of isomorphic fields. Again, since a field is a ring, the definition of isomorphism of fields is the same as that of rings. It can be shown that if f is a field isomorphism, then $f(a^{-1}) = f(a)^{-1}$; that is, inverses are mapped onto inverses under any field isomorphism. A major question to try to solve is: How many different non-isomorphic finite fields are there of any given order? If p is a prime, it seems clear from our discussions that all fields of order p are isomorphic to $\mathbb{Z}_p$. But how many nonisomorphic fields are there, if any, of order 4, 6, 8, 9, etc? The answer is given in the following theorem, whose proof is beyond the scope of this text.

Theorem 16.2.4.

(1) Any finite field F has order p^n for a prime p and a positive integer n.

(2) For any prime p and any positive integer n there is a field of order p^n .

(3) Any two fields of order p^n are isomorphic. This field of order p^n is frequently referred to as the Galois field of order p^n and it is designated by $GF(p^n)$.

Evariste Galois (1811-32) was a pioneer in the field of abstract algebra.

A French stamp honoring Evariste Galois (1811-32)

Theorem 16.2.4 tells us that there is a field of order $2^2 = 4$, and there is only one such field up to isomorphism. That is, all such fields of order 4 are isomorphic to F, which we constructed in Example 16.2.2.

EXERCISES FOR SECTION 16.2

A Exercises

1. Write out the addition, multiplication, and "inverse" tables for each of the following fields'.

(a) $[\mathbb{Z}_2, +_2, \times_2]$

(b) $[\mathbb{Z}_3, +_3, \times_3]$

(c) $[\mathbb{Z}_5, +_5, \times_5]$

2. Show that the set of units of the fields in Exercise 1 form a group under the operation of the multiplication of the given field. Recall that a unit is an element which has a multiplicative inverse.

3. Complete the argument in Example 16.2.2 to show that if $[F, +]$ is isomorphic to $\mathbb{Z}_4$, then F would have a zero divisor.

4. Write out the operation tables for ${\mathbb{Z}_2}^2$. Is ${\mathbb{Z}_2}^2$ a ring? An integral domain? A field? Explain.

5. Determine all values x from the given field that satisfy the given equation:

(a) $x + 1 = -1$ over $\mathbb{Z}_2$, $\mathbb{Z}_3$ and $\mathbb{Z}_5$

(b) $2x + 1 = 2$ over $\mathbb{Z}_3$ and $\mathbb{Z}_5$

(c) $3x + 1 = 2$ over $\mathbb{Z}_5$

6. (a) Prove that if p and q are prime, then $\mathbb{Z}_p \times \mathbb{Z}_q$, is never a field.

(b) Can ${\mathbb{Z}_p}^n$ be a field for any prime p and any positive integer $n \geq 2$?

7. The following are equations over $\mathbb{Z}_2$. Their coefficients come solely from $\mathbb{Z}_2$. Determine all solutions over $\mathbb{Z}_2$; that is, find all numbers in $\mathbb{Z}_2$ that satisfy the equations:

(a) $x^2 + x = 0$

(b) $x^2 + 1 = 0$

(c) $x^3 + x^2 + x + 1 = 0$

(d) $x^3 + x + 1 = 0$

8. Determine the number of different fields, if any, of all orders 2 through 15. Wherever possible, describe these fields via a known field.

B Exercise

9. Let $\mathbb{Q}\left(\sqrt{2}\right) = \left\{a + b\sqrt{2} \;\middle|\; a, b \in \mathbb{Q}\right\}$.

(a) Prove that $\left[\mathbb{Q}\left(\sqrt{2}\right), +, \cdot\right]$ is a field.

(b) Show that $\mathbb{Q}$ is a subfield of $\mathbb{Q}\left(\sqrt{2}\right)$. For this reason, $\mathbb{Q}\left(\sqrt{2}\right)$ is called an extension field of $\mathbb{Q}$.

(c) Show that all the roots of the equation $x^2 - 2 = 0$ lie in the extension field $\mathbb{Q}\left(\sqrt{2}\right)$.

(d) Do the roots of the equation $x^2 - 3 = 0$ lie in this field? Explain.

16.3 Polynomial Rings

In the previous sections we examined the solutions of a few equations over different rings and fields. To solve the equation $x^2 - 2 = 0$ over the field of the real numbers means to find all solutions of this equation that are in this particular field $\mathbb{R}$. This statement can be replaced as follows: Determine all $a \in \mathbb{R}$ such that the polynomial $f(x) = x^2 - 2$ is equal to zero when evaluated at $x = a$. In this section, we will concentrate on the theory of polynomials. We will develop concepts using the general setting of polynomials over rings since results proven over rings are true for fields (and integral domains). The reader should keep in mind that in most cases we are just formalizing concepts that he or she learned in high school algebra over the field of reals.

Definition: Polynomial over R. *Let $[R, +, \cdot]$ be a ring. A polynomial, $f(x)$, over R is an expression of the form*

$$f(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \ , \ n \geq 0,$$

where $a_0, a_1, a_2, \ldots, a_n \in R$. If $a_n \neq 0$, then the degree of $f(x)$ is n, If $f(x) = 0$, then the degree of $f(x)$ is undefined and we assign the value $-\infty$ to the degree. If the degree of $f(x)$ is n, we write deg $f(x) = n$.

Comments:

(1) The symbol x is an object called an *indeterminate*, which is not an element of the ring R.

(2) The set of all polynomials in the indeterminate x with coefficients in R is denoted by $R[x]$.

(3) Note that $R \subseteq R[x]$, The elements of R are called *constant polynomials*, with the nonzero elements of R being the polynomials of degree 0.

(4) R is called the *ground ring* for $R[x]$.

(5) In the definition above, we have written the terms in increasing degree starting with the constant. The ordering of terms can be reversed without changing the polynomial. For example, $1 + 2x - 3x^4$ and $-3x^4 + 2x + 1$ are the same polynomial.

(6) A term of the form x^k in a polynomial is understood to be $1 x^k$,

Example 16.3.1. $f(x) = 3$, $g(x) = 2 - 4x + 7x^2$, and $h(x) = 2 + x^4$ are all polynomials in $\mathbb{Z}[x]$. Their degrees are 0, 2, and 4, respectively.

Addition and multiplication of polynomials are performed as in high school algebra. However, we must do our computations in the ground ring over which we are considering the polynomials.

Example 16.3.2. In $\mathbb{Z}_3[x]$, if $f(x) = 1 + x$ and $g(x) = 2 + x$, then

$$\begin{aligned} f(x) + g(x) &= (1 + x) + (2 + x) \\ &= (1 +_3 2) + (1 +_3 1)x \\ &= 0 + 2x \\ &= 2x \end{aligned}$$

and

$$\begin{aligned} f(x)\,g(x) &= (1 + x) \cdot (2 + x) \\ &= (1 + x) \cdot 2 + (1 + x) \cdot x \\ &= 1 \times_3 2 + 2x + 1x + x \cdot x \\ &= 2 + (2 +_3 1)x + x^2 \\ &= 2 + x^2 \end{aligned} \quad .$$

However, for the same polynomials as above, $f(x)$ and $g(x)$ in $\mathbb{Z}$[x], we have

$$\begin{aligned} f(x) + g(x) &= (1 + x) + (2 + x) \\ &= (1 + 2) + (1 + 1)x \\ &= 3 + 2x \end{aligned}$$

and

$$\begin{aligned} f(x)\,g(x) &= (1 + x) \cdot (2 + x) \\ &= (1 + x) \cdot 2 + (1 + x) \cdot x \\ &= 1 \cdot 2 + 2x + 1x + x \cdot x \\ &= 2 + (2 + 1)x + x^2 \\ &= 2 + 3x + x^2 \end{aligned} \quad .$$

The important fact to keep in mind is that addition and multiplication in $R[x]$ depends on addition and multiplication in R. The x's merely serve the purpose of "place holders." All computations are done over the given ring. We summarize in the following theorem:

Theorem 16.3.1. *Let $[R, +, \cdot]$ be a ring. Then:*

(1) R[x] is a ring under the operations of polynomial addition and multiplication, which depend on (are induced by) the operations in R.

(2) If R is a commutative ring, then R[x] is a commutative ring.

(3) If R is a ring with unity, 1, *then R[x] is a ring with unity (the unity in R[x] is* $1 + 0x + 0x^2 + \cdots$*).*

(4) If R is an integral domain, then R[x] is an integral domain.

(5) If F is a field, then F[x] is not a field. However, F[x] is an integral domain.

The proofs for Parts 1 through 4 are not difficult but rather long, so we omit them. For those inclined to prove them, we include the formal definitions of addition and multiplication in $R[x]$ below.

Proof Of Part 5: $F[x]$ is not a field since for $x \in F[x]$, $x^{-1} = 1/x \notin F[x]$. Hence not all nonzero elements in $F[x]$ have multiplicative inverses in $F[x]$. Every field F is an integral domain. By Part 4, $F[x]$ is an integral domain. ■

Definition: Addition in R[x]. *Let* $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$ *and* $g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n$ *be elements in R[x] so that* $a_i \in R$ *and* $b_i \in R$ *for all i. Let k be the maximum of m and n. Then*

$$f(x) + g(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_k x^k$$

where $c_i = a_i + b_i$ *for* $i = 0, 1, 2, \ldots, k$.

Definition: Multiplication in R[x]. *Let* $f(x)$ *and* $g(x)$ *be as above. Then*

$$f(x) \cdot g\{x) = d_0 + d_1 x + d_2 x^2 + \cdots + d_p x^p \text{ where}$$

$p = m + n$, *and*

$$d_s = \sum_{i=0}^{s} a_i b_{s-i}$$
$$= a_0 b_s + a_1 b_{s-1} + a_2 b_{s-2} + \cdots + a_{s-1} b_1 + a_s b_0$$

for $0 \le s \le p$.

Example 16.3.3. Let $f(x) = 2 + x^2$ and $g(x) = -1 + 4x + 3x^2$. We will compute $f(x) \cdot g(x)$ in $\mathbb{Z}[x]$. Of course this product can be obtained by the usual methods of high school algebra. We will, for illustrative purposes, use the above definition. Using the notation of the above definition, $a_0 = 2$, $a_1 = 0$, $a_2 = 1$, $b_0 = -1$, $b_1 = 4$, and $b_2 = 3$. We want to compute the coefficients d_0, d_1, d_2, d_3, and d_4 . We will compute d_3 , the coefficient of the x^3 term of the product, and leave the remainder to the reader (see Exercise 2 of this section). Since the degrees of both factors is 2, $a_i = b_i = 0$ for $i \ge 3$.

$$d_3 = a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0$$
$$= 2 \cdot 0 + 0 \cdot 3 + 1 \cdot 4 + 0 \cdot (-1) = 4$$

From high school algebra we all learned the standard procedure for dividing a polynomial $f(x)$ by a second polynomial $g(x)$. This process of polynomial long division is referred to as the division property for polynomials. Under this scheme we continue to divide until the result is a quotient $q(x)$ and a remainder $r(x)$ whose degree is strictly less than that of the divisor $g(x)$. This property is valid over any field.

Example 16.3.4. Let $f(x) = 1 + x + x^3$ and $g(x) = 1 + x$ be two polynomials in $\mathbb{Z}_2$[x]. Let us divide $f(x)$ by $g(x)$. Keep in mind that we are in $\mathbb{Z}_2[x]$ and that, in particular, $-1 = 1$ in $\mathbb{Z}_2$. This is a case where reordering the terms in decreasing degree is preferred.

$$\begin{array}{r} x^2 + x \\ x+1 \;\overline{)}\; x^3 + 0x^2 + x + 1 \\ \underline{x^3 + x^2} \\ x^2 + x + 1 \\ \underline{x^2 + x} \\ 1 \end{array}$$

Therefore,

$$\frac{x^3 + x + 1}{x+1} = x^2 + x + \frac{1}{x+1}$$

or equivalently,

$$x^3 + x + 2 = (x^2 + x) \cdot (x + 1) + 1$$

That is $f(x) = g(x) \cdot q(x) + r(x)$ where $q(x) = x^2 + x$ and $r(x) = 1$. Notice that $\deg(r(x)) = 0$, which is strictly less than the $\deg(g(x)) = 1$.

Example 16.3.5. Let $f(x) = 1 + x^4$ and $g(x) = 1 + x$ be polynomials in $\mathbb{Z}_2[x]$. Let us divide f(x) by g(x):

$$
\begin{array}{r}
x^3 + x^2 + x + 1 \\
x+1 \overline{)\, x^4 + 0x^3 + 0x^2 + 0x + 1} \\
\underline{x^4 + x^3} \\
x^3 + 1 \\
\underline{x^3 + x^2} \\
x^2 + 1 \\
\underline{x^2 + x} \\
x + 1 \\
\underline{x + 1} \\
0
\end{array}
$$

Thus $x^4 + 1 = (x^3 + x^2 + x + 1)(x + 1)$.

Since we have 0 as a remainder, $x + 1$ must be a factor of $x^4 + 1$, as in high school algebra. Also, since $x + 1$ is a factor of $x^4 + 1$, 1 is a zero (or root) of $x^4 + 1$. Of course we could have determined that 1 is a root of $f(x)$ simply by computing $f(1) = 1^4 +_2 1 = 1 +_2 1 = 0$.

Before we summarize the main results suggested by the previous examples, we should probably consider what could have happened if we had performed divisions of polynomials in the ring $\mathbb{Z}[x]$ rather than over the field $\mathbb{Z}_2$. For example, $f(x) = x^2 - 1$ and $g(x) = 2x - 2$ are both elements of the ring $\mathbb{Z}[x]$, yet

$$\frac{x^2+1}{2x-1} = \frac{1}{2}x + \frac{1}{2}$$

The quotient is not a polynomial over $\mathbb{Z}$ but a polynomial over the field $\mathbb{Q}$. For this reason it would be wise to describe all results over a field F rather than over an arbitrary ring R.

Theorem 16.3.2. Division Property for F[x]. *Let $[F, +, \cdot]$ be a field and let $f(x)$ and $g(x)$ be two elements of $F[x]$ with $g(x) \neq 0$. Then there exist unique polynomials $q(x)$ and $r(x)$ in $F[x]$ such that $f(x) = g(x)q(x) + r(x)$, where deg $r(x) <$ deg $g(x)$.*

Theorem 16.3.2 can be proven by induction on deg $f(x)$.

Theorem 16.3.3. *Let $[F, +, \cdot]$ be a field. An element $a \in F$ is a zero of $f(x) \in F[x]$ if and only if $x - a$ is a factor of $f(x)$ in $F[x]$.*

Proof: ($\Rightarrow$) Assume that $a \in F$ is a zero of $f(x) \in F[x]$. We wish to show that $x - a$ is a factor of $f(x)$. To do so, apply the division property to $f(x)$ and $g(x) = x - a$. Hence, there exist unique polynomials $q(x)$ and $r(x)$ from $F[x]$ such that $f(x) = (x - a)\cdot q(x) + r(x)$ and the deg $r(x) <$ deg $(x - a) = 1$, so $r(x) = c \in F$, that is, $r(x)$ is a constant. Also a is a zero of $f(x)$ mean $f(a) = 0$. So $f(x) = (x - a)\cdot q(x) + c$ becomes $0 = f(a) = (a - a)q(a) + c$. Hence $c = 0$, so $f(x) = (x - a)\cdot q(x)$, and $x - a$ is a factor of $f(x)$. The reader should note that a critical point of the proof of this half of the theorem was the part of the division property that stated that deg $r(x) <$ deg $g(x)$.

($\Leftarrow$) We leave this half to the reader, exercise 6. ■

Theorem 16.3.4. *A nonzero polynomial $f(x) \in F[x]$ of degree n can have at most n zeros.*

Proof: Let $a \in F$ be a zero of $f(x)$. Then $f(x) = (x - a)\cdot q_1(x)$, $q_1(x) \in F[x]$, by Theorem 16.3.3. If $b \in F$ is a zero of $q_1(x)$, then again by Theorem 16.3.3, $f(x) = (x - a)(x - b)q_2(x)$, $q_2(x) \in F[x]$. Continue this process, which must terminate in at most n steps since the degree of $q_k(x)$ would be $n - k$. ■

From Theorem 16.3.3 we can obtain yet another insight into the problems associated with solving polynomial equations; that is, finding the zeros of a polynomial. The theorem states that an element $a \in F$ is a zero of $f(x) \in F[x]$ if and only if $x - a$ is a factor of $f(x)$. The initial important idea here is that the zero a is from the ground field F. Second, a is a zero only if $(x - a)$ is a factor of $f(x)$ in $F[x]$ —that is, only when $f(x)$ can be factored (or reduced) to the product of $(x - a)$ times some other polynomial in $F[x]$.

Example 16.3.6. Consider the polynomial $f(x) = x^2 - 2$ taken as being in $\mathbb{Q}[x]$. From high school algebra we know that $f(x)$ has two zeros (or roots), namely $\pm\sqrt{2}$, and $x^2 - 2$ can be factored (reduced) as $\left(x - \sqrt{2}\right)\left(x + \sqrt{2}\right)$. However, we are working in $\mathbb{Q}[x]$, these two factors are not in the set of polynomials over the rational numbers, $\mathbb{Q}$ since $\sqrt{2} \notin \mathbb{Q}$. Therefore, $x^2 - 2$ does not have a zero in $\mathbb{Q}$ since it cannot be factored over $\mathbb{Q}$. When this happens, we say that the polynomial is irreducible over $\mathbb{Q}$.

The problem of factoring polynomials is tied hand-in-hand with that of the reducibility of polynomials. We give a precise definition of this concept.

Definition: Irreducible over F. *Let$[F, +, \cdot]$ be a field and let f(x) $\in$ F[x] be a nonconstant polynomial, f(x)) is irreducible over F if and only if f(x) cannot be expressed as a product of two (or more) polynomials, both from F[x] and both of degree lower than that of f(x).*

A polynomial is *reducible over F* if it is not irreducible over F.

Example 16.3.7. The polynomial $f(x) = x^4 + 1$ of Example 16.3.5 is reducible over $\mathbb{Z}_2$ since $x^4 + 1 = (x + 1)(x^3 + x^2 + x - 1)$.

Example 16.3.8. Is the polynomial $f(x) = x^3 + x + 1$ of Example 16.3.4 reducible over $\mathbb{Z}_2$? From Example 16.3.4 we know that $x + 1$ is not a factor of $x^3 + x + 1$, and from high school algebra we realize that a cubic (also second-degree) polynomial is reducible if and only if it has a linear (first-degree) factor. (Why?) Does $f(x) = x^3 + x + 1$ have any other linear factors? Theorem 16.3.1 gives us a quick way of determine this since $x - a$ is a factor of $x^3 + x + 1$ over $\mathbb{Z}_2$ if and only if $a \in \mathbb{Z}_2$ is a zero of $x^3 + x + 1$. So $x^3 + x + 1$ is reducible over $\mathbb{Z}_2$ if and only if it has a zero in $\mathbb{Z}_2$. Since $\mathbb{Z}_2$ has only two elements, 0 and 1, this is easy enough to check.

$f(0) = 0^3 +_2 0 +_2 1 = 1$ and

$f(1) = 1^3 +_2 1 +_2 1 = 1$

so neither 0 nor 1 is a zero of $f(x)$ over $\mathbb{Z}_2$. Hence, $x^3 + x + 1$ is irreducible over $\mathbb{Z}_2$.

From high school algebra we know that $x^3 + x + 1$ has three zeros from some field. Can we find this field? To be more precise, can we construct (find) the field which contains $\mathbb{Z}_2$ and all zeros of $x^3 + x + 1$? We will consider this task in the next section.

We close this section with a final analogy. Prime numbers play an important role in mathematics. The concept of irreducible polynomials (over a field) is analogous to that of a prime number. Just think of the definition of a prime number. A useful fact concerning primes is: If p is a prime and if $p \mid a\,b$, then $p \mid a$ or $p \mid b$. We leave it to the reader to think about the veracity of the following: If $p(x)$ is an irreducible polynomial over F, $a(x)$, $b(x) \in F[x]$ and $p(x) \mid a(x)\,b(x)$, then $p(x) \mid a(x)$ or $p(x) \mid b(x)$.

EXERCISES FOR SECTION 16.3

A Exercises

1. Let $f(x) = 1 + x$ and $g(x) = 1 + x + x^2$. Compute the following sums and products in the indicated rings.
   (a) $f(x) + g(x)$ and $f(x) \cdot g(x)$ in $\mathbb{Z}[x]$
   (b) $f(x) + g(x)$ and $f(x) \cdot g(x)$ in $\mathbb{Z}_2$[x]
   (c) $(f(x) \cdot g(x)) \cdot f(x)$ in $\mathbb{Z}$[x]
   (d) $(f(x) \cdot g(x)) \cdot f(x)$ in $\mathbb{Z}_2[x]$
   (e) $f(x) \cdot f(x) + f(x) \cdot g(x)$ in $\mathbb{Z}_2[x]$
2. Complete Example 16.3.3.
3. Prove that:
   (a) The ring $\mathbb{R}$ is a subring of the ring $\mathbb{R}[x]$.
   (b) The ring $\mathbb{Z}[x]$ is a subring of the $\mathbb{Q}[x]$.
   (c) The ring $\mathbb{Q}[x]$ is a subring of the ring $\mathbb{R}[x]$.
4. (a) Find all zeros of $x^4 + 1$ in $\mathbb{Z}_3$. (b) Find all zeros of $x^5 + 1$ in $\mathbb{Z}_5$.
5. Determine which of the following are reducible over $\mathbb{Z}_2$. Explain.
   (a) $f(x) = x^3 + 1$
   (b) $g(x) = x^3 + x^2 + x$.
   (c) $h(x) = x^3 + x^2 + 1$.
   (d) $k(x) = x^4 + x^2 + 1$. (Be careful.)
6. Prove the second half of Theorem 16.3.3.
7. Give an example of the contention made in the last paragraph of this section.
8. Determine all zeros of $x^4 + 3x^3 + 2x + 4$ in $\mathbb{Z}_5[x]$
9. Show that $x^2 - 3$ is irreducible over $\mathbb{Q}$ but reducible over the field of real numbers.

B Exercises

10. The definition of a vector space given in Chapter 13 holds over any field F, not just over the field of real numbers, where the elements of F are called scalars.
    (a) Show that $F[x]$ is a vector space over F.
    (b) Find a basis for $F[x]$ over F.
    (c) What is the dimension of F[x] over F?
11. Prove Theorem 16.3.2.
    (a) Show that the field $\mathbb{R}$ of real numbers is a vector space over $\mathbb{R}$. Find a basis for this vector space. What is dim $\mathbb{R}$ over $\mathbb{R}$?

(b) Repeat part a for an arbitrary field F.

(c) Show that $\mathbb{R}$ is a vector space over $\mathbb{Q}$.