3

Ring Theory

3.1 Definition and Examples of Rings

As we indicated in Chapter 2, there are certain algebraic systems which serve as the building blocks for the structures comprising the subject which is today called modern algebra. At this stage of the development we have learned something about one of these, namely groups. It is our purpose now to introduce and to study a second such, namely rings. The abstract concept of a group has its origins in the set of mappings, or permutations, of a set onto itself. In contrast, rings stem from another and more familiar source, the set of integers. We shall see that they are patterned after, and are generalizations of, the algebraic aspects of the ordinary integers.

In the next paragraph it will become clear that a ring is quite different from a group in that it is a two-operational system; these operations are usually called addition and multiplication. Yet, despite the differences, the analysis of rings will follow the pattern already laid out for groups. We shall require the appropriate analogs of homomorphism, normal subgroups, factor groups, etc. With the experience gained in our study of groups we shall be able to make the requisite definitions, intertwine them with meaningful theorems, and end up proving results which are both interesting and important about mathematical objects with which we have had long acquaintance. To cite merely one instance, later on in the book, using the tools developed here, we shall prove that it is impossible to trisect an angle of 60° using only a straight-edge and compass.

DEFINITION A nonempty set R is said to be an *associative ring* if in R there are defined two operations, denoted by $+$ and $\cdot$ respectively, such that for all a, b, c in R:

1. $a + b$ is in R.
2. $a + b = b + a$.
3. $(a + b) + c = a + (b + c)$.
4. There is an element 0 in R such that $a + 0 = a$ (for every a in R).
5. There exists an element $-a$ in R such that $a + (-a) = 0$.
6. $a \cdot b$ is in R.
7. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
8. $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$ (the two distributive laws).

Axioms 1 through 5 merely state that R is an abelian group under the operation $+$, which we call addition. Axioms 6 and 7 insist that R be closed under an associative operation $\cdot$, which we call multiplication. Axiom 8 serves to interrelate the two operations of R.

Whenever we speak of ring it will be understood we mean associative ring. Nonassociative rings, that is, those in which axiom 7 may fail to hold, do occur in mathematics and are studied, but we shall have no occasion to consider them.

It may very well happen, or not happen, that there is an element 1 in R such that $a \cdot 1 = 1 \cdot a = a$ for every a in R; if there is such we shall describe R as a *ring with unit element.*

If the multiplication of R is such that $a \cdot b = b \cdot a$ for every a, b in R, then we call R a *commutative ring.*

Before going on to work out some properties of rings, we pause to examine some examples. Motivated by these examples we shall define various special types of rings which are of importance.

Example 3.1.1 R is the set of integers, positive, negative, and 0; $+$ is the usual addition and $\cdot$ the usual multiplication of integers. R is a commutative ring with unit element.

Example 3.1.2 R is the set of even integers under the usual operations of addition and multiplication. R is a commutative ring but has no unit element.

Example 3.1.3 R is the set of rational numbers under the usual addition and multiplication of rational numbers. R is a commutative ring with unit element. But even more than that, note that the elements of R different from 0 form an abelian group under multiplication. A ring with this latter property is called a *field.*

Example 3.1.4 R is the set of integers mod 7 under the addition and multiplication mod 7. That is, the elements of R are the seven symbols $\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}$, where

1. $\bar{i} + \bar{j} = \bar{k}$ where k is the remainder of $i + j$ on division by 7 (thus, for instance, $\bar{4} + \bar{5} = \bar{2}$ since $4 + 5 = 9$, which, when divided by 7, leaves a remainder of 2).
2. $\bar{i} \cdot \bar{j} = \bar{m}$ where m is the remainder of ij on division by 7 (thus, $\bar{5} \cdot \bar{3} = \bar{1}$ since $5 \cdot 3 = 15$ has 1 as a remainder on division by 7).

The student should verify that R is a commutative ring with unit element. However, much more can be shown; namely, since

$$\bar{1} \cdot \bar{1} = \bar{1} = \bar{6} \cdot \bar{6},$$
$$\bar{2} \cdot \bar{4} = \bar{1} = \bar{4} \cdot \bar{2},$$
$$\bar{3} \cdot \bar{5} = \bar{1} = \bar{5} \cdot \bar{3},$$

the nonzero elements of R form an abelian group under multiplication. R is thus a field. Since it only has a finite number of elements it is called a *finite field.*

Example 3.1.5 R is the set of integers mod 6 under addition and multiplication mod 6. If we denote the elements in R by $\bar{0}, \bar{1}, \bar{2}, \ldots, \bar{5}$, one sees that $\bar{2} \cdot \bar{3} = \bar{0}$, yet $\bar{2} \neq \bar{0}$ and $\bar{3} \neq \bar{0}$. *Thus it is possible in a ring R that $a \cdot b = 0$ with neither $a = 0$ nor $b = 0$.* This cannot happen in a field (see Problem 10, end of Section 3.2), thus the ring R in this example is certainly not a field.

Every example given so far has been a commutative ring. We now present a noncommutative ring.

Example 3.1.6 R will be the set of all symbols

$$\alpha_{11}e_{11} + \alpha_{12}e_{12} + \alpha_{21}e_{21} + \alpha_{22}e_{22} = \sum_{i,j=1}^{2} \alpha_{ij}e_{ij},$$

where all the α_{ij} are rational numbers and where we decree

$$\sum_{i,j=1}^{2} \alpha_{ij}e_{ij} = \sum_{i,j=1}^{2} \beta_{ij}e_{ij} \tag{1}$$

if and only if for all $i, j = 1, 2$, $\alpha_{ij} = \beta_{ij}$,

$$\sum_{i,j=1}^{2} \alpha_{ij}e_{ij} + \sum_{i,j=1}^{2} \beta_{ij}e_{ij} = \sum_{i,j=1}^{2} (\alpha_{ij} + \beta_{ij})e_{ij}. \tag{2}$$

$$\left(\sum_{i,j=1}^{2} \alpha_{ij}e_{ij}\right) \cdot \left(\sum_{i,j=1}^{2} \beta_{ij}e_{ij}\right) = \sum_{i,j=1}^{2} \gamma_{ij}e_{ij}, \tag{3}$$

where

$$\gamma_{ij} = \sum_{v=1}^{2} \alpha_{iv}\beta_{vj} = \alpha_{i1}\beta_{1j} + \alpha_{i2}\beta_{2j}.$$

This multiplication, when first seen, looks rather complicated. However, it is founded on relatively simple rules, namely, multiply $\sum \alpha_{ij}e_{ij}$ by $\sum \beta_{ij}e_{ij}$ formally, multiplying out term by term, and collecting terms, and using the relations $e_{ij} \cdot e_{kl} = 0$ for $j \neq k$, $e_{ij} \cdot e_{jl} = e_{il}$ in this term-by-term collecting. (Of course those of the readers who have already encountered some linear algebra will recognize this example as the ring of all 2×2 matrices over the field of rational numbers.)

To illustrate the multiplication, if $a = e_{11} - e_{21} + e_{22}$ and $b = e_{22} + 3e_{12}$, then

$$\begin{aligned} a \cdot b &= (e_{11} - e_{21} + e_{22}) \cdot (e_{22} + 3e_{12}) \\ &= e_{11} \cdot e_{22} + 3e_{11} \cdot e_{12} - e_{21} \cdot e_{22} - 3e_{21} \cdot e_{12} + e_{22} \cdot e_{22} + 3e_{22} \cdot e_{12} \\ &= 0 + 3e_{12} - 0 - 3e_{22} + e_{22} + 0 \\ &= 3e_{12} - 3e_{22} + e_{22} = 3e_{12} - 2e_{22}. \end{aligned}$$

Note that $e_{11} \cdot e_{12} = e_{12}$ whereas $e_{12} \cdot e_{11} = 0$. Thus the multiplication in R is not commutative. Also it is possible for $u \cdot v = 0$ with $u \neq 0$ and $v \neq 0$.

The student should verify that R is indeed a ring. It is called the ring of 2×2 rational matrices. It, and its relative, will occupy a good deal of our time later on in the book.

Example 3.1.7 Let C be the set of all symbols (α, β) where α, β are real numbers. We define

$$(\alpha, \beta) = (\gamma, \delta) \text{ if and only if } \alpha = \gamma \text{ and } \beta = \delta. \tag{1}$$

In C we introduce an addition by defining for $x = (\alpha, \beta)$, $y = (\gamma, \delta)$

$$x + y = (\alpha, \beta) + (\gamma, \delta) = (\alpha + \gamma, \beta + \delta). \tag{2}$$

Note that $x + y$ is again in C. We assert that C is an abelian group under this operation with $(0, 0)$ serving as the identity element for addition, and $(-\alpha, -\beta)$ as the inverse, under addition, of (α, β).

Now that C is endowed with an addition, in order to make of C a ring we still need a multiplication. We achieve this by defining

$$\text{for } X = (\alpha, \beta), \qquad Y = (\gamma, \delta) \text{ in } C,$$

$$X \cdot Y = (\alpha, \beta) \cdot (\gamma, \delta) = (\alpha\gamma - \beta\delta, \alpha\delta + \beta\gamma). \tag{3}$$

Note that $X \cdot Y = Y \cdot X$. Also $X \cdot (1, 0) = (1, 0) \cdot X = X$ so that $(1, 0)$ is a unit element for C.

Again we notice that $X \cdot Y \in C$. Also, if $X = (\alpha, \beta) \neq (0, 0)$ then, since α, β are real and not both 0, $\alpha^2 + \beta^2 \neq 0$; thus

$$Y = \left(\frac{\alpha}{\alpha^2 + \beta^2}, \frac{-\beta}{\alpha^2 + \beta^2}\right)$$

is in C. Finally we see that

$$(\alpha, \beta) \cdot \left(\frac{\alpha}{\alpha^2 + \beta^2}, \frac{-\beta}{\alpha^2 + \beta^2}\right) = (1, 0).$$

All in all we have shown that C is a field. If we write (α, β) as $\alpha + \beta i$, the reader may verify that C is merely a disguised form of the familiar complex numbers.

Example 3.1.8 This last example is often called the ring of *real quaternions.* This ring was first described by the Irish mathematician Hamilton. Initially it was extensively used in the study of mechanics; today its primary interest is that of an important example, although it still plays key roles in geometry and number theory.

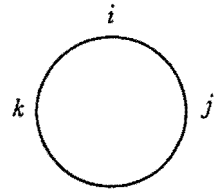
Let Q be the set of all symbols $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, where all the numbers α_0, α_1, α_2, and α_3 are real numbers. We declare two such symbols, $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ and $\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k$, to be equal if and only if $\alpha_t = \beta_t$ for $t = 0, 1, 2, 3$. In order to make Q into a ring we must define a $+$ and a $\cdot$ for its elements. To this end we define

1. For any $X = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, $Y = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 k$ in Q, $X + Y = (\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k) + (\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k) = (\alpha_0 + \beta_0) + (\alpha_1 + \beta_1)i + (\alpha_2 + \beta_2)j + (\alpha_3 + \beta_3)k$

and

2. $X \cdot Y = (\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k) \cdot (\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k) = (\alpha_0\beta_0 - \alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3) + (\alpha_0\beta_1 + \alpha_1\beta_0 + \alpha_2\beta_3 - \alpha_3\beta_2)i + (\alpha_0\beta_2 + \alpha_2\beta_0 + \alpha_3\beta_1 - \alpha_1\beta_3)j + (\alpha_0\beta_3 + \alpha_3\beta_0 + \alpha_1\beta_2 - \alpha_2\beta_1)k.$

Admittedly this formula for the product seems rather formidable; however, it looks much more complicated than it actually is. It comes from multiplying out two such symbols formally and collecting terms using the relations $i^2 = j^2 = k^2 = ijk = 1$, $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$. The latter part of these relations, called the multiplication table of the quaternion units, can be remembered by the little diagram on page 125. As you go around clockwise you read off the product, e.g., $ij = k$, $jk = i$, $ki = j$; while going around counterclockwise you read off the negatives.



Notice that the elements $\pm 1, \pm i, \pm j, \pm k$ form a non-abelian group of order 8 under this product. In fact, this is the group we called the group of quaternion units in Chapter 2.

The reader may prove that Q is a noncommutative ring in which $0 = 0 + 0i + 0j + 0k$ and $1 = 1 + 0i + 0j + 0k$ serve as the zero and unit elements respectively. Now if $X = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ is not 0, then not all of $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ are 0; since they are real, $\beta = \alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 \neq 0$ follows. Thus

$$Y = \frac{\alpha_0}{\beta} - \frac{\alpha_1}{\beta} i - \frac{\alpha_2}{\beta} j - \frac{\alpha_3}{\beta} k \in Q.$$

A simple computation now shows that $X \cdot Y = 1$. Thus the nonzero elements of Q form a non-abelian group under multiplication. A ring in which the nonzero elements form a group is called a *division ring* or *skew-field*. Of course, a commutative division ring is a field. Q affords us a division ring which is not a field. Many other examples of noncommutative division rings exist, but we would be going too far afield to present one here. The investigation of the nature of division rings and the attempts to classify them form an important part of algebra.

3.2 Some Special Classes of Rings

The examples just discussed in Section 3.1 point out clearly that although rings are a direct generalization of the integers, certain arithmetic facts to which we have become accustomed in the ring of integers need not hold in general rings. For instance, we have seen the possibility of $a \cdot b = 0$ with neither a nor b being zero. Natural examples exist where $a \cdot b \neq b \cdot a$. All these run counter to our experience heretofore.

For simplicity of notation we shall henceforth drop the dot in $a \cdot b$ and merely write this product as ab.

DEFINITION If R is a commutative ring, then $a \neq 0 \in R$ is said to be a *zero-divisor* if there exists a $b \in R$, $b \neq 0$, such that $ab = 0$.

DEFINITION A commutative ring is an *integral domain* if it has no zero-divisors.

The ring of integers, naturally enough, is an example of an integral domain.

DEFINITION A ring is said to be a *division ring* if its nonzero elements form a group under multiplication.

The unit element under multiplication will be written as 1, and the inverse of an element a under multiplication will be denoted by a^{-1}.

Finally we make the definition of the ultra-important object known as a field.

DEFINITION A *field* is a commutative division ring.

In our examples in Section 3.1, we exhibited the noncommutative division ring of real quaternions and the following fields: the rational numbers, complex numbers, and the integers mod 7. Chapter 5 will concern itself with fields and their properties.

We wish to be able to compute in rings in much the same manner in which we compute with real numbers, keeping in mind always that there are differences—it may happen that $ab \neq ba$, or that one cannot divide. To this end we prove the next lemma, which asserts that certain things we should like to be true in rings are indeed true.

LEMMA 3.2.1 *If R is a ring, then for all $a, b \in R$*

1. $a0 = 0a = 0$.
2. $a(-b) = (-a)b = -(ab)$.
3. $(-a)(-b) = ab$.

If, in addition, R has a unit element 1, *then*

4. $(-1)a = -a$.
5. $(-1)(-1) = 1$.

Proof.

1. If $a \in R$, then $a0 = a(0 + 0) = a0 + a0$ (using the right distributive law), and since R is a group under addition, this equation implies that $a0 = 0$.

   Similarly, $0a = (0 + 0)a = 0a + 0a$, using the left distributive law, and so here too, $0a = 0$ follows.
2. In order to show that $a(-b) = -(ab)$ we must demonstrate that $ab + a(-b) = 0$. But $ab + a(-b) = a(b + (-b)) = a0 = 0$ by use of

the distributive law and the result of part 1 of this lemma. Similarly $(-a)b = -(ab)$.

3. That $(-a)(-b) = ab$ is really a special case of part 2; we single it out since its analog in the case of real numbers has been so stressed in our early education. So on with it:

$$\begin{aligned}(-a)(-b) &= -(a(-b)) \quad \text{(by part 2)}\\ &= -(-(ab)) \quad \text{(by part 2)}\\ &= ab\end{aligned}$$

since $-(-x) = x$ is a consequence of the fact that in any group $(u^{-1})^{-1} = u$.

4. Suppose that R has a unit element 1; then $a + (-1)a = 1a + (-1)a = (1 + (-1))a = 0a = 0$, whence $(-1)a = -a$. In particular, if $a = -1$, $(-1)(-1) = -(-1) = 1$, which establishes part 5.

With this lemma out of the way we shall, from now on, feel free to compute with negatives and 0 as we always have in the past. The result of Lemma 3.2.1 is our permit to do so. For convenience, $a + (-b)$ will be written $a - b$.

The lemma just proved, while it is very useful and important, is not very exciting. So let us proceed to results of greater interest. Before we do so, we enunciate a principle which, though completely trivial, provides a mighty weapon when wielded properly. This principle says no more or less than the following: if a postman distributes 101 letters to 100 mailboxes then some mailbox must receive at least two letters. It does not sound very promising as a tool, does it? Yet it will surprise us! Mathematical ideas can often be very difficult and obscure, but no such argument can be made against this very simple-minded principle given above. We formalize it and even give it a name.

THE PIGEONHOLE PRINCIPLE *If n objects are distributed over m places, and if $n > m$, then some place receives at least two objects.*

An equivalent formulation, and one which we shall often use is: If n objects are distributed over n places in such a way that no place receives more than one object, then each place receives *exactly* one object.

We immediately make use of this idea in proving

LEMMA 3.2.2 *A finite integral domain is a field.*

Proof. As we may recall, an integral domain is a commutative ring such that $ab = 0$ if and only if at least one of a or b is itself 0. A field, on the other hand, is a commutative ring with unit element in which every non-zero element has a multiplicative inverse in the ring.

Let D be a finite integral domain. In order to prove that D is a field we must

1. Produce an element $1 \in D$ such that $a1 = a$ for every $a \in D$.
2. For every element $a \neq 0 \in D$ produce an element $b \in D$ such that $ab = 1$.

Let $x_1, x_2, \ldots, x_n$ be all the elements of D, and suppose that $a \neq 0 \in D$. Consider the elements $x_1a, x_2a, \ldots, x_na$; they are all in D. We claim that they are all distinct! For suppose that $x_ia = x_ja$ for $i \neq j$; then $(x_i - x_j)a = 0$. Since D is an integral domain and $a \neq 0$, this forces $x_i - x_j = 0$, and so $x_i = x_j$, contradicting $i \neq j$. Thus $x_1a, x_2a, \ldots, x_na$ are n distinct elements lying in D, which has exactly n elements. By the pigeonhole principle these must account for all the elements of D; stated otherwise, every element $y \in D$ can be written as x_ia for some x_i. In particular, since $a \in D$, $a = x_{i_0}a$ for some $x_{i_0} \in D$. Since D is commutative, $a = x_{i_0}a = ax_{i_0}$. We propose to show that x_{i_0} acts as a unit element for every element of D. For, if $y \in D$, as we have seen, $y = x_ia$ for some $x_i \in D$, and so $yx_{i_0} = (x_ia)x_{i_0} = x_i(ax_{i_0}) = x_ia = y$. Thus x_{i_0} is a unit element for D and we write it as 1. Now $1 \in D$, so by our previous argument, it too is realizable as a multiple of a; that is, there exists a $b \in D$ such that $1 = ba$. The lemma is now completely proved.

COROLLARY *If p is a prime number then J_p, the ring of integers* mod p, *is a field.*

Proof. By the lemma it is enough to prove that J_p is an integral domain, since it only has a finite number of elements. If $a, b \in J_p$ and $ab \equiv 0$, then p must divide the ordinary integer ab, and so p, being a prime, must divide a or b. But then either $a \equiv 0 \bmod p$ or $b \equiv 0 \bmod p$, hence in J_p one of these is 0.

The corollary above assures us that we can find an infinity of fields having a finite number of elements. Such fields are called *finite fields*. The fields J_p do not give all the examples of finite fields; there are others. In fact, in Section 7.1 we give a complete description of all finite fields.

We point out a striking difference between finite fields and fields such as the rational numbers, real numbers, or complex numbers, with which we are more familiar.

Let F be a finite field having q elements (if you wish, think of J_p with its p elements). Viewing F merely as a group under addition, since F has q elements, by Corollary 2 to Theorem 2.4.1,

$$\underbrace{a + a + \cdots + a}_{q\text{-times}} = qa = 0$$

for any $a \in F$. Thus, in F, we have $qa = 0$ for some positive integer q, even if $a \neq 0$. This certainly cannot happen in the field of rational numbers, for instance. We formalize this distinction in the definitions we give below. In these definitions, instead of talking just about fields, we choose to widen the scope a little and talk about integral domains.

DEFINITION An integral domain D is said to be of *characteristic* 0 if the relation $ma = 0$, where $a \neq 0$ is in D, and where m is an integer, can hold only if $m = 0$.

The ring of integers is thus of characteristic 0, as are other familiar rings such as the even integers or the rationals.

DEFINITION An integral domain D is said to be of *finite characteristic* if there exists a *positive* integer m such that $ma = 0$ for all $a \in D$.

If D is of finite characteristic, then we define the *characteristic* of D to be the smallest positive integer p such that $pa = 0$ for all $a \in D$. It is not too hard to prove that if D *is of finite characteristic, then its characteristic is a prime number* (see Problem 6 below).

As we pointed out, any finite field is of finite characteristic. However, an integral domain may very well be infinite yet be of finite characteristic (see Problem 7).

One final remark on this question of characteristic: Why define it for integral domains, why not for arbitrary rings? The question is perfectly reasonable. Perhaps the example we give now points out what can happen if we drop the assumption "integral domain."

Let R be the set of all triples (a, b, c), where $a \in J_2$, the integers mod 2, $b \in J_3$, the integers mod 3, and c is any integer. We introduce a $+$ and a $\cdot$ to make of R a ring. We do so by defining $(a_1, b_1, c_1) + (a_2, b_2, c_2) = (a_1 + a_2, b_1 + b_2, c_1 + c_2)$ and $(a_1, b_1, c_1) \cdot (a_2, b_2, c_2) = (a_1a_2, b_1b_2, c_1c_2)$. It is easy to verify that R is a commutative ring. It is not an integral domain since $(1, 2, 0) \cdot (0, 0, 7) = (0, 0, 0)$, the zero-element of R. Note that in R, $2(1, 0, 0) = (1, 0, 0) + (1, 0, 0) = (2, 0, 0) = (0, 0, 0)$ since addition in the first component is in J_2. Similarly $3(0, 1, 0) = (0, 0, 0)$. Finally, for *no* positive integer m is $m(0, 0, 1) = (0, 0, 0)$.

Thus, from the point of view of the definition we gave above for characteristic, the ring R, which we just looked at, is neither fish nor fowl. The definition just doesn't have any meaning for R. We could generalize the notion of characteristic to arbitrary rings by doing it locally, defining it relative to given elements, rather than globally for the ring itself. We say that R has n-torsion, $n > 0$, if there is an element $a \neq 0$ in R such that $na = 0$, and $ma \neq 0$ for $0 < m < n$. For an integral domain D, it turns

out that if D has n-torsion, even for one $n > 0$, then it must be of finite characteristic (see Problem 8).

Problems

R is a ring in all the problems.

1. If $a, b, c, d \in R$, evaluate $(a + b)(c + d)$.
2. Prove that if $a, b \in R$, then $(a + b)^2 = a^2 + ab + ba + b^2$, where by x^2 we mean xx.
3. Find the form of the binomial theorem in a general ring; in other words, find an expression for $(a + b)^n$, where n is a positive integer.
4. If every $x \in R$ satisfies $x^2 = x$, prove that R must be commutative. (A ring in which $x^2 = x$ for all elements is called a *Boolean* ring.)
5. If R is a ring, merely considering it as an abelian group under its addition, we have defined, in Chapter 2, what is meant by na, where $a \in R$ and n is an integer. Prove that if $a, b \in R$ and n, m are integers, then $(na)(mb) = (nm)(ab)$.
6. If D is an integeral domain and D is of finite characteristic, prove that the characteristic of D is a prime number.
7. Give an example of an integral domain which has an infinite number of elements, yet is of finite characteristic.
8. If D is an integral domain and if $na = 0$ for some $a \neq 0$ in D and some integer $n \neq 0$, prove that D is of finite characteristic.
9. If R is a system satisfying all the conditions for a ring with unit element with the possible exception of $a + b = b + a$, prove that the axiom $a + b = b + a$ must hold in R and that R is thus a ring. (*Hint:* Expand $(a + b)(1 + 1)$ in two ways.)
10. Show that the commutative ring D is an integral domain if and only if for $a, b, c \in D$ with $a \neq 0$ the relation $ab = ac$ implies that $b = c$.
11. Prove that Lemma 3.2.2 is false if we drop the assumption that the integral domain is finite.
12. Prove that any field is an integral domain.
13. Useing the pigeonhole principle, prove that if m and n are relatively prime integers and a and b are any integers, there exists an integer x such that $x \equiv a \bmod m$ and $x \equiv b \bmod n$. (*Hint:* Consider the remainders of $a, a + m, a + 2m, \ldots, a + (n - 1)m$ on division by n.)
14. Using the pigeonhole principle, prove that the decimal expansion of a rational number must, after some point, become repeating.

3.3 Homomorphisms

In studying groups we have seen that the concept of a homomorphism turned out to be a fruitful one. This suggests that the appropriate analog for rings could also lead to important ideas. To recall, for groups a homomorphism was defined as a mapping such that $\phi(ab) = \phi(a)\phi(b)$. Since a ring has two operations, what could be a more natural extension of this type of formula than the

DEFINITION A mapping ϕ from the ring R into the ring R' is said to be a *homomorphism* if

1. $\phi(a + b) = \phi(a) + \phi(b)$,
2. $\phi(ab) = \phi(a)\phi(b)$,

for all $a, b \in R$.

As in the case of groups, let us again stress here that the $+$ and $\cdot$ occurring on the left-hand sides of the relations in 1 and 2 are those of R, whereas the $+$ and $\cdot$ occurring on the right-hand sides are those of R'.

A useful observation to make is that a homomorphism of one ring, R, into another, R', is, if we totally ignore the multiplications in both these rings, at least a homomorphism of R into R' when we consider them as abelian groups under their respective additions. Therefore, as far as addition is concerned, all the properties about homomorphisms of groups proved in Chapter 2 carry over. In particular, merely restating Lemma 2.7.2 for the case of the additive group of a ring yields for us

LEMMA 3.3.1 *If ϕ is a homomorphism of R into R', then*

1. $\phi(0) = 0$.
2. $\phi(-a) = -\phi(a)$ *for every* $a \in R$.

A word of caution: if both R and R' have the respective unit elements 1 and 1' for their multiplications it need not follow that $\phi(1) = 1'$. However, if R' is an integral domain, or if R' is arbitrary but ϕ is onto, then $\phi(1) = 1'$ is indeed true.

In the case of groups, given a homomorphism we associated with this homomorphism a certain subset of the group which we called the kernel of the homomorphism. What should the appropriate definition of the kernel of a homomorphism be for rings? After all, the ring has two operations, addition and multiplication, and it might be natural to ask which of these should be singled out as the basis for the definition. However, the choice is clear. Built into the definition of an arbitrary ring is the condition that the ring forms an abelian group under addition. The ring multiplication

was left much more unrestricted, and so, in a sense, much less under our control than is the addition. For this reason the emphasis is given to the operation of addition in the ring, and we make the

DEFINITION If ϕ is a homomorphism of R into R' then the *kernel of* ϕ, $I(\phi)$, is the set of all elements $a \in R$ such that $\phi(a) = 0$, the zero-element of R'.

LEMMA 3.3.2 *If ϕ is a homomorphism of R into R' with kernel $I(\phi)$, then*

1. *$I(\phi)$ is a subgroup of R under addition.*
2. *If $a \in I(\phi)$ and $r \in R$ then both ar and ra are in $I(\phi)$.*

Proof. Since ϕ is, in particular, a homomorphism of R, as an additive group, into R', as an additive group, (1) follows directly from our results in group theory.

To see (2), suppose that $a \in I(\phi)$, $r \in R$. Then $\phi(a) = 0$ so that $\phi(ar) = \phi(a)\phi(r) = 0\phi(r) = 0$ by Lemma 3.2.1. Similarly $\phi(ra) = 0$. Thus by defining property of $I(\phi)$ both ar and ra are in $I(\phi)$.

Before proceeding we examine these concepts for certain examples.

Example 3.3.1 Let R and R' be two arbitrary rings and define $\phi(a) = 0$ for all $a \in R$. Trivially ϕ is a homomorphism and $I(\phi) = R$. ϕ is called the zero-homomorphism.

Example 3.3.2 Let R be a ring, $R' = R$ and define $\phi(x) = x$ for every $x \in R$. Clearly ϕ is a homomorphism and $I(\phi)$ consists only of 0.

Example 3.3.3 Let $J(\sqrt{2})$ be all real numbers of the form $m + n\sqrt{2}$ where m, n are integers; $J(\sqrt{2})$ forms a ring under the usual addition and multiplication of real numbers. (Verify!) Define $\phi: J(\sqrt{2}) \to J(\sqrt{2})$ by $\phi(m + n\sqrt{2}) = m - n\sqrt{2}$. ϕ is a homomorphism of $J(\sqrt{2})$ onto $J(\sqrt{2})$ and its kernel $I(\phi)$, consists only of 0. (Verify!)

Example 3.3.4 Let J be the ring of integers, J_n, the ring of integers modulo n. Define $\phi: J \to J_n$ by $\phi(a) =$ remainder of a on division by n. The student should verify that ϕ is a homomorphism of J onto J_n and that the kernel, $I(\phi)$, of ϕ consists of all multiples of n.

Example 3.3.5 Let R be the set of all continuous, real-valued functions on the closed unit interval. R is made into a ring by the usual addition and multiplication of functions; that it is a ring is a consequence of the fact that the sum and product of two continuous functions are continuous

functions. Let F be the ring of real numbers and define $\phi: R \to F$ by $\phi(f(x)) = f(\frac{1}{2})$. ϕ is then a homomorphism of R onto F and its kernel consists of all functions in R vanishing at $x = \frac{1}{2}$.

All the examples given here have used commutative rings. Many beautiful examples exist where the rings are noncommutative but it would be premature to discuss such an example now.

DEFINITION A homomorphism of R into R' is said to be an *isomorphism* if it is a one-to-one mapping.

DEFINITION Two rings are said to be *isomorphic* if there is an isomorphism of one *onto* the other.

The remarks made in Chapter 2 about the meaning of an isomorphism and of the statement that two groups are isomorphic carry over verbatim to rings. Likewise, the criterion given in Lemma 2.7.4 that a homomorphism be an isomorphism translates directly from groups to rings in the form

LEMMA 3.3.3 *The homomorphism ϕ of R into R' is an isomorphism if and only if $I(\phi) = (0)$.*

3.4 Ideals and Quotient Rings

Once the idea of a homomorphism and its kernel have been set up for rings, based on our experience with groups, it should be fruitful to carry over some analog to rings of the concept of normal subgroup. Once this is achieved, one would hope that this analog would lead to a construction in rings like that of the quotient group of a group by a normal subgroup. Finally, if one were an optimist, one would hope that the homomorphism theorems for groups would come over in their entirety to rings.

Fortunately all this can be done, thereby providing us with an incisive technique for analyzing rings.

The first business at hand, then, seems to be to define a suitable "normal subgroup" concept for rings. With a little hindsight this is not difficult. If you recall, normal subgroups eventually turned out to be nothing else than kernels of homomorphisms, even though their primary defining conditions did not involve homomorphisms. Why not use this observation as the keystone to our definition for rings? Lemma 3.3.2 has already provided us with some conditions that a subset of a ring be the kernel of a homomorphism. We now take the point of view that, since no other information is at present available to us, we shall make the conclusions of Lemma 3.3.2 as the starting point of our endeavor, and so we define

DEFINITION A nonempty subset U of R is said to be a (two-sided) *ideal* of R if

1. U is a subgroup of R under addition.
2. For every $u \in U$ and $r \in R$, both ur and ru are in U.

Condition 2 asserts that U "swallows up" multiplication from the right and left by arbitrary ring elements. For this reason U is usually called a two-sided ideal. Since we shall have no occasion, other than in some of the problems, to use any other derivative concept of ideal, we shall merely use the word ideal, rather than two-sided ideal, in all that follows.

Given an ideal U of a ring R, let R/U be the set of all the distinct cosets of U in R which we obtain by considering U as a subgroup of R under addition. We note that we merely say coset, rather than right coset or left coset; this is justified since R is an abelian group under addition. To restate what we have just said, R/U consists of all the cosets, $a + U$, where $a \in R$. By the results of Chapter 2, R/U is automatically a group under addition; this is achieved by the composition law $(a + U) + (b + U) = (a + b) + U$. In order to impose a ring structure on R/U we must define, in it, a multiplication. What is more natural than to define $(a + U)(b + U) = ab + U$? However, we must make sure that this is meaningful. Otherwise put, we are obliged to show that if $a + U = a' + U$ and $b + U = b' + U$, then under our definition of the multiplication, $(a + U)(b + U) = (a' + U)(b' + U)$. Equivalently, it must be established that $ab + U = a'b' + U$. To this end we first note that since $a + U = a' + U$, $a = a' + u_1$, where $u_1 \in U$; similarly $b = b' + u_2$ where $u_2 \in U$. Thus $ab = (a' + u_1)(b + u_2) = a'b' + u_1b' + a'u_2 + u_1u_2$; *since U is an ideal of* R, $u_1b' \in U$, $a'u_2 \in U$, and $u_1u_2 \in U$. Consequently $u_1b' + a'u_2 + u_1u_2 = u_3 \in U$. But then $ab = a'b' + u_3$, from which we deduce that $ab + U = a'b' + u_3 + U$, and since $u_3 \in U$, $u_3 + U = U$. The net consequence of all this is that $ab + U = a'b' + U$. We at least have achieved the principal step on the road to our goal, namely of introducing a well-defined multiplication. The rest now becomes routine. To establish that R/U is a ring we merely have to go through the various axioms which define a ring and check whether they hold in R/U. All these verifications have a certain sameness to them, so we pick one axiom, the right distributive law, and prove it holds in R/U. The rest we leave to the student as informal exercises. If $X = a + U$, $Y = b + U$, $Z = c + U$ are three elements of R/U, where $a, b, c \in R$, then $(X + Y)Z = ((a + U) + (b + U))(c + U) = ((a + b) + U)(c + U) = (a + b)c + U = ac + bc + U = (ac + U) + (bc + U) = (a + U)(c + U) + (b + U)(c + U) = XZ + YZ$.

R/U has now been made into a ring. Clearly, if R is commutative then so is R/U, for $(a + U)(b + U) = ab + U = ba + U = (b + U)(a + U)$. (The converse to this is false.) If R has a unit element 1, then R/U has a

unit element $1 + U$. We might ask: In what relation is R/U to R? With the experience we now have in hand this is easy to answer. There is a homomorphism ϕ of R *onto* R/U given by $\phi(a) = a + U$ for every $a \in R$, whose kernel is exactly U. (The reader should verify that ϕ so defined is a homomorphism of R onto R/U with kernel U.)

We summarize these remarks in

LEMMA 3.4.1 *If U is an ideal of the ring R, then R/U is a ring and is a homomorphic image of R.*

With this construction of the *quotient ring* of a ring by an ideal satisfactorily accomplished, we are ready to bring over to rings the homomorphism theorems of groups. Since the proof is an exact verbatim translation of that for groups into the language of rings we merely state the theorem without proof, referring the reader to Chapter 2 for the proof.

THEOREM 3.4.1 *Let R, R' be rings and ϕ a homomorphism of R onto R' with kernel U. Then R' is isomorphic to R/U. Moreover there is a one-to-one correspondence between the set of ideals of R' and the set of ideals of R which contain U. This correspondence can be achieved by associating with an ideal W' in R' the ideal W in R defined by $W = \{x \in R \mid \phi(x) \in W'\}$. With W so defined, R/W is isomorphic to R'/W'.*

Problems

1. If U is an ideal of R and $1 \in U$, prove that $U = R$.
2. If F is a field, prove its only ideals are (0) and F itself.
3. Prove that any homomorphism of a field is either an isomorphism or takes each element into 0.
4. If R is a commutative ring and $a \in R$,
   (a) Show that $aR = \{ar \mid r \in R\}$ is a two-sided ideal of R.
   (b) Show by an example that this may be false if R is not commutative.
5. If U, V are ideals of R, let $U + V = \{u + v \mid u \in U, v \in V\}$. Prove that $U + V$ is also an ideal.
6. If U, V are ideals of R let UV be the set of all elements that can be written as finite sums of elements of the form uv where $u \in U$ and $v \in V$. Prove that UV is an ideal of R.
7. In Problem 6 prove that $UV \subset U \cap V$.
8. If R is the ring of integers, let U be the ideal consisting of all multiples of 17. Prove that if V is an ideal of R and $R \supset V \supset U$ then either $V = R$ or $V = U$. Generalize!

9. If U is an ideal of R, let $r(U) = \{x \in R \mid xu = 0 \text{ for all } u \in U\}$. Prove that $r(U)$ is an ideal of R.
10. If U is an ideal of R let $[R:U] = \{x \in R \mid rx \in U \text{ for every } r \in R\}$. Prove that $[R:U]$ is an ideal of R and that it contains U.
11. Let R be a ring with unit element. Using its elements we define a ring $\tilde{R}$ by defining $a \oplus b = a + b + 1$, and $a \cdot b = ab + a + b$, where $a, b \in R$ and where the addition and multiplication on the right-hand side of these relations are those of R.
    (a) Prove that $\tilde{R}$ is a ring under the operations $\oplus$ and $\cdot$.
    (b) What acts as the zero-element of $\tilde{R}$?
    (c) What acts as the unit-element of $\tilde{R}$?
    (d) Prove that R is isomorphic to $\tilde{R}$.

*12. In Example 3.1.6 we discussed the ring of rational 2×2 matrices. Prove that this ring has no ideals other than (0) and the ring itself.

*13. In Example 3.1.8 we discussed the real quaternions. Using this as a model we define the quaternions over the integers mod p, p an odd prime number, in exactly the same way; however, now considering all symbols of the form $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, where $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ are integers mod p.
    (a) Prove that this is a ring with p^4 elements whose only ideals are (0) and the ring itself.
    **(b) Prove that this ring is *not* a division ring.

If R is any ring a subset L of R is called a *left-ideal* of R if
1. L is a subgroup of R under addition.
2. $r \in R$, $a \in L$ implies $ra \in L$.

(One can similarly define a *right-ideal*.) An ideal is thus simultaneously a left- and right-ideal of R.

14. For $a \in R$ let $Ra = \{xa \mid x \in R\}$. Prove that Ra is a left-ideal of R.
15. Prove that the intersection of two left-ideals of R is a left-ideal of R.
16. What can you say about the intersection of a left-ideal and right-ideal of R?
17. If R is a ring and $a \in R$ let $r(a) = \{x \in R \mid ax = 0\}$. Prove that $r(a)$ is a right-ideal of R.
18. If R is a ring and L is a left-ideal of R let $\lambda(L) = \{x \in R \mid xa = 0 \text{ for all } a \in L\}$. Prove that $\lambda(L)$ is a two-sided ideal of R.

*19. Let R be a ring in which $x^3 = x$ for every $x \in R$. Prove that R is a commutative ring.

20. If R is a ring with unit element 1 and ϕ is a homomorphism of R *onto* R' prove that $\phi(1)$ is the unit element of R'.

21. If R is a ring with unit element 1 and ϕ is a homomorphism of R into an integral domain R' such that $I(\phi) \neq R$, prove that $\phi(1)$ is the unit element of R'.

3.5 More Ideals and Quotient Rings

We continue the discussion of ideals and quotient rings.

Let us take the point of view, for the moment at least, that a field is the most desirable kind of ring. Why? If for no other reason, we can divide in a field, so operations and results in a field more closely approximate our experience with real and complex numbers. In addition, as was illustrated by Problem 2 in the preceding problem set, a field has no homomorphic images other than itself or the trivial ring consisting of 0. Thus we cannot simplify a field by applying a homomorphism to it. Taking these remarks into consideration it is natural that we try to link a general ring, in some fashion, with fields. What should this linkage involve? We have a machinery whose component parts are homomorphisms, ideals, and quotient rings. With these we will forge the link.

But first we must make precise the rather vague remarks of the preceding paragraph. We now ask the explicit question: Under what conditions is the homomorphic image of a ring a field? For commutative rings we give a complete answer in this section.

Essential to treating this question is the converse to the result of Problem 2 of the problem list at the end of Section 3.4.

LEMMA 3.5.1 *Let R be a commutative ring with unit element whose only ideals are (0) and R itself. Then R is a field.*

Proof. In order to effect a proof of this lemma for any $a \neq 0 \in R$ we must produce an element $b \neq 0 \in R$ such that $ab = 1$.

So, suppose that $a \neq 0$ is in R. Consider the set $Ra = \{xa \mid x \in R\}$. We claim that Ra is an ideal of R. In order to establish this as fact we must show that it is a subgroup of R under addition and that if $u \in Ra$ and $r \in R$ then ru is also in Ra. (We only need to check that ru is in Ra for then ur also is since $ru = ur$.)

Now, if $u, v \in Ra$, then $u = r_1a$, $v = r_2a$ for some $r_1, r_2 \in R$. Thus $u + v = r_1a + r_2a = (r_1 + r_2)a \in Ra$; similarly $-u = -r_1a = (-r_1)a \in Ra$. Hence Ra is an additive subgroup of R. Moreover, if $r \in R$, $ru = r(r_1a) = (rr_1)a \in Ra$. Ra therefore satisfies all the defining conditions for an ideal of R, hence is an ideal of R. (Notice that both the distributive law and associative law of multiplication were used in the proof of this fact.)

By our assumptions on R, $Ra = (0)$ or $Ra = R$. Since $0 \neq a = 1a \in Ra$, $Ra \neq (0)$; thus we are left with the only other possibility, namely that $Ra = R$. This last equation states that every element in R is a multiple of

a by some element of R. In particular, $1 \in R$ and so it can be realized as a multiple of a; that is, there exists an element $b \in R$ such that $ba = 1$. This completes the proof of the lemma.

DEFINITION An ideal $M \neq R$ in a ring R is said to be a *maximal ideal* of R if whenever U is an ideal of R such that $M \subset U \subset R$, then either $R = U$ or $M = U$.

In other words, an ideal of R is a maximal ideal if it is impossible to squeeze an ideal between it and the full ring. Given a ring R there is no guarantee that it has any maximal ideals! If the ring has a unit element this can be proved, assuming a basic axiom of mathematics, the so-called axiom of choice. Also there may be many distinct maximal ideals in a ring R; this will be illustrated for us below in the ring of integers.

As yet we have made acquaintance with very few rings. Only by considering a given concept in many particular cases can one fully appreciate the concept and its motivation. Before proceeding we therefore examine some maximal ideals in two specific rings. When we come to the discussion of polynomial rings we shall exhibit there all the maximal ideals.

Example 3.5.1 Let R be the ring of integers, and let U be an ideal of R. Since U is a subgroup of R under addition, from our results in group theory, we know that U consists of all the multiples of a fixed integer n_0; we write this as $U = (n_0)$. What values of n_0 lead to maximal ideals?

We first assert that if p is a prime number then $P = (p)$ is a maximal ideal of R. For if U is an ideal of R and $U \supset P$, then $U = (n_0)$ for some integer n_0. Since $p \in P \subset U$, $p = mn_0$ for some integer m; because p is a prime this implies that $n_0 = 1$ or $n_0 = p$. If $n_0 = p$, then $P \subset U = (n_0) \subset P$, so that $U = P$ follows; if $n_0 = 1$, then $1 \in U$, hence $r = 1r \in U$ for all $r \in R$ whence $U = R$ follows. Thus no ideal, other than R or P itself, can be put between P and R, from which we deduce that P is maximal.

Suppose, on the other hand, that $M = (n_0)$ is a maximal ideal of R. We claim that n_0 must be a prime number, for if $n_0 = ab$, where a, b are positive integers, then $U = (a) \supset M$, hence $U = R$ or $U = M$. If $U = R$, then $a = 1$ is an easy consequence; if $U = M$, then $a \in M$ and so $a = rn_0$ for some integer r, since every element of M is a multiple of n_0. But then $n_0 = ab = rn_0b$, from which we get that $rb = 1$, so that $b = 1$, $n_0 = a$. Thus n_0 is a prime number.

In this particular example the notion of maximal ideal comes alive—it corresponds exactly to the notion of prime number. One should not, however, jump to any hasty generalizations; this kind of correspondence does not usually hold for more general rings.

Example 3.5.2 Let R be the ring of all the real-valued, continuous functions on the closed unit interval. (See Example 3.3.5.) Let

$$M = \{f(x) \in R \mid f(\tfrac{1}{2}) = 0\}.$$

M is certainly an ideal of R. Moreover, it is a maximal ideal of R, for if the ideal U contains M and $U \neq M$, then there is a function $g(x) \in U$, $g(x) \notin M$. Since $g(x) \notin M$, $g(\frac{1}{2}) = \alpha \neq 0$. Now $h(x) = g(x) - \alpha$ is such that $h(\frac{1}{2}) = g(\frac{1}{2}) - \alpha = 0$, so that $h(x) \in M \subset U$. But $g(x)$ is also in U; therefore $\alpha = g(x) - h(x) \in U$ and so $1 = \alpha\alpha^{-1} \in U$. Thus for any function $t(x) \in R$, $t(x) = 1t(x) \in U$, in consequence of which $U = R$. M is therefore a maximal ideal of R. Similarly if γ is a real number $0 \leq \gamma \leq 1$, then $M_\gamma = \{f(x) \in R \mid f(\gamma) = 0\}$ is a maximal ideal of R. It can be shown (see Problem 4 at the end of this section) that every maximal ideal is of this form. Thus here the maximal ideals correspond to the points on the unit interval.

Having seen some maximal ideals in some concrete rings we are ready to continue the general development with

THEOREM 3.5.1 *If R is a commutative ring with unit element and M is an ideal of R, then M is a maximal ideal of R if and only if R/M is a field.*

Proof. Suppose, first, that M is an ideal of R such that R/M is a field. Since R/M is a field its only ideals are (0) and R/M itself. But by Theorem 3.4.1 there is a one-to-one correspondence between the set of ideals of R/M and the set of ideals of R which contain M. The ideal M of R corresponds to the ideal (0) of R/M whereas the ideal R of R corresponds to the ideal R/M of R/M in this one-to-one mapping. Thus there is no ideal between M and R other than these two, whence M is a maximal ideal.

On the other hand, if M is a maximal ideal of R, by the correspondence mentioned above R/M has only (0) and itself as ideals. Furthermore R/M is commutative and has a unit element since R enjoys both these properties. All the conditions of Lemma 3.5.1 are fulfilled for R/M so we can conclude, by the result of that lemma, that R/M is a field.

We shall have many occasions to refer back to this result in our study of polynomial rings and in the theory of field extensions.

Problems

1. Let R be a ring with unit element, R not necessarily commutative, such that the only right-ideals of R are (0) and R. Prove that R is a division ring.

*2. Let R be a ring such that the only right ideals of R are (0) and R. Prove that either R is a division ring or that R is a ring with a prime number of elements in which $ab = 0$ for every $a, b \in R$.

3. Let J be the ring of integers, p a prime number, and (p) the ideal of J consisting of all multiples of p. Prove
   (a) $J/(p)$ is isomorphic to J_p, the ring of integers mod p.
   (b) Using Theorem 3.5.1 and part (a) of this problem, that J_p is a field.

**4. Let R be the ring of all real-valued continuous functions on the closed unit interval. If M is a maximal ideal of R, prove that there exists a real number γ, $0 \leq \gamma \leq 1$, such that $M = M_\gamma = \{f(x) \in R \mid f(\gamma) = 0\}$.

3.6 The Field of Quotients of an Integral Domain

Let us recall that an integral domain is a commutative ring D with the additional property that it has no zero-divisors, that is, if $ab = 0$ for some $a, b \in D$ then at least one of a or b must be 0. The ring of integers is, of course, a standard example of an integral domain.

The ring of integers has the attractive feature that we can enlarge it to the set of rational numbers, which is a field. Can we perform a similar construction for any integral domain? We will now proceed to show that indeed we can!

DEFINITION A ring R *can be imbedded* in a ring R' if there is an isomorphism of R into R'. (If R and R' have unit elements 1 and 1′ we insist, in addition, that this isomorphism takes 1 onto 1′.)

R' will be called an *over-ring* or *extension* of R if R can be imbedded in R'.

With this understanding of imbedding we prove

THEOREM 3.6.1 *Every integral domain can be imbedded in a field.*

Proof. Before becoming explicit in the details of the proof let us take an informal approach to the problem. Let D be our integral domain; roughly speaking the field we seek should be all quotients a/b, where $a, b \in D$ and $b \neq 0$. Of course in D, a/b may very well be meaningless. What should we require of these symbols a/b? Clearly we must have an answer to the following three questions:

1. When is $a/b = c/d$?
2. What is $(a/b) + (c/d)$?
3. What is $(a/b)(c/d)$?

In answer to 1, what could be more natural than to insist that $a/b = c/d$

if and only if $ad = bc$? As for 2 and 3, why not try the obvious, that is, define

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \quad \text{and} \quad \frac{a}{b}\frac{c}{d} = \frac{ac}{bd}.$$

In fact in what is to follow we make these considerations our guide. So let us leave the heuristics and enter the domain of mathematics, with precise definitions and rigorous deductions.

Let $\mathscr{M}$ be the set of all ordered pairs (a, b) where $a, b \in D$ and $b \neq 0$. (Think of (a, b) as a/b.) In $\mathscr{M}$ we now define a relation as follows:

$$(a, b) \sim (c, d) \text{ if and only if } ad = bc.$$

We claim that this defines an equivalence relation on $\mathscr{M}$. To establish this we check the three defining conditions for an equivalence relation for this particular relation.

1. If $(a, b) \in \mathscr{M}$, then $(a, b) \sim (a, b)$ since $ab = ba$.
2. If $(a, b), (c, d) \in \mathscr{M}$ and $(a, b) \sim (c, d)$, then $ad = bc$, hence $cb = da$, and so $(c, d) \sim (a, b)$.
3. If $(a, b), (c, d), (e, f)$ are all in $\mathscr{M}$ and $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$, then $ad = bc$ and $cf = de$. Thus $bcf = bde$, and since $bc = ad$, it follows that $adf = bde$. Since D is commutative, this relation becomes $afd = bed$; since, moreover, D is an integral domain and $d \neq 0$, this relation further implies that $af = be$. But then $(a, b) \sim (e, f)$ and our relation is transitive.

Let $[a, b]$ be the equivalence class in $\mathscr{M}$ of (a, b), and let F be the set of all such equivalence classes $[a, b]$ where $a, b \in D$ and $b \neq 0$. F is the candidate for the field we are seeking. In order to create out of F a field we must introduce an addition and a multiplication for its elements and then show that under these operations F forms a field.

We first dispose of the addition. Motivated by our heuristic discussion at the beginning of the proof we define

$$[a, b] + [c, d] = [ad + bc, bd].$$

Since D is an integral domain and both $b \neq 0$ and $d \neq 0$ we have that $bd \neq 0$; this, at least, tells us that $[ad + bc, bd] \in F$. We now assert that this addition is well defined, that is, if $[a, b] = [a', b']$ and $[c, d] = [c', d']$, then $[a, b] + [c, d] = [a', b'] + [c', d']$. To see that this is so, from $[a, b] = [a', b']$ we have that $ab' = ba'$; from $[c, d] = [c', d']$ we have that $cd' = dc'$. What we need is that these relations force the equality of $[a, b] + [c, d]$ and $[a', b'] + [c', d']$. From the definition of addition this boils down to showing that $[ad + bc, bd] = [a'd' + b'c', b'd']$, or, in equivalent terms, that $(ad + bc)b'd' = bd(a'd' + b'c')$. Using $ab' = ba'$, $cd' = dc'$

this becomes: $(ad + bc)b'd' = adb'd' + bcb'd' = ab'dd' + bb'cd' = ba'dd' + bb'dc' = bd(a'd' + b'c')$, which is the desired equality.

Clearly $[0, b]$ acts as a zero-element for this addition and $[-a, b]$ as the negative of $[a, b]$. It is a simple matter to verify that F is an abelian group under this addition.

We now turn to the multiplication in F. Again motivated by our preliminary heuristic discussion we define $[a, b][c, d] = [ac, bd]$. As in the case of addition, since $b \neq 0$, $d \neq 0$, $bd \neq 0$ and so $[ac, bd] \in F$. A computation, very much in the spirit of the one just carried out, proves that if $[a, b] = [a', b']$ and $[c, d] = [c', d']$ then $[a, b][c, d] = [a', b'][c', d']$. One can now show that the nonzero elements of F (that is, all the elements $[a, b]$ where $a \neq 0$) form an abelian group under multiplication in which $[d, d]$ acts as the unit element and where

$$[c, d]^{-1} = [d, c] \text{ (since } c \neq 0, [d, c] \text{ is in } F).$$

It is a routine computation to see that the distributive law holds in F. F is thus a field.

All that remains is to show that D can be imbedded in F. We shall exhibit an explicit isomorphism of D into F. Before doing so we first notice that for $x \neq 0, y \neq 0$ in D, $[ax, x] = [ay, y]$ because $(ax)y = x(ay)$; let us denote $[ax, x]$ by $[a, 1]$. Define $\phi: D \to F$ by $\phi(a) = [a, 1]$ for every $a \in D$. We leave it to the reader to verify that ϕ is an isomorphism of D into F, and that if D has a unit element 1, then $\phi(1)$ is the unit element of F. The theorem is now proved in its entirety.

F is usually called the *field of quotients* of D. In the special case in which D is the ring of integers, the F so constructed is, of course, the field of rational numbers.

Problems

1. Prove that if $[a, b] = [a', b']$ and $[c, d] = [c', d']$ then $[a, b][c, d] = [a', b'][c', d']$.
2. Prove the distributive law in F.
3. Prove that the mapping $\phi: D \to F$ defined by $\phi(a) = [a, 1]$ is an isomorphism of D into F.
4. Prove that if K is any field which contains D then K contains a subfield isomorphic to F. (*In this sense F is the smallest field containing D.*)

*5. Let R be a commutative ring with unit element. A nonempty subset S of R is called a multiplicative system if
   1. $0 \notin S$.
   2. $s_1, s_2 \in S$ implies that $s_1 s_2 \in S$.

Let $\mathcal{M}$ be the set of all ordered pairs (r, s) where $r \in R$, $s \in S$. In $\mathcal{M}$ define $(r, s) \sim (r', s')$ if there exists an element $s'' \in S$ such that

$$s''(rs' - sr') = 0.$$

(a) Prove that this defines an equivalence relation on $\mathcal{M}$.

Let the equivalence class of (r, s) be denoted by $[r, s]$, and let R_S be the set of all the equivalence classes. In R_S define $[r_1, s_1] + [r_2, s_2] = [r_1s_2 + r_2s_1, s_1s_2]$ and $[r_1, s_1][r_2, s_2] = [r_1r_2, s_1s_2]$.

(b) Prove that the addition and multiplication described above are well defined and that R_S forms a ring under these operations.

(c) Can R be imbedded in R_S?

(d) Prove that the mapping $\phi: R \to R_s$ defined by $\phi(a) = [as, s]$ is a homomorphism of R into R_S and find the kernel of ϕ.

(e) Prove that this kernel has no element of S in it.

(f) Prove that every element of the form $[s_1, s_2]$(where $s_1, s_2 \in S$) in R_S has an inverse in R_S.

6. Let D be an integral domain, $a, b \in D$. Suppose that $a^n = b^n$ and $a^m = b^m$ for two relatively prime positive integers m and n. Prove that $a = b$.

7. Let R be a ring, possibly noncommutative, in which $xy = 0$ implies $x = 0$ or $y = 0$. If $a, b \in R$ and $a^n = b^n$ and $a^{n+1} = b^m$ for two relatively prime positive integers m and n, prove that $a = b$.

3.7 Euclidean Rings

The class of rings we propose to study now is motivated by several existing examples—the ring of integers, the Gaussian integers (Section 3.8), and polynomial rings (Section 3.9). The definition of this class is designed to incorporate in it certain outstanding characteristics of the three concrete examples listed above.

DEFINITION An integral domain R is said to be a *Euclidean ring* if for every $a \neq 0$ in R there is defined a nonnegative integer $d(a)$ such that

1. For all $a, b \in R$, both nonzero, $d(a) \leq d(ab)$.
2. For any $a, b \in R$, both nonzero, there exist $t, r \in R$ such that $a = tb + r$ where either $r = 0$ or $d(r) < d(b)$.

We do not assign a value to $d(0)$. The integers serve as an example of a Euclidean ring, where $d(a) =$ absolute value of a acts as the required function. In the next section we shall see that the Gaussian integers also form a Euclidean ring. Out of that observation, and the results developed in this part, we shall prove a classic theorem in number theory due to

Fermat, namely, that every prime number of the form $4n + 1$ can be written as the sum of two squares.

We begin with

THEOREM 3.7.1 *Let R be a Euclidean ring and let A be an ideal of R. Then there exists an element $a_0 \in A$ such that A consists exactly of all a_0x as x ranges over R.*

Proof. If A just consists of the element 0, put $a_0 = 0$ and the conclusion of the theorem holds.

Thus we may assume that $A \neq (0)$; hence there is an $a \neq 0$ in A. Pick an $a_0 \in A$ such that $d(a_0)$ is minimal. (Since d takes on nonnegative integer values this is always possible.)

Suppose that $a \in A$. By the properties of Euclidean rings there exist $t, r \in R$ such that $a = ta_0 + r$ where $r = 0$ or $d(r) < d(a_0)$. Since $a_0 \in A$ and A is an ideal of R, ta_0 is in A. Combined with $a \in A$ this results in $a - ta_0 \in A$; but $r = a - ta_0$, whence $r \in A$. If $r \neq 0$ then $d(r) < d(a_0)$, giving us an element r in A whose d-value is smaller than that of a_0, in contradiction to our choice of a_0 as the element in A of minimal d-value. Consequently $r = 0$ and $a = ta_0$, which proves the theorem.

We introduce the notation $(a) = \{xa \mid x \in R\}$ to represent the ideal of all multiples of a.

DEFINITION An integral domain R with unit element is a *principal ideal ring* if every ideal A in R is of the form $A = (a)$ for some $a \in R$.

Once we establish that a Euclidean ring has a unit element, in virtue of Theorem 3.7.1, we shall know that a Euclidean ring is a principal ideal ring. The converse, however, is false; there are principal ideal rings which are not Euclidean rings. [See the paper by T. Motzkin, *Bulletin of the American Mathematical Society*, Vol. 55 (1949), pages 1142–1146, entitled "The Euclidean algorithm."]

COROLLARY TO THEOREM 3.7.1 *A Euclidean ring possesses a unit element.*

Proof. Let R be a Euclidean ring; then R is certainly an ideal of R, so that by Theorem 3.7.1 we may conclude that $R = (u_0)$ for some $u_0 \in R$. Thus every element in R is a multiple of u_0. Therefore, in particular, $u_0 = u_0c$ for some $c \in R$. If $a \in R$ then $a = xu_0$ for some $x \in R$, hence $ac = (xu_0)c = x(u_0c) = xu_0 = a$. Thus c is seen to be the required unit element.

DEFINITION If $a \neq 0$ and b are in a commutative ring R then a is said to *divide* b if there exists a $c \in R$ such that $b = ac$. We shall use the symbol

$a \mid b$ to represent the fact that a divides b and $a \nmid b$ to mean that a does not divide b.

The proof of the next remark is so simple and straightforward that we omit it.

REMARK 1. *If* $a \mid b$ *and* $b \mid c$ *then* $a \mid c$.
2. *If* $a \mid b$ *and* $a \mid c$ *then* $a \mid (b \pm c)$.
3. *If* $a \mid b$ *then* $a \mid bx$ *for all* $x \in R$.

DEFINITION If $a, b \in R$ then $d \in R$ is said to be a *greatest common divisor* of a and b if

1. $d \mid a$ and $d \mid b$.
2. Whenever $c \mid a$ and $c \mid b$ then $c \mid d$.

We shall use the notation $d = (a, b)$ to denote that d is a greatest common divisor of a and b.

LEMMA 3.7.1 *Let R be a Euclidean ring. Then any two elements a and b in R have a greatest common divisor d. Moreover $d = \lambda a + \mu b$ for some $\lambda, \mu \in R$.*

Proof. Let A be the set of all elements $ra + sb$ where r, s range over R. We claim that A is an ideal of R. For suppose that $x, y \in A$; therefore $x = r_1 a + s_1 b$, $y = r_2 a + s_2 b$, and so $x \pm y = (r_1 \pm r_2)a + (s_1 \pm s_2)b \in A$. Similarly, for any $u \in R$, $ux = u(r_1 a + s_1 b) = (ur_1)a + (us_1)b \in A$.

Since A is an ideal of R, by Theorem 3.7.1 there exists an element $d \in A$ such that every element in A is a mutiple of d. By dint of the fact that $d \in A$ and that every element of A is of the form $ra + sb$, $d = \lambda a + \mu b$ for some $\lambda, \mu \in R$. Now by the corollary to Theorem 3.7.1, R has a unit element 1; thus $a = 1a + 0b \in A$, $b = 0a + 1b \in A$. Being in A, they are both multiples of d, whence $d \mid a$ and $d \mid b$.

Suppose, finally, that $c \mid a$ and $c \mid b$; then $c \mid \lambda a$ and $c \mid \mu b$ so that c certainly divides $\lambda a + \mu b = d$. Therefore d has all the requisite conditions for a greatest common divisor and the lemma is proved.

DEFINITION Let R be a commutative ring with unit element. An element $a \in R$ is a *unit* in R if there exists an element $b \in R$ such that $ab = 1$.

Do not confuse a unit with a unit element! A unit in a ring is an element whose inverse is also in the ring.

LEMMA 3.7.2 *Let R be an integral domain with unit element and suppose that for $a, b \in R$ both $a \mid b$ and $b \mid a$ are true. Then $a = ub$, where u is a unit in R.*

Proof. Since $a \mid b$, $b = xa$ for some $x \in R$; since $b \mid a$, $a = yb$ for some $y \in R$. Thus $b = x(yb) = (xy)b$; but these are elements of an integral domain, so that we can cancel the b and obtain $xy = 1$; y is thus a unit in R and $a = yb$, proving the lemma.

DEFINITION Let R be a commutative ring with unit element. Two elements a and b in R are said to be *associates* if $b = ua$ for some unit u in R.

The relation of being associates is an equivalence relation. (Problem 1 at the end of this section.) Note that in a Euclidean ring any two greatest common divisors of two given elements are associates (Problem 2).

Up to this point we have, as yet, not made use of condition 1 in the definition of a Euclidean ring, namely that $d(a) \leq d(ab)$ for $b \neq 0$. We now make use of it in the proof of

LEMMA 3.7.3 *Let R be a Euclidean ring and $a, b \in R$. If $b \neq 0$ is not a unit in R, then $d(a) < d(ab)$.*

Proof. Consider the ideal $A = (a) = \{xa \mid x \in R\}$ of R. By condition 1 for a Euclidean ring, $d(a) \leq d(xa)$ for $x \neq 0$ in R. Thus the d-value of a is the minimum for the d-value of any element in A. Now $ab \in A$; if $d(ab) = d(a)$, by the proof used in establishing Theorem 3.7.1, since the d-value of ab is minimal in regard to A, every element in A is a multiple of ab. In particular, since $a \in A$, a must be a multiple of ab; whence $a = abx$ for some $x \in R$. Since all this is taking place in an integral domain we obtain $bx = 1$. In this way b is a unit in R, in contradiction to the fact that it was not a unit. The net result of this is that $d(a) < d(ab)$.

DEFINITION In the Euclidean ring R a nonunit π is said to be a *prime element* of R if whenever $\pi = ab$, where a, b are in R, then one of a or b is a unit in R.

A prime element is thus an element in R which cannot be factored in R in a nontrivial way.

LEMMA 3.7.4 *Let R be a Euclidean ring. Then every element in R is either a unit in R or can be written as the product of a finite number of prime elements of R.*

Proof. The proof is by induction on $d(a)$.

If $d(a) = d(1)$ then a is a unit in R (Problem 3), and so in this case, the assertion of the lemma is correct.

We assume that the lemma is true for all elements x in R such that $d(x) < d(a)$. On the basis of this assumption we aim to prove it for a. This would complete the induction and prove the lemma.

If a is a prime element of R there is nothing to prove. So suppose that $a = bc$ where neither b nor c is a unit in R. By Lemma 3.7.3, $d(b) < d(bc) = d(a)$ and $d(c) < d(bc) = d(a)$. Thus by our induction hypothesis b and c can be written as a product of a finite number of prime elements of R; $b = \pi_1\pi_2 \cdots \pi_n$, $c = \pi_1'\pi_2' \cdots \pi_m'$ where the π's and π''s are prime elements of R. Consequently $a = bc = \pi_1\pi_2 \cdots \pi_n\pi_1'\pi_2' \cdots \pi_m'$ and in this way a has been factored as a product of a finite number of prime elements. This completes the proof.

DEFINITION In the Euclidean ring R, a and b in R are said to be *relatively prime* if their greatest common divisor is a unit of R.

Since any associate of a greatest common divisor is a greatest common divisor, and since 1 is an associate of any unit, if a and b are relatively prime we may assume that $(a, b) = 1$.

LEMMA 3.7.5 *Let R be a Euclidean ring. Suppose that for $a, b, c \in R$, $a \mid bc$ but $(a, b) = 1$. Then $a \mid c$.*

Proof. As we have seen in Lemma 3.7.1, the greatest common divisor of a and b can be realized in the form $\lambda a + \mu b$. Thus by our assumptions, $\lambda a + \mu b = 1$. Multiplying this relation by c we obtain $\lambda ac + \mu bc = c$. Now $a \mid \lambda ac$, always, and $a \mid \mu bc$ since $a \mid bc$ by assumption; therefore $a \mid (\lambda ac + \mu bc) = c$. This is, of course, the assertion of the lemma.

We wish to show that prime elements in a Euclidean ring play the same role that prime numbers play in the integers. If π in R is a prime element of R and $a \in R$, then either $\pi \mid a$ or $(\pi, a) = 1$, for, in particular, (π, a) is a divisor of π so it must be π or 1 (or any unit). If $(\pi, a) = 1$, one-half our assertion is true; if $(\pi, a) = \pi$, since $(\pi, a) \mid a$ we get $\pi \mid a$, and the other half of our assertion is true.

LEMMA 3.7.6 *If π is a prime element in the Euclidean ring R and $\pi \mid ab$ where $a, b \in R$ then π divides at least one of a or b.*

Proof. Suppose that π does not divide a; then $(\pi, a) = 1$. Applying Lemma 3.7.5 we are led to $\pi \mid b$.

COROLLARY *If π is a prime element in the Euclidean ring R and $\pi \mid a_1a_2 \cdots a_n$ then π divides at least one $a_1, a_2, \ldots, a_n$.*

We carry the analogy between prime elements and prime numbers further and prove

THEOREM 3.7.2 (UNIQUE FACTORIZATION THEOREM) *Let R be a Euclidean ring and $a \neq 0$ a nonunit in R. Suppose that $a = \pi_1\pi_2 \cdots \pi_n = \pi'_1\pi'_2 \cdots \pi'_m$ where the π_i and π'_j are prime elements of R. Then $n = m$ and each π_i, $1 \leq i \leq n$ is an associate of some π'_j, $1 \leq j \leq m$ and conversely each π'_k is an associate of some π_q.*

Proof. Look at the relation $a = \pi_1\pi_2 \cdots \pi_n = \pi'_1\pi'_2 \cdots \pi'_m$. But $\pi_1 \mid \pi_1\pi_2 \cdots \pi_n$, hence $\pi_1 \mid \pi'_1\pi'_2 \cdots \pi'_m$. By Lemma 3.7.6, π_1 must divide some π'_i; since π_1 and π'_i are both prime elements of R and $\pi_1 \mid \pi'_i$ they must be associates and $\pi'_i = u_1\pi_1$, where u_1 is a unit in R. Thus $\pi_1\pi_2 \cdots \pi_n = \pi'_1\pi'_2 \cdots \pi'_m = u_1\pi_1\pi'_2 \cdots \pi'_{i-1}\pi'_{i+1} \cdots \pi'_m$; cancel off π_1 and we are left with $\pi_2 \cdots \pi_n = u_1\pi'_2 \cdots \pi'_{i\ 1}\pi'_{i+1} \cdots \pi'_m$. Repeat the argument on this relation with π_2. After n steps, the left side becomes 1, the right side a product of a certain number of π' (the excess of m over n). This would force $n \leq m$ since the π' are not units. Similarly, $m \leq n$, so that $n = m$. In the process we have also showed that every π_i has some π'_i as an associate and conversely.

Combining Lemma 3.7.4 and Theorem 3.7.2 we have that *every nonzero element in a Euclidean ring R can be uniquely written (up to associates) as a product of prime elements or is a unit in R.*

We finish the section by determining all the maximal ideals in a Euclidean ring.

In Theorem 3.7.1 we proved that any ideal A in the Euclidean ring R is of the form $A = (a_0)$ where $(a_0) = \{xa_0 \mid x \in R\}$. We now ask: What conditions imposed on a_0 insure that A is a maximal ideal of R? For this question we have a simple, precise answer, namely

LEMMA 3.7.7 *The ideal $A = (a_0)$ is a maximal ideal of the Euclidean ring R if and only if a_0 is a prime element of R.*

Proof. We first prove that if a_0 is not a prime element, then $A = (a_0)$ is not a maximal ideal. For, suppose that $a_0 = bc$ where $b, c \in R$ and neither b nor c is a unit. Let $B = (b)$; then certainly $a_0 \in B$ so that $A \subset B$. We claim that $A \neq B$ and that $B \neq R$.

If $B = R$ then $1 \in B$ so that $1 = xb$ for some $x \in R$, forcing b to be a unit in R, which it is not. On the other hand, if $A = B$ then $b \in B = A$ whence $b = xa_0$ for some $x \in R$. Combined with $a_0 = bc$ this results in $a_0 = xca_0$, in consequence of which $xc = 1$. But this forces c to be a unit in R, again contradicting our assumption. Therefore B is neither A nor R and since $A \subset B$, A cannot be a maximal ideal of R.

Conversely, suppose that a_0 is a prime element of R and that U is an ideal of R such that $A = (a_0) \subset U \subset R$. By Theorem 3.7.1, $U = (u_0)$. Since $a_0 \in A \subset U = (u_0)$, $a_0 = xu_0$ for some $x \in R$. But a_0 is a prime element of R, from which it follows that either x or u_0 is a unit in R. If u_0

is a unit in R then $U = R$ (see Problem 5). If, on the other hand, x is a unit in R, then $x^{-1} \in R$ and the relation $a_0 = xu_0$ becomes $u_0 = x^{-1}a_0 \in A$ since A is an ideal of R. This implies that $U \subset A$; together with $A \subset U$ we conclude that $U = A$. Therefore there is no ideal of R which fits strictly between A and R. This means that A is a maximal ideal of R.

Problems

1. In a commutative ring with unit element prove that the relation a is an associate of b is an equivalence relation.
2. In a Euclidean ring prove that any two greatest common divisors of a and b are associates.
3. Prove that a necessary and sufficient condition that the element a in the Euclidean ring be a unit is that $d(a) = d(1)$.
4. Prove that in a Euclidean ring (a, b) can be found as follows:

$$\begin{aligned} b &= q_0 a + r_1, \quad \text{where} \quad d(r_1) < d(a) \\ a &= q_1 r_1 + r_2, \quad \text{where} \quad d(r_2) < d(r_1) \\ r_1 &= q_2 r_2 + r_3, \quad \text{where} \quad d(r_3) < d(r_2) \\ &\vdots \qquad\qquad\qquad\qquad \vdots \\ r_{n-1} &= q_n r_n \end{aligned}$$

and $r_n = (a, b)$.

5. Prove that if an ideal U of a ring R contains a unit of R, then $U = R$.
6. Prove that the units in a commutative ring with a unit element form an abelian group.
7. Given two elements a, b in the Euclidean ring R their *least common multiple* $c \in R$ is an element in R such that $a \mid c$ and $b \mid c$ and such that whenever $a \mid x$ and $b \mid x$ for $x \in R$ then $c \mid x$. Prove that any two elements in the Euclidean ring R have a least common multiple in R.
8. In Problem 7, if the least common multiple of a and b is denoted by $[a, b]$, prove that $[a, b] = ab/(a, b)$.

3.8 A Particular Euclidean Ring

An abstraction in mathematics gains in substance and importance when, particularized to a specific example, it sheds new light on this example. We are about to particularize the notion of a Euclidean ring to a concrete ring, the ring of Gaussian integers. Applying the general results obtained about Euclidean rings to the Gaussian integers we shall obtain a highly nontrivial theorem about prime numbers due to Fermat.

Let $J[i]$ denote the set of all complex numbers of the form $a + bi$ where a and b are integers. Under the usual addition and multiplication of complex numbers $J[i]$ forms an integral domain called the domain of *Gaussian integers.*

Our first objective is to exhibit $J[i]$ as a Euclidean ring. In order to do this we must first introduce a function $d(x)$ defined for every nonzero element in $J[i]$ which satisfies

1. $d(x)$ is a nonnegative integer for every $x \neq 0 \in J[i]$.
2. $d(x) \leq d(xy)$ for every $y \neq 0$ in $J[i]$.
3. Given $u, v \in J[i]$ there exist $t, r \in J[i]$ such that $v = tu + r$ where $r = 0$ or $d(r) < d(u)$.

Our candidate for this function d is the following: if $x = a + bi \in J[i]$, then $d(x) = a^2 + b^2$. The $d(x)$ so defined certainly satisfies property 1; in fact, if $x \neq 0 \in J[i]$ then $d(x) \geq 1$. As is well known, for any two complex numbers (not necessarily in $J[i]$) x, y, $d(xy) = d(x)d(y)$; thus if x and y are in addition in $J[i]$ and $y \neq 0$, then since $d(y) \geq 1$, $d(x) = d(x)1 \leq d(x)d(y) = d(xy)$, showing that condition 2 is satisfied. All our effort now will be to show that condition 3 also holds for this function d in $J[i]$. This is done in the proof of

THEOREM 3.8.1 *$J[i]$ is a Euclidean ring.*

Proof. As was remarked in the discussion above, to prove Theorem 3.8.1 we merely must show that, given $x, y \in J[i]$ there exists $t, r \in J[i]$ such that $y = tx + r$ where $r = 0$ or $d(r) < d(x)$.

We first establish this for a very special case, namely, where y is arbitrary in $J[i]$ but where x is an (ordinary) positive integer n. Suppose that $y = a + bi$; by the division algorithm for the ring of integers we can find integers u, v such that $a = un + u_1$ and $b = vn + v_1$ where u_1 and v_1 are integers satisfying $|u_1| \leq \frac{1}{2}n$ and $|v_1| \leq \frac{1}{2}n$. Let $t = u + vi$ and $r = u_1 + v_1 i$; then $y = a + bi = un + u_1 + (vn + v_1)i = (u + vi)n + u_1 + v_1 i = tn + r$. Since $d(r) = d(u_1 + v_1 i) = {u_1}^2 + {v_1}^2 \leq n^2/4 + n^2/4 < n^2 = d(n)$, we see that in this special case we have shown that $y = tn + r$ with $r = 0$ or $d(r) < d(n)$.

We now go to the general case; let $x \neq 0$ and y be arbitrary elements in $J[i]$. Thus $x\bar{x}$ is a positive integer n where $\bar{x}$ is the complex conjugate of x. Applying the result of the paragraph above to the elements $y\bar{x}$ and n we see that there are elements $t, r \in J[i]$ such that $y\bar{x} = tn + r$ with $r = 0$ or $d(r) < d(n)$. Putting into this relation $n = x\bar{x}$ we obtain $d(y\bar{x} - tx\bar{x}) < d(n) = d(x\bar{x})$; applying to this the fact that $d(y\bar{x} - tx\bar{x}) = d(y - tx)d(\bar{x})$ and $d(x\bar{x}) = d(x)d(\bar{x})$ we obtain that $d(y - tx)d(\bar{x}) < d(x)d(\bar{x})$. Since $x \neq 0$, $d(\bar{x})$ is a positive integer, so this inequality simplifies to $d(y - tx) < d(x)$. We represent $y = tx + r_0$, where $r_0 = y - tx$; thus t and r_0 are in

$J[i]$ and as we saw above, $r_0 = 0$ or $d(r_0) = d(y - tx) < d(x)$. This proves the theorem.

Since $J[i]$ has been proved to be a Euclidean ring, we are free to use the results established about this class of rings in the previous section to the Euclidean ring we have at hand, $J[i]$.

LEMMA 3.8.1 *Let p be a prime integer and suppose that for some integer c relatively prime to p we can find integers x and y such that $x^2 + y^2 = cp$. Then p can be written as the sum of squares of two integers, that is, there exist integers a and b such that $p = a^2 + b^2$.*

Proof. The ring of integers is a subring of $J[i]$. Suppose that the integer p is also a prime element of $J[i]$. Since $cp = x^2 + y^2 = (x + yi)(x - yi)$, by Lemma 3.7.6, $p \mid (x + yi)$ or $p \mid (x - yi)$ in $J[i]$. But if $p \mid (x + yi)$ then $x + yi = p(u + vi)$ which would say that $x = pu$ and $y = pv$ so that p also would divide $x - yi$. But then $p^2 \mid (x + yi)(x - yi) = cp$ from which we would conclude that $p \mid c$ contrary to assumption. Similarly if $p \mid (x - yi)$. Thus p is *not* a prime element in $J[i]$! In consequence of this,

$$p = (a + bi)(g + di)$$

where $a + bi$ and $g + di$ are in $J[i]$ and where neither $a + bi$ nor $g + di$ is a unit in $J[i]$. But this means that neither $a^2 + b^2 = 1$ nor $g^2 + d^2 = 1$. (See Problem 2.) From $p = (a + bi)(g + di)$ it follows easily that $p = (a - bi)(g - di)$. Thus

$$p^2 = (a + bi)(g + di)(a - bi)\ (g - di) = (a^2 + b^2)(g^2 + d^2).$$

Therefore $(a^2 + b^2) \mid p^2$ so $a^2 + b^2 = 1$, p or p^2; $a^2 + b^2 \neq 1$ since $a + bi$ is not a unit, in $J[i]$; $a^2 + b^2 \neq p^2$, otherwise $g^2 + d^2 = 1$, contrary to the fact that $g + di$ is not a unit in $J[i]$. Thus the only feasibility left is that $a^2 + b^2 = p$ and the lemma is thereby established.

The odd prime numbers divide into two classes, those which have a remainder of 1 on division by 4 and those which have a remainder of 3 on division by 4. We aim to show that every prime number of the first kind can be written as the sum of two squares, whereas no prime in the second class can be so represented.

LEMMA 3.8.2 *If p is a prime number of the form $4n + 1$, then we can solve the congruence $x^2 \equiv -1 \bmod p$.*

Proof. Let $x = 1 \cdot 2 \cdot 3 \cdots (p - 1)/2$. Since $p - 1 = 4n$, in this product for x there are an even number of terms, in consequence of which

$$x = (-1)(-2)(-3) \cdots \left(-\left(\frac{p-1}{2}\right)\right).$$

But $p - k \equiv -k \bmod p$, so that

$$x^2 \equiv \left(1 \cdot 2 \cdots \frac{p-1}{2}\right)(-1)(-2)\cdots\left(-\left(\frac{p-1}{2}\right)\right)$$

$$\equiv 1 \cdot 2 \cdots \frac{p-1}{2}\frac{p+1}{2}\cdots(p-1)$$

$$\equiv (p-1)! = -1 \bmod p.$$

We are using here Wilson's theorem, proved earlier, namely that if p is a prime number $(p-1)! \equiv -1(p)$.

To illustrate this result, if $p = 13$,

$$x = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720 = 5 \bmod 13 \text{ and } 5^2 = -1 \bmod 13.$$

THEOREM 3.8.2 (FERMAT) *If p is a prime number of the form $4n + 1$, then $p = a^2 + b^2$ for some integers a, b.*

Proof. By Lemma 3.8.2 there exists an x such that $x^2 \equiv -1 \bmod p$. The x can be chosen so that $0 \leq x \leq p - 1$ since we only need to use the remainder of x on division by p. We can restrict the size of x even further, namely to satisfy $|x| \leq p/2$. For if $x > p/2$, then $y = p - x$ satisfies $y^2 \equiv -1 \bmod p$ but $|y| \leq p/2$. Thus we may assume that we have an integer x such that $|x| \leq p/2$ and $x^2 + 1$ is a multiple of p, say cp. Now $cp = x^2 + 1 \leq p^2/4 + 1 < p^2$, hence $c < p$ and so $p \nmid c$. Invoking Lemma 3.8.1 we obtain that $p = a^2 + b^2$ for some integers a and b, proving the theorem.

Problems

1. Find all the units in $J[i]$.
2. If $a + bi$ is not a unit of $J[i]$ prove that $a^2 + b^2 > 1$.
3. Find the greatest common divisor in $J[i]$ of
   (a) $3 + 4i$ and $4 - 3i$. (b) $11 + 7i$ and $18 - i$.
4. Prove that if p is a prime number of the form $4n + 3$, then there is no x such that $x^2 \equiv -1 \bmod p$.
5. Prove that no prime of the form $4n + 3$ can be written as $a^2 + b^2$ where a and b are integers.
6. Prove that there is an infinite number of primes of the form $4n + 3$.

*7. Prove there exists an infinite number of primes of the form $4n + 1$.

*8. Determine all the prime elements in $J[i]$.

*9. Determine all positive integers which can be written as a sum of two squares (of integers).

3.9 Polynomial Rings

Very early in our mathematical education—in fact in junior high school or early in high school itself—we are introduced to polynomials. For a seemingly endless amount of time we are drilled, to the point of utter boredom, in factoring them, multiplying them, dividing them, simplifying them. Facility in factoring a quadratic becomes confused with genuine mathematical talent.

Later, at the beginning college level, polynomials make their appearance in a somewhat different setting. Now they are functions, taking on values, and we become concerned with their continuity, their derivatives, their integrals, their maxima and minima.

We too shall be interested in polynomials but from neither of the above viewpoints. To us polynomials will simply be elements of a certain ring and we shall be concerned with algebraic properties of this ring. Our primary interest in them will be that they give us a Euclidean ring whose properties will be decisive in discussing fields and extensions of fields.

Let F be a field. By the *ring of polynomials* in the indeterminate, x, written as $F[x]$, we mean the set of all symbols $a_0 + a_1x + \cdots + a_nx^n$, where n can be any nonnegative integer and where the coefficients $a_1, a_2, \ldots, a_n$ are all in F. In order to make a ring out of $F[x]$ we must be able to recognize when two elements in it are equal, we must be able to add and multiply elements of $F[x]$ so that the axioms defining a ring hold true for $F[x]$. This will be our initial goal.

We could avoid the phrase "the set of all symbols" used above by introducing an appropriate apparatus of sequences but it seems more desirable to follow a path which is somewhat familiar to most readers.

DEFINITION If $p(x) = a_0 + a_1x + \cdots + a_mx^m$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$ are in $F[x]$, then $p(x) = q(x)$ if and only if for every integer $i \geq 0$, $a_i = b_i$.

Thus two polynomials are declared to be equal if and only if their corresponding coefficients are equal.

DEFINITION If $p(x) = a_0 + a_1x + \cdots + a_mx^m$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$ are both in $F[x]$, then $p(x) + q(x) = c_0 + c_1x + \cdots + c_tx^t$ where for each i, $c_i = a_i + b_i$.

In other words, add two polynomials by adding their coefficients and collecting terms. To add $1 + x$ and $3 - 2x + x^2$ we consider $1 + x$ as $1 + x + 0x^2$ and add, according to the recipe given in the definition, to obtain as their sum $4 - x + x^2$.

The most complicated item, and the only one left for us to define for $F[x]$, is the multiplication.

DEFINITION If $p(x) = a_0 + a_1x + \cdots + a_mx^m$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$, then $p(x)q(x) = c_0 + c_1x + \cdots + c_kx^k$ where $c_t = a_tb_0 + a_{t-1}b_1 + a_{t-2}b_2 + \cdots + a_0b_t$.

This definition says nothing more than: multiply the two polynomials by multiplying out the symbols formally, use the relation $x^\alpha x^\beta = x^{\alpha+\beta}$, and collect terms. Let us illustrate the definition with an example:

$$p(x) = 1 + x - x^2, \qquad q(x) = 2 + x^2 + x^3.$$

Here $a_0 = 1$, $a_1 = 1$, $a_2 = -1$, $a_3 = a_4 = \cdots = 0$, and $b_0 = 2$, $b_1 = 0$, $b_2 = 1$, $b_3 = 1$, $b_4 = b_5 = \cdots = 0$. Thus

$$c_0 = a_0b_0 = 1.2 = 2,$$

$$c_1 = a_1b_0 + a_0b_1 = 1.2 + 1.0 = 2,$$

$$c_2 = a_2b_0 + a_1b_1 + a_0b_2 = (-1)(2) + 1.0 + 1.1 = -1,$$

$$c_3 = a_3b_0 + a_2b_1 + a_1b_2 + a_0b_3 = (0)(2) + (-1)(0) + 1.1 + 1.1 = 2,$$

$$\begin{aligned} c_4 &= a_4b_0 + a_3b_1 + a_2b_2 + a_1b_3 + a_0b_4 \\ &= (0)(2) + (0)(0) + (-1)(1) + (1)(1) + 1(0) = 0, \end{aligned}$$

$$\begin{aligned} c_5 &= a_5b_0 + a_4b_1 + a_3b_2 + a_2b_3 + a_1b_4 + a_0b_5 \\ &= (0)(2) + (0)(0) + (0)(1) + (-1)(1) + (1)(0) + (0)(0) = -1, \end{aligned}$$

$$\begin{aligned} c_6 &= a_6b_0 + a_5b_1 + a_4b_2 + a_3b_3 + a_2b_4 + a_1b_5 + a_0b_6 \\ &= (0)(2) + (0)(0) + (0)(1) + (0)(1) + (-1)(0) + (1)(0) + (1)(0) = 0, \end{aligned}$$

$$c_7 = c_8 = \cdots = 0.$$

Therefore according to our definition,

$$(1 + x - x^2)(2 + x^2 + x^3) = c_0 + c_1x + \cdots = 2 + 2x - x^2 + 2x^3 - x^5.$$

If you multiply these together high-school style you will see that you get the same answer. Our definition of product is the one the reader has always known.

Without further ado we assert that $F[x]$ is a ring with these operations, its multiplication is commutative, and it has a unit element. We leave the verification of the ring axioms to the reader.

DEFINITION If $f(x) = a_0 + a_1x + \cdots + a_nx^n \neq 0$ and $a_n \neq 0$ then the *degree* of $f(x)$, written as $\deg f(x)$, is n.

That is, the degree of $f(x)$ is the largest integer i for which the ith coefficient of $f(x)$ is not 0. We do not define the degree of the zero polynomial. We say a polynomial is a *constant* if its degree is 0. The degree

function defined on the nonzero elements of $F[x]$ will provide us with the function $d(x)$ needed in order that $F[x]$ be a Euclidean ring.

LEMMA 3.9.1 *If $f(x)$, $g(x)$ are two nonzero elements of $F[x]$, then*

$$\deg (f(x)g(x)) = \deg f(x) + \deg g(x).$$

Proof. Suppose that $f(x) = a_0 + a_1x + \cdots + a_mx^m$ and $g(x) = b_0 + b_1x + \cdots + b_nx^n$ and that $a_m \neq 0$ and $b_n \neq 0$. Therefore $\deg f(x) = m$ and $\deg g(x) = n$. By definition, $f(x)g(x) = c_0 + c_1x + \cdots + c_kx^k$ where $c_t = a_tb_0 + a_{t-1}b_1 + \cdots + a_1b_{t-1} + a_0b_t$. We claim that $c_{m+n} = a_mb_n \neq 0$ and $c_i = 0$ for $i > m + n$. That $c_{m+n} = a_mb_n$ can be seen at a glance by its definition. What about c_i for $i > m + n$? c_i is the sum of terms of the form a_jb_{i-j}; since $i = j + (i - j) > m + n$ then either $j > m$ or $(i - j) > n$. But then one of a_j or b_{i-j} is 0, so that $a_jb_{i-j} = 0$; since c_i is the sum of a bunch of zeros it itself is 0, and our claim has been established. Thus the highest nonzero coefficient of $f(x)g(x)$ is c_{m+n}, whence $\deg f(x)g(x) = m + n = \deg f(x) + \deg g(x)$.

COROLLARY *If $f(x)$, $g(x)$ are nonzero elements in $F[x]$ then* $\deg f(x) \leq \deg f(x)g(x)$.

Proof. Since $\deg f(x)g(x) = \deg f(x) + \deg g(x)$, and since $\deg g(x) \geq 0$, this result is immediate from the lemma.

COROLLARY *$F[x]$ is an integral domain.*

We leave the proof of this corollary to the reader.

Since $F[x]$ is an integral domain, in light of Theorem 3.6.1 we can construct for it its field of quotients. This field merely consists of all quotients of polynomials and is called the field of *rational functions* in x over F.

The function $\deg f(x)$ defined for all $f(x) \neq 0$ in $F[x]$ satisfies

1. $\deg f(x)$ is a nonnegative integer.
2. $\deg f(x) \leq \deg f(x)g(x)$ for all $g(x) \neq 0$ in $F[x]$.

In order for $F[x]$ to be a Euclidean ring with the degree function acting as the d-function of a Euclidean ring we still need that given $f(x)$, $g(x) \in F[x]$, there exist $t(x)$, $r(x)$ in $F[x]$ such that $f(x) = t(x)g(x) + r(x)$ where either $r(x) = 0$ or $\deg r(x) < \deg g(x)$. This is provided us by

LEMMA 3.9.2 (The Division Algorithm) *Given two polynomials $f(x)$ and $g(x) \neq 0$ in $F[x]$, then there exist two polynomials $t(x)$ and $r(x)$ in $F[x]$ such that $f(x) = t(x)g(x) + r(x)$ where $r(x) = 0$ or $\deg r(x) < \deg g(x)$.*

Proof. The proof is actually nothing more than the "long-division" process we all used in school to divide one polynomial by another.

If the degree of $f(x)$ is smaller than that of $g(x)$ there is nothing to prove, for merely put $t(x) = 0$, $r(x) = f(x)$, and we certainly have that $f(x) = 0g(x) + f(x)$ where $\deg f(x) < \deg g(x)$ or $f(x) = 0$.

So we may assume that $f(x) = a_0 + a_1x + \cdots + a_mx^m$ and $g(x) = b_0 + b_1x + \cdots + b_nx^n$ where $a_m \neq 0$, $b_n \neq 0$ and $m \geq n$.

Let $f_1(x) = f(x) - (a_m/b_n)x^{m-n}g(x)$; thus $\deg f_1(x) \leq m - 1$, so by induction on the degree of $f(x)$ we may assume that $f_1(x) = t_1(x)g(x) + r(x)$ where $r(x) = 0$ or $\deg r(x) < \deg g(x)$. But then $f(x) - (a_m/b_n)x^{m-n}g(x) = t_1(x)g(x) + r(x)$, from which, by transposing, we arrive at $f(x) = ((a_m/b_n)x^{m-n} + t_1(x))g(x) + r(x)$. If we put $t(x) = (a_m/b_n)x^{m-n} + t_1(x)$ we do indeed have that $f(x) = t(x)g(x) + r(x)$ where $t(x)$, $r(x) \in F[x]$ and where $r(x) = 0$ or $\deg r(x) < \deg g(x)$. This proves the lemma.

This last lemma fills the gap needed to exhibit $F[x]$ as a Euclidean ring and we now have the right to say

THEOREM 3.9.1 *$F[x]$ is a Euclidean ring.*

All the results of Section 3.7 now carry over and we list these, for our particular case, as the following lemmas. It could be very instructive for the reader to try to prove these directly, adapting the arguments used in Section 3.7 for our particular ring $F[x]$ and its Euclidean function, the degree.

LEMMA 3.9.3 *$F[x]$ is a principal ideal ring.*

LEMMA 3.9.4 *Given two polynomials $f(x)$, $g(x)$ in $F[x]$ they have a greatest common divisor $d(x)$ which can be realized as $d(x) = \lambda(x)f(x) + \mu(x)g(x)$.*

What corresponds to a prime element?

DEFINITION A polynomial $p(x)$ in $F[x]$ is said to be *irreducible* over F if whenever $p(x) = a(x)b(x)$ with $a(x)$, $b(x) \in F[x]$, then one of $a(x)$ or $b(x)$ has degree 0 (i.e., is a constant).

Irreducibility depends on the field; for instance the polynomial $x^2 + 1$ is irreducible over the real field but not over the complex field, for there $x^2 + 1 = (x + i)(x - i)$ where $i^2 = -1$.

LEMMA 3.9.5 *Any polynomial in $F[x]$ can be written in a unique manner as a product of irreducible polynomials in $F[x]$.*

LEMMA 3.9.6 *The ideal $A = (p(x))$ in $F[x]$ is a maximal ideal if and only if $p(x)$ is irreducible over F.*

In Chapter 5 we shall return to take a much closer look at this field $F[x]/(p(x))$, but for now we should like to compute an example.

Let F be the field of rational numbers and consider the polynomial $p(x) = x^3 - 2$ in $F[x]$. As is easily verified, it is irreducible over F, whence $F[x]/(x^3 - 2)$ is a field. What do its elements look like? Let $A = (x^3 - 2)$, the ideal in $F[x]$ generated by $x^3 - 2$.

Any element in $F[x]/(x^3 - 2)$ is a coset of the form $f(x) + A$ of the ideal A with $f(x)$ in $F[x]$. Now, given any polynomial $f(x) \in F[x]$, by the division algorithm, $f(x) = t(x)(x^3 - 2) + r(x)$, where $r(x) = 0$ or $\deg r(x) < \deg (x^3 - 2) = 3$. Thus $r(x) = a_0 + a_1x + a_2x^2$ where a_0, a_1, a_2 are in F; consequently $f(x) + A = a_0 + a_1x + a_2x^2 + t(x)(x^3 - 2) + A = a_0 + a_1x + a_2x^2 + A$ since $t(x)(x^3 - 2)$ is in A, hence by the addition and multiplication in $F[x]/(x^3 - 2)$, $f(x) + A = (a_0 + A) + a_1(x + A) + a_2(x + A)^2$. If we put $t = x + A$, then every element in $F[x]/(x^3 - 2)$ is of the form $a_0 + a_1t + a_2t^2$ with a_0, a_1, a_2 in F. What about t? Since $t^3 - 2 = (x + A)^3 - 2 = x^3 - 2 + A = A = 0$ (since A is the zero element of $F[x]/(x^3 - 2)$) we see that $t^3 = 2$.

Also, if $a_0 + a_1t + a_2t^2 = b_0 + b_1t + b_2t^2$, then $(a_0 - b_0) + (a_1 - b_1)t + (a_2 - b_2)t^2 = 0$, whence $(a_0 - b_0) + (a_1 - b_1)x + (a_2 - b_2)x^2$ is in $A = (x^3 - 2)$. How can this be, since every element in A has degree at least 3? Only if $a_0 - b_0 + (a_1 - b_1)x + (a_2 - b_2)x^2 = 0$, that is, only if $a_0 = b_0$, $a_1 = b_1$, $a_2 = b_2$. Thus every element in $F[x]/(x^3 - 2)$ has a *unique* representation as $a_0 + a_1t + a_2t^2$ where $a_0, a_1, a_2 \in F$. By Lemma 3.9.6, $F[x]/(x^3 - 2)$ is a field. It would be instructive to see this directly; all that it entails is proving that if $a_0 + a_1t + a_2t^2 \neq 0$ then it has an inverse of the form $\alpha + \beta t + \gamma t^2$. Hence we must solve for α, β, γ in the relation $(a_0 + a_1t + a_2t^2)(\alpha + \beta t + \gamma t^2) = 1$, where not all of a_0 a_1, a_2 are 0. Multiplying the relation out and using $t^3 = 2$ we obtain $(a_0\alpha + 2a_2\beta + 2a_1\gamma) + (a_1\alpha + a_0\beta + 2a_2\gamma)t + (a_2\alpha + a_1\beta + a_0\gamma)t^2 = 1$; thus

$$a_0\alpha + 2a_2\beta + 2a_1\gamma = 1,$$

$$a_1\alpha + a_0\beta + 2a_2\gamma = 0,$$

$$a_2\alpha + a_1\beta + a_0\gamma = 0.$$

We can try to solve these three equations in the three unknowns α, β, γ. When we do so we find that a solution exists if and only if

$$a_0^3 + 2a_1^3 + 4a_2^3 - 6a_0a_1a_2 \neq 0.$$

Therefore the problem of proving directly that $F[x]/(x^3 - 2)$ is a field boils down to proving that the only solution in *rational* numbers of

$$a_0^3 + 2a_1^3 + 4a_2^3 = 6a_0a_1a_2 \tag{1}$$

is the solution $a_0 = a_1 = a_2 = 0$. We now proceed to show this. If a solution exists in rationals, by clearing of denominators we can show that a solution exists where a_0, a_1, a_2 are integers. Thus we may assume that a_0, a_1, a_2 are integers satisfying (1). We now assert that we may assume that a_0, a_1, a_2 have no common divisor other than 1, for if $a_0 = b_0 d$, $a_1 = b_1 d$, and $a_2 = b_2 d$, where d is their greatest common divisor, then substituting in (1) we obtain $d^3(b_0^3 + 2b_1^3 + 4b_2^3) = d^3(6b_0b_1b_2)$, and so $b_0^3 + 2b_1^3 + 4b_2^3 = 6b_0b_1b_2$. The problem has thus been reduced to proving that (1) has no solutions in integers which are relatively prime. But then (1) implies that a_0^3 is even, so that a_0 is even; substituting $a_0 = 2\alpha_0$ in (1) gives us $4\alpha_0^3 + a_1^3 + 2a_2^3 = 6\alpha_0 a_1 a_2$. Thus a_1^3, and so, a_1 is even; $a_1 = 2\alpha_1$. Substituting in (1) we obtain $2\alpha_0^3 + 4\alpha_1^3 + a_2^3 = 6\alpha_0\alpha_1 a_2$. Thus a_2^3, and so a_2, is even! But then a_0, a_1, a_2 have 2 as a common factor! This contradicts that they are relatively prime, and we have proved that the equation $a_0^3 + 2a_1^3 + 4a_2^3 = 6a_0a_1a_2$ has no rational solution other than $a_0 = a_1 = a_2 = 0$. Therefore we can solve for α, β, γ and $F[x]/(x^3 - 2)$ is seen, directly, to be a field.

Problems

1. Find the greatest common divisor of the following polynomials over F, the field of rational numbers:
   (a) $x^3 - 6x^2 + x + 4$ and $x^5 - 6x + 1$.
   (b) $x^2 + 1$ and $x^6 + x^3 + x + 1$.
2. Prove that
   (a) $x^2 + x + 1$ is irreducible over F, the field of integers mod 2.
   (b) $x^2 + 1$ is irreducible over the integers mod 7.
   (c) $x^3 - 9$ is irreducible over the integers mod 31.
   (d) $x^3 - 9$ is reducible over the integers mod 11.
3. Let F, K be two fields $F \subset K$ and suppose $f(x), g(x) \in F[x]$ are relatively prime in $F[x]$. Prove that they are relatively prime in $K[x]$.
4. (a) Prove that $x^2 + 1$ is irreducible over the field F of integers mod 11 and prove directly that $F[x]/(x^2 + 1)$ is a field having 121 elements.
   (b) Prove that $x^2 + x + 4$ is irreducible over F, the field of integers mod 11 and prove directly that $F[x]/(x^2 + x + 4)$ is a field having 121 elements.
   *(c) Prove that the fields of part (a) and part (b) are isomorphic.
5. Let F be the field of real numbers. Prove that $F[x]/(x^2 + 1)$ is a field isomorphic to the field of complex numbers.

*6. Define the *derivative* $f'(x)$ of the polynomial

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n$$

as

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1}.$$

Prove that if $f(x) \in F[x]$, where F is the field of rational numbers, then $f(x)$ is divisible by the square of a polynomial if and only if $f(x)$ and $f'(x)$ have a greatest common divisor $d(x)$ of positive degree.

7. If $f(x)$ is in $F[x]$, where F is the field of integers mod p, p a prime, and $f(x)$ is irreducible over F of degree n prove that $F[x]/(f(x))$ is a field with p^n elements.

3.10 Polynomials over the Rational Field

We specialize the general discussion to that of polynomials whose coefficients are rational numbers. Most of the time the coefficients will actually be integers. For such polynomials we shall be concerned with their irreducibility.

DEFINITION The polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n$, where the $a_0, a_1, a_2, \ldots, a_n$ are integers is said to be *primitive* if the greatest common divisor of $a_0, a_1, \ldots, a_n$ is 1.

LEMMA 3.10.1 *If $f(x)$ and $g(x)$ are primitive polynomials, then $f(x)g(x)$ is a primitive polynomial.*

Proof. Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ and $g(x) = b_0 + b_1x + \cdots + b_mx^m$. Suppose that the lemma was false; then all the coefficients of $f(x)g(x)$ would be divisible by some integer larger than 1, hence by some prime number p. Since $f(x)$ is primitive, p does not divide some coefficient a_i. Let a_j be the first coefficient of $f(x)$ which p does not divide. Similarly let b_k be the first coefficient of $g(x)$ which p does not divide. In $f(x)g(x)$ the coefficient of x^{j+k}, c_{j+k}, is

$$c_{j+k} = a_jb_k + (a_{j+1}b_{k-1} + a_{j+2}b_{k-2} + \cdots + a_{j+k}b_0) + (a_{j-1}b_{k+1} + a_{j-2}b_{k+2} + \cdots + a_0b_{j+k}). \tag{1}$$

Now by our choice of b_k, $p \mid b_{k-1}, b_{k-2}, \ldots$ so that $p \mid (a_{j+1}b_{k-1} + a_{j+2}b_{k-2} + \cdots + a_{j+k}b_0)$. Similarly, by our choice of a_j, $p \mid a_{j-1}, a_{j-2}, \ldots$ so that $p \mid (a_{j-1}b_{k+1} + a_{j-2}b_{k+2} + \cdots + a_0b_{k+j})$. By assumption, $p \mid c_{j+k}$. Thus by (1), $p \mid a_jb_k$, which is nonsense since $p \nmid a_j$ and $p \nmid b_k$. This proves the lemma.

DEFINITION The *content* of the polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n$, where the a's are integers, is the greatest common divisor of the integers $a_0, a_1, \ldots, a_n$.

Clearly, given any polynomial $p(x)$ with integer coefficients it can be written as $p(x) = dq(x)$ where d is the content of $p(x)$ and where $q(x)$ is a primitive polynomial.

THEOREM 3.10.1 (GAUSS' LEMMA) *If the primitive polynomial $f(x)$ can be factored as the product of two polynomials having rational coefficients, it can be factored as the product of two polynomials having integer coefficients.*

Proof. Suppose that $f(x) = u(x)v(x)$ where $u(x)$ and $v(x)$ have rational coefficients. By clearing of denominators and taking out common factors we can then write $f(x) = (a/b)\lambda(x)\mu(x)$ where a and b are integers and where both $\lambda(x)$ and $\mu(x)$ have integer coefficients and are primitive. Thus $bf(x) = a\lambda(x)\mu(x)$. The content of the left-hand side is b, since $f(x)$ is primitive; since both $\lambda(x)$ and $\mu(x)$ are primitive, by Lemma 3.10.1 $\lambda(x)\mu(x)$ is primitive, so that the content of the right-hand side is a. Therefore $a = b$, $(a/b) = 1$, and $f(x) = \lambda(x)\mu(x)$ where $\lambda(x)$ and $\mu(x)$ have integer coefficients. This is the assertion of the theorem.

DEFINITION A polynomial is said to be *integer monic* if all its coefficients are integers and its highest coefficient is 1.

Thus an integer monic polynomial is merely one of the form $x^n + a_1x^{n-1} + \cdots + a_n$ where the a's are integers. Clearly an integer monic polynomial is primitive.

COROLLARY *If an integer monic polynomial factors as the product of two non-constant polynomials having rational coefficients then it factors as the product of two integer monic polynomials.*

We leave the proof of the corollary as an exercise for the reader.

The question of deciding whether a given polynomial is irreducible or not can be a difficult and laborious one. Few criteria exist which declare that a given polynomial is or is not irreducible. One of these few is the following result:

THEOREM 3.10.2 (THE EISENSTEIN CRITERION) *Let $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ be a polynomial with integer coefficients. Suppose that for some prime number p, $p \nmid a_n$, $p \mid a_1, p \mid a_2, \ldots, p \mid a_0$, $p^2 \nmid a_0$. Then $f(x)$ is irreducible over the rationals.*

Proof. Without loss of generality we may assume that $f(x)$ is primitive, for taking out the greatest common factor of its coefficients does not disturb the hypotheses, since $p \nmid a_n$. If $f(x)$ factors as a product of two rational polynomials, by Gauss' lemma it factors as the product of two polynomials having integer coefficients. Thus if we assume that $f(x)$ is reducible, then

$$f(x) = (b_0 + b_1x + \cdots + b_rx^r)(c_0 + c_1x + \cdots + c_sx^s),$$

where the b's and c's are integers and where $r > 0$ and $s > 0$. Reading off

the coefficients we first get $a_0 = b_0c_0$. Since $p \mid a_0$, p must divide one of b_0 or c_0. Since $p^2 \nmid a_0$, p cannot divide both b_0 and c_0. Suppose that $p \mid b_0$, $p \nmid c_0$. Not all the coefficients $b_0, \ldots, b_r$ can be divisible by p; otherwise all the coefficients of $f(x)$ would be divisible by p, which is manifestly false since $p \nmid a_n$. Let b_k be the first b not divisible by p, $k \leq r < n$. Thus $p \mid b_{k-1}$ and the earlier b's. But $a_k = b_kc_0 + b_{k-1}c_1 + b_{k-2}c_2 + \cdots + b_0c_k$, and $p \mid a_k, p \mid b_{k-1}, b_{k-2}, \ldots, b_0$, so that $p \mid b_kc_0$. However, $p \nmid c_0, p \nmid b_k$, which conflicts with $p \mid b_kc_0$. This contradiction proves that we could not have factored $f(x)$ and so $f(x)$ is indeed irreducible.

Problems

1. Let D be a Euclidean ring, F its field of quotients. Prove the Gauss Lemma for polynomials with coefficients in D factored as products of polynomials with coefficients in F.
2. If p is a prime number, prove that the polynomial $x^n - p$ is irreducible over the rationals.
3. Prove that the polynomial $1 + x + \cdots + x^{p-1}$, where p is a prime number, is irreducible over the field of rational numbers. (*Hint:* Consider the polynomial $1 + (x + 1) + (x + 1)^2 + \cdots + (x + 1)^{p-1}$, and use the Eisenstein criterion.)
4. If m and n are relatively prime integers and if

$$\left(x - \frac{m}{n}\right) \mid (a_0 + a_1x + \cdots + a_rx^r),$$

where the a's are integers, prove that $m \mid a_0$ and $n \mid a_r$.

5. If a is rational and $x - a$ divides an integer monic polynomial, prove that a must be an integer.

3.11 Polynomial Rings over Commutative Rings

In defining the polynomial ring in one variable over a field F, no essential use was made of the fact that F was a field; all that was used was that F was a commutative ring. The field nature of F only made itself felt in proving that $F[x]$ was a Euclidean ring.

Thus we can imitate what we did with fields for more general rings. While some properties may be lost, such as "Euclideanism," we shall see that enough remain to lead us to interesting results. The subject could have been developed in this generality from the outset, and we could have obtained the particular results about $F[x]$ by specializing the ring to be a field. However, we felt that it would be healthier to go from the concrete to the abstract rather than from the abstract to the concrete. The price we

pay for this is repetition, but even that serves a purpose, namely, that of consolidating the ideas. Because of the experience gained in treating polynomials over fields, we can afford to be a little sketchier in the proofs here.

Let R be a commutative ring with unit element. By *the polynomial ring in x over R*, $R[x]$, we shall mean the set of formal symbols $a_0 + a_1x + \cdots + a_mx^m$, where $a_0, a_1, \ldots, a_m$ are in R, and where equality, addition, and multiplication are defined exactly as they were in Section 3.9. As in that section, $R[x]$ is a commutative ring with unit element.

We now define the *ring of polynomials in the n-variables* $x_1, \ldots, x_n$ over R, $R[x_1, \ldots, x_n]$, as follows: Let $R_1 = R[x_1]$, $R_2 = R_1[x_2]$, the polynomial ring in x_2 over $R_1, \ldots, R_n = R_{n-1}[x_n]$. R_n is called the ring of polynomials in $x_1, \ldots, x_n$ over R. Its elements are of the form $\sum a_{i_1 i_2 \ldots i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$, where equality and addition are defined coefficientwise and where multiplication is defined by use of the distributive law and the rule of exponents $(x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n})(x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}) = x_1^{i_1+j_1} x_2^{i_2+j_2} \cdots x_n^{i_n+j_n}$. Of particular importance is the case in which $R = F$ is a field; here we obtain the ring of polynomials in n-variables over a field.

Of interest to us will be the influence of the structure of R on that of $R[x_1, \ldots, x_n]$. The first result in this direction is

LEMMA 3.11.1 *If R is an integral domain, then so is $R[x]$.*

Proof. For $0 \neq f(x) = a_0 + a_1x + \cdots + a_mx^m$, where $a_m \neq 0$, in $R[x]$, we define the *degree* of $f(x)$ to be m; thus $\deg f(x)$ is the index of the highest nonzero coefficient of $f(x)$. If R is an integral domain we leave it as an exercise to prove that $\deg(f(x)g(x)) = \deg f(x) + \deg g(x)$. But then, for $f(x) \neq 0$, $g(x) \neq 0$, it is impossible to have $f(x)g(x) = 0$. That is, $R[x]$ is an integral domain.

Making successive use of the lemma immediately yields the

COROLLARY *If R is an integral domain, then so is $R[x_1, \ldots, x_n]$.*

In particular, when F is a field, $F[x_1, \ldots, x_n]$ must be an integral domain. As such, we can construct its field of quotients; *we call this the field of rational functions in $x_1, \ldots, x_n$ over F* and denote it by $F(x_1, \ldots, x_n)$. This field plays a vital role in algebraic geometry. For us it shall be of utmost importance in our discussion, in Chapter 5, of Galois theory.

However, we want deeper interrelations between the structures of R and of $R[x_1, \ldots, x_n]$ than that expressed in Lemma 3.11.1. Our development now turns in that direction.

Exactly in the same way as we did for Euclidean rings, we can speak about divisibility, units, etc., in arbitrary integral domains, R, with unit element. Two elements a, b in R are said to be *associates* if $a = ub$ where u

is a unit in R. An element a which is not a unit in R will be called *irreducible* (or a *prime element*) if, whenever $a = bc$ with b, c both in R, then one of b or c must be a unit in R. An irreducible element is thus an element which cannot be factored in a "nontrivial" way.

DEFINITION An integral domain, R, with unit element is a *unique factorization domain* if

a. Any nonzero element in R is either a unit or can be written as the product of a finite number of irreducible elements of R.
b. The decomposition in part (a) is unique up to the order and associates of the irreducible elements.

Theorem 3.7.2 asserts that a Euclidean ring is a unique factorization domain. The converse, however, is false; for example, the ring $F[x_1, x_2]$, where F is a field, is not even a principal ideal ring (hence is certainly not Euclidean), but as we shall soon see it is a unique factorization domain.

In general commutative rings we may speak about the greatest common divisors of elements; the main difficulty is that these, in general, might not exist. However, in unique factorization domains their existence is assured. This fact is not difficult to prove and we leave it as an exercise; equally easy are the other parts of

LEMMA 3.11.2 *If R is a unique factorization domain and if a, b are in R, then a and b have a greatest common divisor (a, b) in R. Moreover, if a and b are relatively prime (i.e., $(a, b) = 1$), whenever $a \mid bc$ then $a \mid c$.*

COROLLARY *If $a \in R$ is an irreducible element and $a \mid bc$, then $a \mid b$ or $a \mid c$.*

We now wish to transfer the appropriate version of the Gauss lemma (Theorem 3.10.1), which we proved for polynomials with integer coefficients, to the ring $R[x]$, where R is a unique factorization domain.

Given the polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$ in $R[x]$, then the *content of* $f(x)$ is defined to be the greatest common divisor of $a_0, a_1, \ldots, a_m$. It is unique within units of R. We shall denote the content of $f(x)$ by $c(f)$. A polynomial in $R[x]$ is said to be *primitive* if its content is 1 (that is, is a unit in R). Given any polynomial $f(x) \in R[x]$, we can write $f(x) = af_1(x)$ where $a = c(f)$ and where $f_1(x) \in R[x]$ is primitive. (Prove!) Except for multiplication by units of R this decomposition of $f(x)$, as an element of R by a primitive polynomial in $R[x]$, is unique. (Prove!)

The proof of Lemma 3.10.1 goes over completely to our present situation; the only change that must be made in the proof is to replace the prime number p by an irreducible element of R. Thus we have

LEMMA 3.11.3 *If R is a unique factorization domain, then the product of two primitive polynomials in $R[x]$ is again a primitive polynomial in $R[x]$.*

Given $f(x)$, $g(x)$ in $R[x]$ we can write $f(x) = af_1(x)$, $g(x) = bg_1(x)$, where $a = c(f)$, $b = c(g)$ and where $f_1(x)$ and $g_1(x)$ are primitive. Thus $f(x)g(x) = abf_1(x)g_1(x)$. By Lemma 3.11.3, $f_1(x)g_1(x)$ is primitive. Hence the content of $f(x)g(x)$ is ab, that is, it is $c(f)c(g)$. We have proved the

COROLLARY *If R is a unique factorization domain and if $f(x)$, $g(x)$ are in $R[x]$, then $c(fg) = c(f)c(g)$ (up to units).*

By a simple induction, the corollary extends to the product of a finite number of polynomials to read $c(f_1f_2\cdots f_k) = c(f_1)c(f_2)\cdots c(f_k)$.

Let R be a unique factorization domain. Being an integral domain, by Theorem 3.6.1, it has a field of quotients F. We can consider $R[x]$ to be a subring of $F[x]$. Given any polynomial $f(x) \in F[x]$, then $f(x) = (f_0(x)/a)$, where $f_0(x) \in R[x]$ and where $a \in R$. (Prove!) It is natural to ask for the relation, in terms of reducibility and irreducibility, of a polynomial in $R[x]$ considered as a polynomial in the larger ring $F[x]$

LEMMA 3.11.4 *If $f(x)$ in $R[x]$ is both primitive and irreducible as an element of $R[x]$, then it is irreducible as an element of $F[x]$. Conversely, if the primitive element $f(x)$ in $R[x]$ is irreducible as an element of $F[x]$, it is also irreducible as an element of $R[x]$.*

Proof. Suppose that the primitive element $f(x)$ in $R[x]$ is irreducible in $R[x]$ but is reducible in $F[x]$. Thus $f(x) = g(x)h(x)$, where $g(x)$, $h(x)$ are in $F[x]$ and are of positive degree. Now $g(x) = (g_0(x)/a)$, $h(x) = (h_0(x)/b)$, where $a, b \in R$ and where $g_0(x), h_0(x) \in R[x]$. Also $g_0(x) = \alpha g_1(x)$, $h_0(x) = \beta h_1(x)$, where $\alpha = c(g_0)$, $\beta = c(h_0)$, and $g_1(x)$, $h_1(x)$ are primitive in $R[x]$. Thus $f(x) = (\alpha\beta/ab)g_1(x)h_1(x)$, whence $abf(x) = \alpha\beta g_1(x)h_1(x)$. By Lemma 3.11.3, $g_1(x)h_1(x)$ is primitive, whence the content of the right-hand side is $\alpha\beta$. Since $f(x)$ is primitive, the content of the left-hand side is ab; but then $ab = \alpha\beta$; the implication of this is that $f(x) = g_1(x)h_1(x)$, and we have obtained a nontrivial factorization of $f(x)$ in $R[x]$, contrary to hypothesis. (Note: this factorization is nontrivial since each of $g_1(x)$, $h_1(x)$ are of the same degree as $g(x)$, $h(x)$, so cannot be units in $R[x]$ (see Problem 4).) We leave the converse half of the lemma as an exercise.

LEMMA 3.11.5 *If R is a unique factorization domain and if $p(x)$ is a primitive polynomial in $R[x]$, then it can be factored in a unique way as the product of irreducible elements in $R[x]$.*

Proof. When we consider $p(x)$ as an element in $F[x]$, by Lemma 3.9.5, we can factor it as $p(x) = p_1(x)\cdots p_k(x)$, where $p_1(x), p_2(x), \ldots, p_k(x)$ are

irreducible polynomials in $F[x]$. Each $p_i(x) = (f_i(x)/a_i)$, where $f_i(x) \in R[x]$ and $a_i \in R$; moreover, $f_i(x) = c_i q_i(x)$, where $c_i = c(f_i)$ and where $q_i(x)$ is primitive in $R[x]$. Thus each $p_i(x) = (c_i q_i(x)/a_i)$, where $a_i, c_i \in R$ and where $q_i(x) \in R[x]$ is primitive. Since $p_i(x)$ is irreducible in $F[x]$, $q_i(x)$ must also be irreducible in $F[x]$, hence by Lemma 3.11.4 *it is irreducible in* $R[x]$.

Now

$$p(x) = p_1(x) \cdots p_k(x) = \frac{c_1 c_2 \cdots c_k}{a_1 a_2 \cdots a_k} q_1(x) \cdots q_k(x),$$

whence $a_1 a_2 \cdots a_k p(x) = c_1 c_2 \cdots c_k q_1(x) \cdots q_k(x)$. Using the primitivity of $p(x)$ and of $q_1(x) \cdots q_k(x)$, we can read off the content of the left-hand side as $a_1 a_2 \cdots a_k$ and that of the right-hand side as $c_1 c_2 \cdots c_k$. Thus $a_1 a_2 \cdots a_k = c_1 c_2 \cdots c_k$, hence $p(x) = q_1(x) \cdots q_k(x)$. We have factored $p(x)$, in $R[x]$, as a product of irreducible elements.

Can we factor it in another way? If $p(x) = r_1(x) \cdots r_k(x)$, where the $r_i(x)$ are irreducible in $R[x]$, by the primitivity of $p(x)$, each $r_i(x)$ must be primitive, *hence irreducible in* $F[x]$ by Lemma 3.11.4. But by Lemma 3.9.5 we know unique factorization in $F[x]$; the net result of this is that the $r_i(x)$ and the $q_i(x)$ are equal (up to associates) in some order, hence $p(x)$ has a unique factorization as a product of irreducibles in $R[x]$.

We now have all the necessary information to prove the principal theorem of this section.

THEOREM 3.11.1 *If R is a unique factorization domain, then so is $R[x]$.*

Proof. Let $f(x)$ be an arbitrary element in $R[x]$. We can write $f(x)$ in a unique way as $f(x) = cf_1(x)$ where $c = c(f)$ is in R and where $f_1(x)$, in $R[x]$, is primitive. By Lemma 3.11.5 we can decompose $f_1(x)$ in a unique way as the product of irreducible elements of $R[x]$. What about c? Suppose that $c = a_1(x)a_2(x) \cdots a_m(x)$ in $R[x]$; then $0 = \deg c = \deg(a_1(x)) + \deg(a_2(x)) + \cdots + \deg(a_m(x))$. Therefore, each $a_i(x)$ must be of degree 0, that is, it must be an element of R. In other words, the only factorizations of c as an element of $R[x]$ are those it had as an element of R. In particular, an irreducible element in R is still irreducible in $R[x]$. Since R is a unique factorization domain, c has a unique factorization as a product of irreducible elements of R, hence of $R[x]$.

Putting together the unique factorization of $f(x)$ in the form $cf_1(x)$ where $f_1(x)$ is primitive and where $c \in R$ with the unique factorizability of c and of $f_1(x)$ we have proved the theorem.

Given R as a unique factorization domain, then $R_1 = R[x_1]$ is also a unique factorization domain. Thus $R_2 = R_1[x_2] = R[x_1, x_2]$ is also a unique factorization domain. Continuing in this pattern we obtain

COROLLARY 1 *If R is a unique factorization domain then so is $R[x_1, \ldots, x_n]$.*

A special case of Corollary 1 but of independent interest and importance is

COROLLARY 2 *If F is a field then $F[x_1, \ldots, x_n]$ is a unique factorization domain.*

Problems

1. Prove that $R[x]$ is a commutative ring with unit element whenever R is.
2. Prove that $R[x_1, \ldots, x_n] = R[x_{i_1}, \ldots, x_{i_n}]$, where $(i_1, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$.
3. If R is an integral domain, prove that for $f(x)$, $g(x)$ in $R[x]$, $\deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x))$.
4. If R is an integral domain with unit element, prove that any unit in $R[x]$ must already be a unit in R.
5. Let R be a commutative ring with no nonzero *nilpotent* elements (that is, $a^n = 0$ implies $a = 0$). If $f(x) = a_0 + a_1x + \cdots + a_mx^m$ in $R[x]$ is a zero-divisor, prove that there is an element $b \neq 0$ in R such that $ba_0 = ba_1 = \cdots = ba_m = 0$.

*6. Do Problem 5 dropping the assumption that R has no nonzero nilpotent elements.

*7. If R is a commutative ring with unit element, prove that $a_0 + a_1x + \cdots + a_nx^n$ in $R[x]$ has an inverse in $R[x]$ (i.e., is a unit in $R[x]$) if and only if a_0 is a unit in R and $a_1, \ldots, a_n$ are nilpotent elements in R.

8. Prove that when F is a field, $F[x_1, x_2]$ is not a principal ideal ring.
9. Prove, completely, Lemma 3.11.2 and its corollary.
10. (a) If R is a unique factorization domain, prove that every $f(x) \in R[x]$ can be written as $f(x) = af_1(x)$, where $a \in R$ and where $f_1(x)$ is primitive.
    (b) Prove that the decomposition in part (a) is unique (up to associates).
11. If R is an integral domain, and if F is its field of quotients, prove that any element $f(x)$ in $F[x]$ can be written as $f(x) = (f_0(x)/a)$, where $f_0(x) \in R[x]$ and where $a \in R$.
12. Prove the converse part of Lemma 3.11.4.
13. Prove Corollary 2 to Theorem 3.11.1.
14. Prove that a principal ideal ring is a unique factorization domain.
15. If J is the ring of integers, prove that $J[x_1, \ldots, x_n]$ is a unique factorization domain.

Supplementary Problems

1. Let R be a commutative ring; an ideal P of R is said to be a *prime ideal* of R if $ab \in P$, $a, b \in R$ implies that $a \in P$ or $b \in P$. Prove that P is a prime ideal of R if and only if R/P is an integral domain.
2. Let R be a commutative ring with unit element; prove that every maximal ideal of R is a prime ideal.
3. Give an example of a ring in which some prime ideal is not a maximal ideal.
4. If R is a finite commutative ring (i.e., has only a finite number of elements) with unit element, prove that every prime ideal of R is a maximal ideal of R.
5. If F is a field, prove that $F[x]$ is isomorphic to $F[t]$.
6. Find all the automorphisms σ of $F[x]$ with the property that $\sigma(f) = f$ for every $f \in F$.
7. If R is a commutative ring, let $N = \{x \in R \mid x^n = 0 \text{ for some integer } n\}$. Prove
   (a) N is an ideal of R.
   (b) In $\overline{R} = R/N$ if $\bar{x}^m = 0$ for some m then $\bar{x} = 0$.
8. Let R be a commutative ring and suppose that A is an ideal of R. Let $N(A) = \{x \in R \mid x^n \in A \text{ for some } n\}$. Prove
   (a) $N(A)$ is an ideal of R which contains A.
   (b) $N(N(A)) = N(A)$.
   $N(A)$ is often called the *radical* of A.
9. If n is an integer, let J_n be the ring of integers mod n. Describe N (see Problem 7) for J_n in terms of n.
10. If A and B are ideals in a ring R such that $A \cap B = (0)$, prove that for every $a \in A$, $b \in B$, $ab = 0$.
11. If R is a ring, let $Z(R) = \{x \in R \mid xy = yx \text{ all } y \in R\}$. Prove that $Z(R)$ is a subring of R.
12. If R is a division ring, prove that $Z(R)$ is a field.
13. Find a polynomial of degree 3 irreducible over the ring of integers, J_3, mod 3. Use it to construct a field having 27 elements.
14. Construct a field having 625 elements.
15. If F is a field and $p(x) \in F[x]$, prove that in the ring

$$R = \frac{F[x]}{(p(x))},$$

N (see Problem 7) is (0) if an only if $p(x)$ is not divisible by the square of any polynomial.

16. Prove that the polynomial $f(x) = 1 + x + x^3 + x^4$ is not irreducible over any field F.
17. Prove that the polynomial $f(x) = x^4 + 2x + 2$ is irreducible over the field of rational numbers.
18. Prove that if F is a finite field, its characteristic must be a prime number p and F contains p^n elements for some integer. Prove further that if $a \in F$ then $a^{p^n} = a$.
19. Prove that any nonzero ideal in the Gaussian integers $J[i]$ must contain some positive integer.
20. Prove that if R is a ring in which $a^4 = a$ for every $a \in R$ then R must be commutative.
21. Let R and R' be rings and ϕ a mapping from R into R' satisfying
    (a) $\phi(x + y) = \phi(x) + \phi(y)$ for every $x, y \in R$.
    (b) $\phi(xy) = \phi(x)\phi(y)$ or $\phi(y)\phi(x)$.
    Prove that for all $a, b \in R$, $\phi(ab) = \phi(a)\phi(b)$ or that, for all $a, b \in R$, $\phi(a) = \phi(b)\phi(a)$. (*Hint:* If $a \in R$, let
    $$W_a = \{x \in R \mid \phi(ax) = \phi(a)\phi(x)\}$$
    and
    $$U_a = \{x \in R \mid \phi(ax) = \phi(x)\phi(a)\}.)$$
22. Let R be a ring with a unit element, 1, in which $(ab)^2 = a^2b^2$ for all $a, b \in R$. Prove that R must be commutative.
23. Give an example of a noncommutative ring (of course, without 1) in which $(ab)^2 = a^2b^2$ for all elements a and b.
24. (a) Let R be a ring with unit element 1 such that $(ab)^2 = (ba)^2$ for all $a, b \in R$. If in R, $2x = 0$ implies $x = 0$, prove that R must be commutative.
    (b) Show that the result of (a) may be false if $2x = 0$ for some $x \neq 0$ in R.
    (c) Even if $2x = 0$ implies $x = 0$ in R, show that the result of (a) may be false if R does not have a unit element.
25. Let R be a ring in which $x^n = 0$ implies $x = 0$. If $(ab)^2 = a^2b^2$ for all $a, b \in R$, prove that R is commutative.
26. Let R be a ring in which $x^n = 0$ implies $x = 0$. If $(ab)^2 = (ba)^2$ for all $a, b \in R$, prove that R must be commutative.
27. Let $p_1, p_2, \ldots, p_k$ be distinct primes, and let $n = p_1p_2 \cdots p_k$. If R is the ring of integers modulo n, show that there are exactly 2^k elements a in R such that $a^2 = a$.
28. Construct a polynomial $q(x) \neq 0$ with integer coefficients which has no rational roots but is such that for any prime p we can solve the congruence $q(x) \equiv 0 \bmod p$ in the integers.

Supplementary Reading

ZARISKI, OSCAR, and SAMUEL, PIERRE, *Commutative Algebra*, Vol. 1. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1958.

MCCOY, N. H., *Rings and Ideals*, Carus Monograph No. 8. La Salle, Illinois: Open Court Publishing Company, 1948.

Topic for Class Discussion

MOTZKIN, T., "The Euclidean algorithm," *Bulletin of the American Mathematical Society*, Vol. 55 (1949), pages 1142–1146.

7

Selected Topics

In this final chapter we have set ourselves two objectives. Our first is to present some mathematical results which cut deeper than most of the material up to now, results which are more sophisticated, and are a little apart from the general development which we have followed. Our second goal is to pick results of this kind whose discussion, in addition, makes vital use of a large cross section of the ideas and theorems expounded earlier in the book. To this end we have decided on three items to serve as the focal points of this chapter.

The first of these is a celebrated theorem proved by Wedderburn in 1905 ("A Theorem on Finite Algebras," *Transactions of the American Mathematical Society*, Vol. 6 (1905), pages 349–352) which asserts that a division ring which has only a finite number of elements must be a commutative field. We shall give two proofs of this theorem, differing totally from each other. The first one will closely follow Wedderburn's original proof and will use a counting argument; it will lean heavily on results we developed in the chapter on group theory. The second one will use a mixture of group-theoretic and field-theoretic arguments, and will draw incisively on the material we developed in both these directions. The second proof has the distinct advantage that in the course of executing the proof certain side-results will fall out which will enable us to proceed to the proof, in the division ring case, of a beautiful theorem due to Jacobson ("Structure Theory for Algebraic Algebras of Bounded Degree," *Annals of Mathematics*, Vol. 46 (1945), pages 695–707) which is a far-reaching generalization of Wedderburn's theorem.

Our second high spot is a theorem due to Frobenius ("Über lineare Substitutionen und bilineare Formen," *Journal für die Reine und Angewandte Mathematik*, Vol. 84 (1877), especially pages 59–63) which states that the only division rings algebraic over the field of all real numbers are the field of real numbers, the field of complex numbers, and the division ring of real quaternions. The theorem points out a unique role for the quaternions, and makes it somewhat amazing that Hamilton should have discovered them in his somewhat ad hoc manner. Our proof of the Frobenius theorem, now quite elementary, is a variation of an approach laid out by Dickson and Albert; it will involve the theory of polynomials and fields.

Our third goal is the theorem that every positive integer can be represented as the sum of four squares. This famous result apparently was first conjectured by the early Greek mathematician Diophantos. Fermat grappled unsuccessfully with it and sadly announced his failure to solve it (in a paper where he did, however, solve the two-square theorem which we proved in Section 3.8). Euler made substantial inroads on the problem; basing his work on that of Euler, Lagrange in 1770 finally gave the first complete proof. Our approach will be entirely different from that of Lagrange. It is rooted in the work of Adolf Hurwitz and will involve a generalization of Euclidean rings. Using our ring-theoretic techniques on a certain ring of quaternions, the Lagrange theorem will drop out as a consequence.

En route to establishing these theorems many ideas and results, interesting in their own right, will crop up. This is characteristic of a good theorem—its proof invariably leads to side results of almost equal interest.

7.1 Finite Fields

Before we can enter into a discussion of Wedderburn's theorem and finite division rings, it is essential that we investigate the nature of fields having only a finite number of elements. Such fields are called *finite fields*. Finite fields do exist, for the ring J_p of integers modulo any prime p, provides us with an example of such. In this section we shall determine all possible finite fields and many of the important properties which they possess.

We begin with

LEMMA 7.1.1 *Let F be a finite field with q elements and suppose that $F \subset K$ where K is also a finite field. Then K has q^n elements where $n = [K:F]$.*

Proof. K is a vector space over F and since K is finite it is certainly finite dimensional as a vector space over F. Suppose that $[K:F] = n$; then K has a basis of n elements over F. Let such a basis be $v_1, v_2, \ldots, v_n$. Then every element in K has a unique representation in the form $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are all in F. Thus the number of

elements in K is the number of $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ as the α_1, $\alpha_2, \ldots, \alpha_n$ range over F. Since each coefficient can have q values K must clearly have q^n elements.

COROLLARY 1 *Let F be a finite field; then F has p^m elements where the prime number p is the characteristic of F.*

Proof. Since F has a finite number of elements, by Corollary 2 to Theorem 2.4.1, $f1 = 0$ where f is the number of elements in F. Thus F has characteristic p for some prime number p. Therefore F contains a field F_0 isomorphic to J_p. Since F_0 has p elements, F has p^m elements where $m = [F:F_0]$, by Lemma 7.1.1.

COROLLARY 2 *If the finite field F has p^m elements then every $a \in F$ satisfies $a^{p^m} = a$.*

Proof. If $a = 0$ the assertion of the corollary is trivially true.

On the other hand, the nonzero elements of F form a group under multiplication of order $p^m - 1$ thus by Corollary 2 to Theorem 2.4.1, $a^{p^m - 1} = 1$ for all $a \neq 0$ in F. Multiplying this relation by a we obtain that $a^{p^m} = a$.

From this last corollary we can easily pass to

LEMMA 7.1.2 *If the finite field F has p^m elements then the polynomial $x^{p^m} - x$ in $F[x]$ factors in $F[x]$ as $x^{p^m} - x = \prod_{\lambda \in F} (x - \lambda)$.*

Proof. By Lemma 5.3.2 the polynomial $x^{p^m} - x$ has at most p^{p^m} roots in F. However, by Corollary 2 to Lemma 7.1.1 we know p^m such roots, namely all the elements of F. By the corollary to Lemma 5.3.1 we can conclude that $x^{p^m} - x = \prod_{\lambda \in F} (x - \lambda)$.

COROLLARY *If the field F has p^m elements then F is the splitting field of the polynomial $x^{p^m} - x$.*

Proof. By Lemma 7.1.2, $x^{p^m} - x$ certainly splits in F. However, it cannot split in any smaller field for that field would have to have all the roots of this polynomial and so would have to have at least p^m elements. Thus F is the splitting field of $x^{p^m} - x$.

As we have seen in Chapter 5 (Theorem 5.3.4) any two splitting fields over a given field of a given polynomial are isomorphic. In light of the corollary to Lemma 7.1.2 we can state

LEMMA 7.1.3 *Any two finite fields having the same number of elements are isomorphic.*

Proof. If these fields have p^m elements, by the above corollary they are both splitting fields of the polynomial $x^{p^m} - x$, over J_p whence they are isomorphic.

Thus for any integer m and any prime number p there is, up to isomorphism, at most one field having p^m elements. The purpose of the next lemma is to demonstrate that for any prime number p and any integer m there is a field having p^m elements. When this is done we shall know that there is exactly one field having p^m elements where p is an arbitrary prime and m an arbitrary integer.

LEMMA 7.1.4 *For every prime number p and every positive integer m there exists a field having p^m elements.*

Proof. Consider the polynomial $x^{p^m} - x$ in $J_p[x]$, the ring of polynomials in x over J_p, the field of integers mod p. Let K be the splitting field of this polynomial. In K let $F = \{a \in K \mid a^{p^m} = a\}$. The elements of F are thus the roots of $x^{p^m} - x$, which by Corollary 2 to Lemma 5.5.2 are distinct; whence F has p^m elements. We now claim that F is a field. If $a, b \in F$ then $a^{p^m} = a$, $b^{p^m} = b$ and so $(ab)^{p^m} = a^{p^m}b^{p^m} = ab$; thus $ab \in F$. Also since the characteristic is p, $(a \pm b)^{p^m} = a^{p^m} \pm b^{p^m} = a \pm b$, hence $a \pm b \in F$. Consequently F is a subfield of K and so is a field. Having exhibited the field F having p^m elements we have proved Lemma 7.1.4.

Combining Lemmas 7.1.3 and 7.1.4 we have

THEOREM 7.1.1 *For every prime number p and every positive integer m there is a unique field having p^m elements.*

We now return to group theory for a moment. The group-theoretic result we seek will determine the structure of any finite multiplicative subgroup of the group of nonzero elements of any field, and, in particular, it will determine the multiplicative structure of any finite field.

LEMMA 7.1.5 *Let G be a finite abelian group enjoying the property that the relation $x^n = e$ is satisfied by at most n elements of G, for every integer n. Then G is a cyclic group.*

Proof. If the order of G is a power of some prime number q then the result is very easy. For suppose that $a \in G$ is an element whose order is as large as possible; its order must be q^r for some integer r. The elements $e, a, a^2, \ldots, a^{q^r - 1}$ give us q^r distinct solutions of the equation $x^{q^r} = e$, which, by our hypothesis, implies that these are all the solutions of this equation. Now if $b \in G$ its order is q^s where $s \leq r$, hence $b^{q^r} = (b^{q^s})^{q^{r-s}} = e$.

By the observation made above this forces $b = a^i$ for some i, and so G is cyclic.

The general finite abelian group G can be realized as $G = S_{q_1}S_{q_2}\ldots, S_{q_k}$ where the q_i are the distinct prime divisors of $o(G)$ and where the S_{q_i} are the Sylow subgroups of G. Moreover, every element $g \in G$ can be written in a *unique* way as $g = s_1 s_2, \ldots, s_k$ where $s_i \in S_{q_i}$ (see Section 2.7). Any solution of $x^n = e$ in S_{q_i} is one of $x^n = e$ in G so that each S_{q_i} inherits the hypothesis we have imposed on G. By the remarks of the first paragraph of the proof, each S_{q_i} is a cyclic group; let a_i be a generator of S_{q_i}. We claim that $c = a_1 a_2, \ldots, a_k$ is a cyclic generator of G. To verify this all we must do is prove that $o(G)$ divides m, the order of c. Since $c^m = e$, we have that $a_1{}^m a_2{}^m \cdots a_k{}^m = e$. By the uniqueness of representation of an element of G as a product of elements in the S_{q_i}, we conclude that each $a_i{}^m = e$. Thus $o(S_{q_i}) \mid m$ for every i. Thus $o(G) = o(S_{q_1})o(S_{q_2}) \cdots o(S_{q_k}) \mid m$. However, $m \mid o(G)$ and so $o(G) = m$. This proves that G is cyclic.

Lemma 7.1.5 has as an important consequence

LEMMA 7.1.6 *Let K be a field and let G be a finite subgroup of the multiplicative group of nonzero elements of K. Then G is a cyclic group.*

Proof. Since K is a field, any polynomial of degree n in $K[x]$ has at most n roots in K. Thus in particular, for any integer n, the polynomial $x^n - 1$ has at most n roots in K, and all the more so, at most n roots in G. The hypothesis of Lemma 7.1.5 is satisfied, so G is cyclic.

Even though the situation of a finite field is merely a special case of Lemma 7.1.6, it is of such widespread interest that we single it out as

THEOREM 7.1.2 *The multiplicative group of nonzero elements of a finite field is cyclic.*

Proof. Let F be a finite field. By merely applying Lemma 7.1.6 with $F = K$ and $G =$ the group of nonzero elements of F, the result drops out.

We conclude this section by using a counting argument to prove the existence of solutions of certain equations in a finite field. We shall need the result in one proof of the Wedderburn theorem.

LEMMA 7.1.7 *If F is a finite field and $\alpha \neq 0$, $\beta \neq 0$ are two elements of F then we can find elements a and b in F such that $1 + \alpha a^2 + \beta b^2 = 0$.*

Proof. If the characteristic of F is 2, F has 2^n elements and every element x in F satisfies $x^{2^n} = x$. Thus every element in F is a square. In particular $\alpha^{1} = a^2$ for some $a \in F$. Using this a and $b = 0$, we have

$1 + \alpha a^2 + \beta b^2 = 1 + \alpha\alpha^{-1} + 0 = 1 + 1 = 0$, the last equality being a consequence of the fact that the characteristic of F is 2.

If the characteristic of F is an odd prime p, F has p^n elements. Let $W_\alpha = \{1 + \alpha x^2 \mid x \in F\}$. How many elements are there in W_α? We must check how often $1 + \alpha x^2 = 1 + \alpha y^2$. But this relation forces $\alpha x^2 = \alpha y^2$ and so, since $\alpha \neq 0$, $x^2 = y^2$. Finally this leads to $x = \pm y$. Thus for $x \neq 0$ we get from each pair x and $-x$ one element in W_α, and for $x = 0$ we get $1 \in W_\alpha$. Thus W_α has $1 + (p^n - 1)/2 = (p^n + 1)/2$ elements. Similarly $W_\beta = \{-\beta x^2 \mid x \in F\}$ has $(p^n + 1)/2$ elements. Since each of W_α and W_β has more than half the elements of F they must have a nonempty intersection. Let $c \in W_\alpha \cap W_\beta$. Since $c \in W_\alpha$, $c = 1 + \alpha a^2$ for some $a \in F$; since $c \in W_\beta$, $c = -\beta b^2$ for some $b \in F$. Therefore $1 + \alpha a^2 = -\beta b^2$, which, on transposing yields the desired result $1 + \alpha a^2 + \beta b^2 = 0$.

Problems

1. By Theorem 7.1.2 the nonzero elements of J_p form a cyclic group under multiplication. Any generator of this group is called a *primitive root* of p.
   (a) Find primitive roots of: 17, 23, 31.
   (b) How many primitive roots does a prime p have?
2. Using Theorem 7.1.2 prove that $x^2 \equiv -1 \bmod p$ is solvable if and only if the odd prime p is of the form $4n + 1$.
3. If a is an integer not divisible by the odd prime p, prove that $x^2 \equiv a \bmod p$ is solvable for some integer x if and only if $a^{(p-1)/2} \equiv 1 \bmod p$. (This is called the *Euler criterion* that a be a quadratic residue mod p.)
4. Using the result of Problem 3 determine if:
   (a) 3 is a square mod 17.
   (b) 10 is a square mod 13.
5. If the field F has p^n elements prove that the automorphisms of F form a cyclic group of order n.
6. If F is a finite field, by the quaternions over F we shall mean the set of all $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ where $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in F$ and where addition and multiplication are carried out as in the real quaternions (i.e., $i^2 = j^2 = k^2 = ijk = -1$, etc.). Prove that the quaternions over a finite field *do not* form a division ring.

7.2 Wedderburn's Theorem on Finite Division Rings

In 1905 Wedderburn proved the theorem, now considered a classic, that a finite division ring must be a commutative field. This result has caught the imagination of most mathematicians because it is so unexpected, interrelating two seemingly unrelated things, namely the number of elements in a certain

algebraic system and the multiplication of that system. Aside from its intrinsic beauty the result has been very important and useful since it arises in so many contexts. To cite just one instance, the only known proof of the purely geometric fact that in a finite geometry the Desargues configuration implies that of Pappus (for the definition of these terms look in any good book on projective geometry) is to reduce the geometric problem to an algebraic one, and this algebraic question is then answered by invoking the Wedderburn theorem. For algebraists the Wedderburn theorem has served as a jumping-off point for a large area of research, in the 1940s and 1950s, concerned with the commutativity of rings.

THEOREM 7.2.1 (WEDDERBURN) *A finite division ring is necessarily a commutative field.*

First Proof. Let K be a finite division ring and let $Z = \{z \in K \mid zx = xz$ for all $x \in K\}$ be its center. If Z has q elements then, as in the proof of Lemma 7.1.1, it follows that K has q^n elements. Our aim is to prove that $Z = K$, or, equivalently, that $n = 1$.

If $a \in K$ let $N(a) = \{x \in K \mid xa = ax\}$. $N(a)$ clearly contains Z, and, as a simple check reveals, $N(a)$ is a subdivision ring of K. Thus $N(a)$ contains $q^{n(a)}$ elements for some integer $n(a)$. We claim that $n(a) \mid n$. For, the nonzero elements of $N(a)$ form a subgroup of order $q^{n(a)} - 1$ of the group of nonzero elements, under multiplication, of K which has $q^n - 1$ elements. By Lagrange's theorem (Theorem 2.4.1) $q^{n(a)} - 1$ is a divisor of $q^n - 1$; but this forces $n(a)$ to be a divisor of n (see Problem 1 at the end of this section).

In the group of nonzero elements of K we have the conjugacy relation used in Chapter 2, namely a is a conjugate of b if $a = x^{-1}bx$ for some $x \neq 0$ in K.

By Theorem 2.11.1 the number of elements in K conjugate to a is the index of the normalizer of a in the group of nonzero elements of K. Therefore the number of conjugates of a in K is $(q^n - 1)/(q^{n(a)} - 1)$. Now $a \in Z$ if and only if $n(a) = n$, thus by the class equation (see the corollary to Theorem 2.11.1)

$$q^n - 1 = q - 1 + \sum_{\substack{n(a) \mid n \\ n(a) \neq n}} \frac{q^n - 1}{q^{n(a)} - 1} \tag{1}$$

where the sum is carried out over one a in each conjugate class for a's *not* in the center.

The problem has been reduced to proving that no equation such as (1) can hold in the integers. Up to this point we have followed the proof in Wedderburn's original paper quite closely. He went on to rule out the possibility of equation (1) by making use of the following number-theoretic

result due to Birkhoff and Vandiver: for $n > 1$ there exists a prime number which is a divisor of $q^n - 1$ but is not a divisor of *any* $q^m - 1$ where m is a proper divisor of n, with the exceptions of $2^6 - 1 = 63$ whose prime factors already occur as divisors of $2^2 - 1$ and $2^3 - 1$, and $n = 2$, and q a prime of the form $2^k - 1$. If we grant this result, how would we finish the proof? This prime number would be a divisor of the left-hand side of (1) and also a divisor of each term in the sum occurring on the right-hand side since it divides $q^n - 1$ but not $q^{n(a)} - 1$; thus this prime would then divide $q - 1$ giving us a contradiction. The case $2^6 - 1$ still would need ruling out but that is simple. In case $n = 2$, the other possibility not covered by the above argument, there can be no subfield between Z and K and this forces $Z = K$. (Prove!—See Problem 2.)

However, we do not want to invoke the result of Birkhoff and Vandiver without proving it, and its proof would be too large a digression here. So we look for another artifice. Our aim is to find an integer which divides $(q^n - 1)/(q^{n(a)} - 1)$, for all divisors $n(a)$ of n except $n(a) = n$, but does not divide $q - 1$. Once this is done, equation (1) will be impossible unless $n = 1$ and, therefore, Wedderburn's theorem will have been proved. The means to this end is the theory of cyclotomic polynomials. (These have been mentioned in the problems at the end of Section 5.6.)

Consider the polynomial $x^n - 1$ considered as an element of $C[x]$ where C is the field of complex numbers. In $C[x]$

$$x^n - 1 = \prod (x - \lambda), \tag{2}$$

where this product is taken over all λ satisfying $\lambda^n = 1$.

A complex number θ is said to be a *primitive nth root of unity* if $\theta^n = 1$ but $\theta^m \neq 1$ for any positive integer $m < n$. The complex numbers satisfying $x^n = 1$ form a finite subgroup, under multiplication, of the complex numbers, so by Theorem 7.1.2 this group is cyclic. Any cyclic generator of this group must then be a primitive nth root of unity, so we know that such primitive roots exist. (Alternatively, $\theta = e^{2\pi i/n}$ yields us a primitive nth root of unity.)

Let $\Phi_n(x) = \prod (x - \theta)$ where this product is taken over all the primitive nth roots of unity. This polynomial is called a *cyclotomic* polynomial. We list the first few cyclotomic polynomials: $\Phi_1(x) = x - 1$, $\Phi_2(x) = x + 1$, $\Phi_3(x) = x^2 + x + 1$, $\Phi_4(x) = x^2 + 1$, $\Phi_5(x) = x^4 + x^3 + x^2 + x + 1$, $\Phi_6(x) = x^2 - x + 1$. Notice that these are all monic polynomials with integer coefficients.

Our first aim is to prove that in general $\Phi_n(x)$ is a monic polynomial with integer coefficients. We regroup the factored form of $x^n - 1$ as given in (2), and obtain

$$x^n - 1 = \prod_{d|n} \Phi_d(x). \tag{3}$$

By induction we assume that $\Phi_d(x)$ is a monic polynomial with integer coefficients for $d \mid n$, $d \neq n$. Thus $x^n - 1 = \Phi_n(x)\, g(x)$ where $g(x)$ is a monic polynomial with integer coefficients. Therefore,

$$\Phi_n(x) = \frac{x^n - 1}{g(x)},$$

which, on actual division (or by comparing coefficients), tells us that $\Phi_n(x)$ is a monic polynomial with integer coefficients.

We now claim that for any divisor d of n, where $d \neq n$,

$$\Phi_n(x) \left| \frac{x^n - 1}{x^d - 1} \right.$$

in the sense that the quotient is a polynomial with integer coefficients. To see this, first note that

$$x^d - 1 = \prod_{k \mid d} \Phi_k(x),$$

and since every divisor of d is also a divisor of n, by regouping terms on the right-hand side of (3) we obtain $x^d - 1$ on the right-hand side; also since $d < n$, $x^d - 1$ does not involve $\Phi_n(x)$. Therefore, $x^n - 1 = \Phi_n(x)(x^d - 1) f(x)$ where

$$f(x) = \prod_{\substack{k \mid n \\ k \nmid d}} \Phi_k(x)$$

has integer coefficients, and so

$$\Phi_n(x) \left| \frac{x^n - 1}{x^d - 1} \right.$$

in the sense that the quotient is a polynomial with integer coefficients. This establishes our claim.

For any integer t, $\Phi_n(t)$ is an integer and from the above as an integer divides $(t^n - 1)/(t^d - 1)$. In particular, returning to equation (1),

$$\Phi_n(q) \left| \frac{q^n - 1}{q^{n(a)} - 1} \right.$$

and $\Phi_n(q) \mid (q^n - 1)$; thus by (1), $\Phi_n(q) \mid (q - 1)$. We claim, however, that if $n > 1$ then $|\Phi_n(q)| > q - 1$. For $\Phi_n(q) = \prod (q - \theta)$ where θ runs over all primitive nth roots of unity and $|q - \theta| > q - 1$ for all $\theta \neq 1$ a root of unity (Prove!) whence $|\Phi_n(q)| = \prod |q - \theta| > q - 1$. Clearly, then $\Phi_n(q)$ cannot divide $q - 1$, leading us to a contradiction. We must, therefore, assume that $n = 1$, forcing the truth of the Wedderburn theorem.

Second Proof. Before explicitly examining finite division rings again, we prove some preliminary lemmas.

LEMMA 7.2.1 *Let R be a ring and let $a \in R$. Let T_a be the mapping of R into itself defined by $xT_a = xa - ax$. Then*

$$xT_a^m = xa^m - maxa^{m-1} + \frac{m(m-1)}{2} a^2xa^{m-2} - \frac{m(m-1)(m-2)}{3!} a^3xa^{m-3} + \cdots.$$

Proof. What is xT_a^2? $xT_a^2 = (xT_a)T_a = (xa - ax)T_a = (xa - ax)a - a(xa - ax) = xa^2 - 2axa + a^2x$. What about xT_a^3? $xT_a^3 = (xT_a^2)T_a = (xa^2 - 2axa + a^2x)a - a(xa^2 - 2axa + a^2x) = xa^3 - 3axa^2 + 3a^2xa - a^3x$. Continuing in this way, or by the use of induction, we get the result of Lemma 7.2.1.

COROLLARY *If R is a ring in which $px = 0$ for all $x \in R$, where p is a prime number, then $xT_a^{p^m} = xa^{p^m} - a^{p^m}x$.*

Proof. By the formula of Lemma 7.2.1, if $p = 2$, $xT_a^2 = xa^2 - a^2x$, since $2axa = 0$. Thus, $xT_a^4 = (xa^2 - a^2x)a^2 - a^2(xa^2 - a^2x) = xa^4 - a^4x$, and so on for $xT_a^{2^m}$.

If p is an odd prime, again by the formula of Lemma 7.2.1,

$$xT_a^p = xa^p - paxa^{p-1} + \frac{p(p-1)}{2} a^2xa^{p-2} + \cdots - a^px,$$

and since

$$p \left| \frac{p(p-1)\cdots(p-i+1)}{i!} \right.$$

for $i < p$, all the middle terms drop out and we are left with $xT_a^p = xa^p - a^px = xT_{a^p}$. Now $xT_a^{p^2} = x(T_{a^p})^p = xT_{a^{p^2}}$, and so on for the higher powers of p.

LEMMA 7.2.2 *Let D be a division ring of characteristic $p > 0$ with center Z, and let $P = \{0, 1, 2, \ldots, (p-1)\}$ be the subfield of Z isomorphic to J_p. Suppose that $a \in D$, $a \notin Z$ is such that $a^{p^n} = a$ for some $n \geq 1$. Then there exists an $x \in D$ such that*

1. $xax^{-1} \neq a$.
2. $xax^{-1} \in P(a)$ *the field obtained by adjoining a to P.*

Proof. Define the mapping T_a of D into itself by $yT_a = ya - ay$ for every $y \in D$.

$P(a)$ is a finite field, since a is algebraic over P and has, say, p^m elements. These all satisfy $u^{p^m} = u$. By the corollary to Lemma 7.2.1, $yT_a^{p^m} = ya^{p^m} - a^{p^m}y = ya - ay = yT_a$, and so $T_a^{p^m} = T_a$.

Now, if $\lambda \in P(a)$, $(\lambda x)T_a = (\lambda x)a - a(\lambda x) = \lambda xa - \lambda ax = \lambda(xa - ax) = \lambda(xT_a)$, since λ commutes with a. Thus the mapping λI of D into itself defined by $\lambda I: y \to \lambda y$ commutes with T_a for every $\lambda \in P(a)$. Now the polynomial

$$u^{p^m} - u = \prod_{\lambda \in P(a)} (u - \lambda)$$

by Lemma 7.2.1. *Since T_a commutes with λI for every $\lambda \in P(a)$*, and since $T_a{}^{p^m} = T_a$, we have that

$$0 = T_a{}^{p^m} - T_a = \prod_{\lambda \in P(a)} (T_a - \lambda I).$$

If for every $\lambda \neq 0$ in $P(a)$, $T_a - \lambda I$ annihilates no nonzero element in D (if $y(T_a - \lambda I) = 0$ implies $y = 0$), since $T_a(T_a - \lambda_1 I) \cdots (T_a - \lambda_k I) = 0$, where $\lambda_1, \ldots, \lambda_k$ are the nonzero elements of $P(a)$, we would get $T_a = 0$. That is, $0 = yT_a = ya - ay$ for every $y \in D$ forcing $a \in Z$ contrary to hypothesis. Thus there is a $\lambda \neq 0$ in $P(a)$ and an $x \neq 0$ in D such that $x(T_a - \lambda I) = 0$. Writing this out explicitly, $xa - ax - \lambda x = 0$; hence, $xax^{-1} = a + \lambda$ is in $P(a)$ and is not equal to a since $\lambda \neq 0$. This proves the lemma.

COROLLARY *In Lemma 7.2.2, $xax^{-1} = a^i \neq a$ for some integer i.*

Proof. Let a be of order s; then in the field $P(a)$ all the roots of the polynomial $u^s - 1$ are $1, a, a^2, \ldots, a^{s-1}$ since these are all distinct roots and they are s in number. Since $(xax^{-1})^s = xa^sx^{-1} = 1$, and since $xax^{-1} \in P(a)$, xax^{-1} is a root in $P(a)$ of $u^s - 1$, hence $xax^{-1} = a^i$.

We now have all the pieces that we need to carry out our second proof of Wedderburn's theorem.

Let D be a finite division ring and let Z be its center. By induction we may assume that any division ring having fewer elements than D is a commutative field.

We first remark that if $a, b \in D$ are such that $b^ta = ab^t$ but $ba \neq ab$, then $b^t \in Z$. For, consider $N(b^t) = \{x \in D \mid b^tx = xb^t\}$. $N(b^t)$ is a subdivision ring of D; if it were not D, by our induction hypothesis, it would be commutative. However, both a and b are in $N(b^t)$ and these do not commute; consequently, $N(b^t)$ is not commutative so must be all of D. Thus $b^t \in Z$.

Every nonzero element in D has finite order, so some positive power of it falls in Z. Given $w \in D$ let the *order of w relative to Z* be the smallest positive integer $m(w)$ such that $w^{m(w)} \in Z$. Pick an element a in D but not in Z having minimal possible order relative to Z, and let this order be r. *We claim that r is a prime number*, for if $r = r_1r_2$ with $1 < r_1 < r$ then a^{r_1} is not in Z. Yet $(a^{r_1})^{r_2} = a^r \in Z$, implying that a^{r_1} has an order relative to Z smaller than that of a.

By the corollary to Lemma 7.2.2 there is an $x \in D$ such that $xax^{-1} = a^i \neq a$; thus $x^2ax^{-2} = x(xax^{-1})x^{-1} = xa^ix^{-1} = (xax^{-1})^i = (a^i)^i = a^{i^2}$. Similarly, we get $x^r{}^{-1}ax^{(r-1)} = a^{i^r}{}^{-1}$. However, r is a prime number, thus by the little Fermat theorem (corollary to Theorem 2.4.1), $i^r{}^{-1} = 1 + u_0r$, hence $a^{i^{r-1}} = a^{1+u_0r} = aa^{u_0r} = \lambda a$ where $\lambda = a^{u_0r} \in Z$. Thus $x^{r-1}a = \lambda ax^{r-1}$. Since $x \notin Z$, by the minimal nature of r, x^{r-1} cannot be in Z. By the remark of the earlier paragraph, since $xa \neq ax$, $x^{r-1}a \neq ax^{r-1}$ and so $\lambda \neq 1$. Let $b = x^r{}^{-1}$; thus $bab^{-1} = \lambda a$; consequently, $\lambda^r a^r = (bab^{-1})^r = ba^rb^{-1} = a^r$ since $a^r \in Z$. This relation forces $\lambda^r = 1$.

We claim that if $y \in D$ then whenever $y^r = 1$, then $y = \lambda^i$ for some i, for in the *field* $Z(y)$ there are at most r roots of the polynomial $u^r - 1$; the elements $1, \lambda, \lambda^2, \ldots, \lambda^{r-1}$ in Z are all distinct since λ is of the prime order r and they already account for r roots of $u^r - 1$ in $Z(y)$, in consequence of which $y = \lambda^i$.

Since $\lambda^r = 1$, $b^r = \lambda^r b^r = (\lambda b)^r = (a^{-1}ba)^r = a^{-1}b^ra$ from which we get $ab^r = b^ra$. Since a commutes with b^r but does not commute with b, by the remark made earlier, b^r must be in Z. By Theorem 7.1.2 the multiplicative group of nonzero elements of Z is cyclic; let $\gamma \in Z$ by a generator. Thus $a^r = \gamma^j$, $b^r = \gamma^k$; if $j = sr$ then $a^r = \gamma^{sr}$, whence $(a/\gamma^s)^r = 1$; this would imply that $a/\gamma^s = \lambda^i$, leading to $a \in Z$, contrary to $a \notin Z$. Hence, $r \nmid j$; similarly $r \nmid k$. Let $a_1 = a^k$ and $b_1 = b^j$; a direct computation from $ba = \lambda ab$ leads to $a_1b_1 = \mu b_1a_1$ where $\mu = \lambda^{-jk} \in Z$. Since the prime number r which is the order of λ does not divide j or k, $\lambda^{jk} \neq 1$ hence $\mu \neq 1$. Note that $\mu^r = 1$.

Let us see where we are. We have produced two elements a_1, b_1 such that

1. $a_1^r = b_1^r = \alpha \in Z$.
2. $a_1b_1 = \mu b_1a_1$ with $\mu \neq 1$ in Z.
3. $\mu^r = 1$.

We compute $(a_1{}^{1}b_1)^r$; $(a_1{}^{1}b_1)^2 = a_1{}^{-1}b_1a_1{}^{-1}b_1 = a_1{}^{1}(b_1a_1{}^{-1})b_1 = a_1{}^{1}(\mu a_1{}^{-1}b_1)b_1 = \mu a_1{}^{2}b_1{}^{2}$. If we compute $(a_1{}^{1}b_1)^3$ we find it equal to $\mu^{1+2}a_1{}^{3}b_1{}^{3}$. Continuing, we obtain $(a_1{}^{1}b_1)^r = \mu^{1+2+\cdots+(r-1)}a_1{}^{r}b_1{}^{r} = \mu^{1+2+\cdots+(r-1)} = \mu^{r(r-1)/2}$. If r is an odd prime, since $\mu^r = 1$, we get $\mu^{r(r-1)/2} = 1$, whence $(a_1{}^{-1}b_1)^r = 1$. Being a solution of $y^r = 1$, $a_1{}^{1}b_1 = \lambda^i$ so that $b_1 = \lambda^i a_1$; but then $\mu b_1a_1 = a_1b_1 = b_1a_1$, contradicting $\mu \neq 1$. Thus if r is an odd prime number, the theorem is proved.

We must now rule out the case $r = 2$. In that special situation we have two elements $a_1, b_1 \in D$ such that $a_1{}^2 = b_1{}^2 = \alpha \in Z$, $a_1b_1 = \mu b_1a_1$ where $\mu^2 = 1$ and $\mu \neq 1$. Thus $\mu = -1$ and $a_1b_1 = -b_1a_1 \neq b_1a_1$; in consequence, the characteristic of D is *not* 2. By Lemma 7.1.7 we can find elements $\zeta, \eta \in Z$ such that $1 + \zeta^2 - \alpha\eta^2 = 0$. Consider $(a_1 + \zeta b_1 + \eta a_1b_1)^2$; on computing this out we find that $(a_1 + \zeta b_1 + \eta a_1b_1)^2 = \alpha(1 + \zeta^2 - \alpha\eta^2) = 0$. Being in a division ring this yields that $a_1 + \zeta b_1 + \eta a_1b_1 = 0$; thus $0 \neq$

$2a_1^2 = a_1(a_1 + \zeta b_1 + \eta a_1 b_1) + (a_1 + \zeta b_1 + \eta a_1 b_1)a_1 = 0$. This contradiction finishes the proof and Wedderburn's theorem is established.

This second proof has some advantages in that we can use parts of it to proceed to a remarkable result due to Jacobson, namely,

THEOREM 7.2.2 (JACOBSON) *Let D be a division ring such that for every $a \in D$ there exists a positive integer $n(a) > 1$, depending on a, such that $a^{n(a)} = a$. Then D is a commutative field.*

Proof. If $a \neq 0$ is in D then $a^n = a$ and $(2a)^m = 2a$ for some integers $n, m > 1$. Let $s = (n-1)(m-1) + 1$; $s > 1$ and a simple calculation shows that $a^s = a$ and $(2a)^s = 2a$. But $(2a)^s = 2^s a^s = 2^s a$, whence $2^s a = 2a$ from which we get $(2^s - 2)a = 0$. Thus D has characteristic $p > 0$. If $P \subset Z$ is the field having p elements (isomorphic to J_p), since a is algebraic over P, $P(a)$ has a finite number of elements, in fact, p^h elements for some integer h. Thus, since $a \in P(a)$, $a^{p^h} = a$. Therefore, if $a \notin Z$ all the conditions of Lemma 7.2.2 are satisfied, hence there exists a $b \in D$ such that

$$bab^{-1} = a^\mu \neq a. \tag{1}$$

By the same argument, $b^{p^k} = b$ for some integer $k > 1$. Let

$$W = \left\{ x \in D \mid x = \sum_{i=1}^{p^h} \sum_{j=1}^{p^k} p_{ij} a^i b^j \text{ where } p_{ij} \in P \right\}.$$

W is finite and is closed under addition. By virtue of (1) it is also closed under multiplication. (Verify!) Thus W is a finite ring, and being a subring of the division ring D, it itself must be a division ring (Problem 3). Thus W is a finite division ring; by Wedderburn's theorem it is commutative. But a and b are both in W; therefore, $ab = ba$ contrary to $a^\mu b = ba$. This proves the theorem.

Jacobson's theorem actually holds for *any* ring R satisfying $a^{n(a)} = a$ for every $a \in R$, not just for division rings. The transition from the division ring case to the general case, while not difficult, involves the axiom of choice, and to discuss it would take us too far afield.

Problems

1. If $t > 1$ is an integer and $(t^m - 1) | (t^n - 1)$, prove that $m \mid n$.
2. If D is a division ring, prove that its dimension (as a vector space) over its center cannot be 2.
3. Show that any finite subring of a division ring is a division ring.

4. (a) Let D be a division ring of characteristic $p \neq 0$ and let G be a finite subgroup of the group of nonzero elements of D under multiplication. Prove that G is abelian. (*Hint:* consider the subset $\{x \in D \mid x = \sum \lambda_i g_i,\ \lambda_i \in P,\ g_i \in G\}$.)
   (b) In part (a) prove that G is actually cyclic.

*5. (a) If R is a finite ring in which $x^n = x$, for all $x \in R$ where $n > 1$ prove that R is commutative.
   (b) If R is a finite ring in which $x^2 = 0$ implies that $x = 0$, prove that R is commutative.

*6. Let D be a division ring and suppose that $a \in D$ only has a finite number of conjugates (i.e., only a finite number of distinct $x^{-1}ax$). Prove that a has only one conjugate and must be in the center of D.

7. Use the result of Problem 6 to prove that if a polynomial of degree n having coefficients in the center of a division ring has $n + 1$ roots in the division ring then it has an infinite number of roots in that division ring.

*8. Let D be a division ring and K a subdivision ring of D such that $xKx^{-1} \subset K$ for every $x \neq 0$ in D. Prove that either $K \subset Z$, the center of D or $K = D$. (This result is known as the *Brauer-Cartan-Hua theorem.*)

*9. Let D be a division ring and K a subdivision ring of D. Suppose that the group of nonzero elements of K is a subgroup of finite index in the group (under multiplication) of nonzero elements of D. Prove that either D is finite or $K = D$.

10. If $\theta \neq 1$ is a root of unity and if q is a positive integer, prove that $|q - \theta| > q - 1$.

7.3 A Theorem of Frobenius

In 1877 Frobenius classified all division rings having the field of real numbers in their center and satisfying, in addition, one other condition to be described below. The aim of this section is to present this result of Frobenius.

In Chapter 6 we brought attention to two important facts about the field of complex numbers. We recall them here:

FACT 1 Every polynomial of degree n over the field of complex numbers has all its n roots in the field of complex numbers.

FACT 2 The only irreducible polynomials over the field of real numbers are of degree 1 or 2.

DEFINITION A division algebra D is said to be *algebraic over a field* F if

1. F is contained in the center of D;
2. every $a \in D$ satisfies a nontrivial polynomial with coefficients in F.

If D, as a vector space, is finite-dimensional over the field F which is contained in its center, it can easily be shown that D is algebraic over F (see Problem 1, end of this section). However, it can happen that D is algebraic over F yet is not finite-dimensional over F.

We start our investigation of division rings algebraic over the real field by first finding those algebraic over the complex field.

LEMMA 7.3.1 *Let C be the field of complex numbers and suppose that the division ring D is algebraic over C. Then $D = C$.*

Proof. Suppose that $a \in D$. Since D is algebraic over C, $a^n + \alpha_1 a^{n-1} + \cdots + \alpha_{n-1}a + \alpha_n = 0$ for some $\alpha_1, \alpha_2, \ldots, \alpha_n$ in C.

Now the polynomial $p(x) = x^n + \alpha_1 x^{n-1} + \cdots + \alpha_{n-1}x + \alpha_n$ in $C[x]$, by Fact 1, can be factored, in $C[x]$, into a product of linear factors; that is, $p(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_n)$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all in C. Since C is in the center of D, every element of C commutes with a, hence $p(a) = (a - \lambda_1)(a - \lambda_2)\cdots(a - \lambda_n)$. But, by assumption, $p(a) = 0$, thus $(a - \lambda_1)(a - \lambda_2)\cdots(a - \lambda_n) = 0$. Since a product in a division ring is zero only if one of the terms of the product is zero, we conclude that $a - \lambda_k = 0$ for some k, hence $a = \lambda_k$, from which we get that $a \in C$. Therefore, every element of D is in C; since $C \subset D$, we obtain $D = C$.

We are now in a position to prove the classic result of Frobenius, namely,

THEOREM 7.3.1 (FROBENIUS) *Let D be a division ring algebraic over F, the field of real numbers. Then D is isomorphic to one of: the field of real numbers, the field of complex numbers, or the division ring of real quaternions.*

Proof. The proof consists of three parts. In the first, and easiest, we dispose of the commutative case; in the second, assuming that D is not commutative, we construct a replica of the real quaternions in D; in the third part we show that this replica of the quaternions fills out all of D.

Suppose that $D \neq F$ and that a is in D but not in F. By our assumptions, a satisfies some polynomial over F, hence some irreducible polynomial over F. In consequence of Fact 2, a satisfies either a linear or quadratic equation over F. If this equation is linear, a must be in F contrary to assumption. So we may suppose that $a^2 - 2\alpha a + \beta = 0$ where $\alpha, \beta \in F$. Thus $(a - \alpha)^2 = \alpha^2 - \beta$; we claim that $\alpha^2 - \beta < 0$ for, otherwise, it would have a real square root δ and we would have $a - \alpha = \pm\delta$ and so a would be in F. Since $\alpha^2 - \beta < 0$ it can be written as $-\gamma^2$ where $\gamma \in F$. Consequently $(a - \alpha)^2 = -\gamma^2$, whence $[(a - \alpha)/\gamma]^2 = -1$. *Thus if $a \in D$, $a \notin F$ we can find real α, γ such that $[(a - \alpha)/\gamma]^2 = -1$.*

If D is commutative, pick $a \in D$, $a \notin F$ and let $i = (a - \alpha)/\gamma$ where α, γ in F are chosen so as to make $i^2 = -1$. Therefore D contains $F(i)$, a field isomorphic to the field of complex numbers. Since D is commutative and

algebraic over F it is, all the more so, algebraic over $F(i)$. By Lemma 7.3.1 we conclude that $D = F(i)$. Thus if D is commutative it is either F or $F(i)$.

Assume, then, that D is *not* commutative. We claim that the center of D must be exactly F. If not, there is an a in the center, a not in F. But then for some $\alpha, \gamma \in F$, $[(a - \alpha)/\gamma]^2 = -1$ so that the center contains a field isomorphic to the complex numbers. However, by Lemma 7.3.1 if the complex numbers (or an isomorph of them) were in the center of D then $D = C$ forcing D to be commutative. Hence F is the center of D.

Let $a \in D$, $a \notin F$; for some $\alpha, \gamma \in F$, $i = (a - \alpha)/\gamma$ satisfies $i^2 = -1$. Since $i \notin F$, i is not in the center of F. Therefore there is an element $b \in D$ such that $c = bi - ib \neq 0$. We compute $ic + ci$; $ic + ci = i(bi - ib) + (bi - ib)i = ibi - i^2b + bi^2 - ibi = 0$ since $i^2 = -1$. Thus $ic = -ci$; from this we get $ic^2 = -c(ic) = -c(-ci) = c^2i$, and so c^2 commutes with i. Now c satisfies some quadratic equation over F, $c^2 + \lambda c + \mu = 0$. Since c^2 and μ commute with i, λc must commute with i; that is, $\lambda ci = i\lambda c = \lambda ic = -\lambda ci$, hence $2\lambda ci = 0$, and since $2ci \neq 0$ we have that $\lambda = 0$. Thus $c^2 = -\mu$; since $c \notin F$ (for $ci = -ic \neq ic$) we can say, as we have before, that μ is positive and so $\mu = \nu^2$ where $\nu \in F$. Therefore $c^2 = -\nu^2$; let $j = c/\nu$. Then j satisfies

1. $j^2 = \dfrac{c^2}{\nu^2} = -1.$

2. $ji + ij = \dfrac{c}{\nu} i + i \dfrac{c}{\nu} = \dfrac{ci + ic}{\nu} = 0.$

Let $k = ij$. The i, j, k we have constructed behave like those for the quaternions, whence $T = \{\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k \mid \alpha_0, \alpha_1, \alpha_2, \alpha_3 \in F\}$ forms a subdivision ring of D isomorphic to the real quaternions. We have produced a replica, T, of the division ring of real quaternions in D!

Our last objective is to demonstrate that $T = D$.

If $r \in D$ satisfies $r^2 = -1$ let $N(r) = \{x \in D \mid xr = rx\}$. $N(r)$ is a subdivision ring of D; moreover r, and so all $\alpha_0 + \alpha_1 r$, $\alpha_0, \alpha_1 \in F$, are in the center of $N(r)$. By Lemma 7.3.1 it follows that $N(r) = \{\alpha_0 + \alpha_1 r \mid \alpha_0, \alpha_1 \in F\}$. Thus if $xr = rx$ then $x = \alpha_0 + \alpha_1 r$ for some α_0, α_1 in F.

Suppose that $u \in D$, $u \notin F$. For some $\alpha, \beta \in F$, $w = (u - \alpha)/\beta$ satisfies $w^2 = -1$. We claim that $wi + iw$ commutes with both i and w; for $i(wi + iw) = iwi + i^2w = iwi + wi^2 = (iw + wi)i$ since $i^2 = -1$. Similarly $w(wi + iw) = (wi + iw)w$. By the remark of the preceding paragraph, $wi + iw = \alpha_0' + \alpha_1' i = \alpha_0 + \alpha_1 w$. If $w \notin T$ this last relation forces $\alpha_1 = 0$ (for otherwise we could solve for w in terms of i). Thus $wi + iw = \alpha_0 \in F$. Similarly $wj + jw = \beta_0 \in F$ and $wk + kw = \gamma_0 \in F$. Let

$$z = w + \frac{\alpha_0}{2} i + \frac{\beta_0}{2} j + \frac{\gamma_0}{2} k.$$

Then

$$zi + iz = wi + iw + \frac{\alpha_0}{2}(i^2 + i^2) + \frac{\beta_0}{2}(ji + ij) + \frac{\gamma_0}{2}(ki + ik)$$
$$= \alpha_0 - \alpha_0 = 0;$$

similarly $zj + jz = 0$ and $zk + kz = 0$. We claim these relations force z to be 0. For $0 = zk + kz = zij + ijz = (zi + iz)j + i(jz - zj) = i(jz - zj)$ since $zi + iz = 0$. However $i \neq 0$, and since we are in a division ring, it follows that $jz - zj = 0$. But $jz + zj = 0$. Thus $2jz = 0$, and since $2j \neq 0$ we have that $z = 0$. Going back to the expression for z we get

$$w + \frac{\alpha_0}{2}i + \frac{\beta_0}{2}j + \frac{\gamma_0}{2}k = 0,$$

hence $w \in T$, contradicting $w \notin T$. Thus, indeed, $w \in T$. Since $w = (u - \alpha)/\beta$, $u = \beta w + \alpha$ and so $u \in T$. We have proved that any element in D is in T. Since $T \subset D$ we conclude that $D = T$; because T is isomorphic to the real quaternions we now get that D is isomorphic to the division ring of real quaternions. This, however, is just the statement of the theorem.

Problems

1. If the division ring D is finite-dimensional, as a vector space, over the field F contained in the center of D, prove that D is algebraic over F.
2. Give an example of a field K algebraic over another field F but not finite-dimensional over F.
3. If A is a ring algebraic over a field F and A has no zero divisors prove that A is a division ring.

7.4 Integral Quaternions and the Four-Square Theorem

In Chapter 3 we considered a certain special class of integral domains called Euclidean rings. When the results about this class of rings were applied to the ring of Gaussian integers, we obtained, as a consequence, the famous result of Fermat that every prime number of the form $4n + 1$ is the sum of two squares.

We shall now consider a particular subring of the quaternions which, in all ways except for its lack of commutativity, will look like a Euclidean ring. Because of this it will be possible to explicitly characterize all its left-ideals. This characterization of the left-ideals will lead us quickly to a proof of the classic theorem of Lagrange that every positive integer is a sum of four squares.

Let Q be the division ring of real quaternions. In Q we now proceed to introduce an adjoint operation, $*$, by making the

DEFINITION For $x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ in Q the *adjoint* of x, denoted by x^*, is defined by $x^* = \alpha_0 - \alpha_1 i - \alpha_2 j - \alpha_3 k$.

LEMMA 7.4.1 *The adjoint in Q satisfies*

1. $x^{**} = x$;
2. $(\delta x + \gamma y)^* = \delta x^* + \gamma y^*$;
3. $(xy)^* = y^* x^*$;

for all x, y in Q and all real δ and γ.

Proof. If $x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ then $x^* = \alpha_0 - \alpha_1 i - \alpha_2 j - \alpha_3 k$, whence $x^{**} = (x^*)^* = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, proving part 1.

Let $x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ and $y = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 k$ be in Q and let δ and γ be arbitrary real numbers. Thus $\delta x + \gamma y = (\delta\alpha_0 + \gamma\beta_0) + (\delta\alpha_1 + \gamma\beta_1)i + (\delta\alpha_2 + \gamma\beta_2)j + (\delta\alpha_3 + \gamma\beta_3)k$; therefore by the definition of the $*$, $(\delta x + \gamma y)^* = (\delta\alpha_0 + \gamma\beta_0) - (\delta\alpha_1 + \gamma\beta_1)i - (\delta\alpha_2 + \gamma\beta_2)j - (\delta\alpha_3 + \gamma\beta_3)k = \delta(\alpha_0 - \alpha_1 i - \alpha_2 j - \alpha_3 k) + \gamma(\beta_0 - \beta_1 i - \beta_2 j - \beta_3 k) = \delta x^* + \gamma y^*$. This, of course, proves part 2.

In light of part 2, to prove 3 it is enough to do so for a basis of Q over the reals. We prove it for the particular basis $1, i, j, k$. Now $ij = k$, hence $(ij)^* = k^* = -k = ji = (-j)(-i) = j^* i^*$. Similarly $(ik)^* = k^* i^*$, $(jk)^* = k^* j^*$. Also $(i^2)^* = (-1)^* = -1 = (i^*)^2$, and similarly for j and k. Since part 3 is true for the basis elements and part 2 holds, 3 is true for all linear combinations of the basis elements with real coefficients, hence 3 holds for arbitrary x and y in Q.

DEFINITION If $x \in Q$ then the *norm* of x, denoted by $N(x)$, is defined by $N(x) = xx^*$.

Note that if $x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ then $N(x) = xx^* = (\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k)(\alpha_0 - \alpha_1 i - \alpha_2 j - \alpha_3 k) = {\alpha_0}^2 + {\alpha_1}^2 + {\alpha_2}^2 + {\alpha_3}^2$; therefore $N(0) = 0$ and $N(x)$ is a *positive* real number for $x \neq 0$ in Q. In particular, for any real number α, $N(\alpha) = \alpha^2$. If $x \neq 0$ note that $x^{-1} = [1/N(x)]x^*$.

LEMMA 7.4.2 *For all $x, y \in Q$, $N(xy) = N(x)N(y)$.*

Proof. By the very definition of norm, $N(xy) = (xy)(xy)^*$; by part 3 of Lemma 7.4.1, $(xy)^* = y^* x^*$ and so $N(xy) = xyy^* x^*$. However, $yy^* = N(y)$ is a real number, and thereby it is in the center of Q; in particular it must commute with x^*. Consequently $N(xy) = x(yy^*)x^* = (xx^*)(yy^*) = N(x)N(y)$.

As an immediate consequence of Lemma 7.4.2 we obtain

LEMMA 7.4.3 (LAGRANGE IDENTITY) *If* $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ *and* $\beta_0, \beta_1, \beta_2, \beta_3$ *are real numbers then* $(\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2)(\beta_0^2 + \beta_1^2 + \beta_2^2 + \beta_3^2) = (\alpha_0\beta_0 - \alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3)^2 + (\alpha_0\beta_1 + \alpha_1\beta_0 + \alpha_2\beta_3 - \alpha_3\beta_2)^2 + (\alpha_0\beta_2 - \alpha_1\beta_3 + \alpha_2\beta_0 + \alpha_3\beta_1)^2 + (\alpha_0\beta_3 + \alpha_1\beta_2 - \alpha_2\beta_1 + \alpha_3\beta_0)^2$.

Proof. Of course there is one obvious proof of this result, namely, multiply everything out and compare terms.

However, an easier way both to reconstruct the result at will and, at the same time, to prove it, is to notice that the left-hand side is $N(x)N(y)$ while the right-hand side is $N(xy)$ where $x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ and $y = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 k$. By Lemma 7.4.2, $N(x)N(y) = N(xy)$, ergo the Lagrange identity.

The Lagrange identity says that the sum of four squares times the sum of four squares is again, in a very specific way, the sum of four squares. A very striking result of Adolf Hurwitz says that if the sum of n squares times the sum of n squares is again a sum of n squares, where this last sum has terms computed bilinearly from the other two sums, then $n = 1, 2, 4$, or 8. There is, in fact, an identity for the product of sums of eight squares but it is too long and cumbersome to write down here.

Now is the appropriate time to introduce the Hurwitz ring of integral quaternions. Let $\zeta = \frac{1}{2}(1 + i + j + k)$ and let

$$H = \{m_0\zeta + m_1 i + m_2 j + m_3 k \mid m_0, m_1, m_2, m_3 \text{ integers}\}.$$

LEMMA 7.4.4 *H is a subring of Q. If* $x \in H$ *then* $x^* \in H$ *and* $N(x)$ *is a positive* integer *for every nonzero x in H.*

We leave the proof of Lemma 7.4.4 to the reader. It should offer no difficulties.

In some ways H might appear to be a rather contrived ring. Why use the quaternions ζ? Why not merely consider the more natural ring $Q_0 = \{m_0 + m_1 i + m_2 j + m_3 k \mid m_0, m_1, m_2, m_3 \text{ are integers}\}$? The answer is that Q_0 is not large enough, whereas H is, for the key lemma which follows to hold in it. But we want this next lemma to be true in the ring at our disposal for it allows us to characterize its left-ideals. This, perhaps, indicates why we (or rather Hurwitz) chose to work in H rather than in Q_0.

LEMMA 7.4.5 (LEFT-DIVISION ALGORITHM) *Let a and b be in H with* $b \neq 0$. *Then there exist two elements c and d in H such that* $a = cb + d$ *and* $N(d) < N(b)$.

Proof. Before proving the lemma, let's see what it tells us. If we look back in the section in Chapter 3 which deals with Euclidean rings, we can see that Lemma 7.4.5 assures us that except for its lack of commutativity H has all the properties of a Euclidean ring. The fact that elements in H may fail to commute will not bother us. True, we must be a little careful not to jump to erroneous conclusions; for instance $a = cb + d$ but we have no right to assume that a is also equal to $bc + d$, for b and c might not commute. But this will not influence any argument that we shall use.

In order to prove the lemma we first do so for a very special case, namely, that one in which a is an arbitrary element of H but b is a positive integer n. Suppose that $a = t_0\zeta + t_1 i + t_2 j + t_3 k$ where t_0, t_1, t_2, t_3 are integers and that $b = n$ where n is a positive integer. Let $c = x_0\zeta + x_1 i + x_2 j + x_3 k$ where x_0, x_1, x_2, x_3 are integers yet to be determined. We want to choose them in such a manner as to force $N(a - cn) < N(n) = n^2$. But

$$\begin{aligned} a - cn &= \left(t_0\left(\frac{1 + i + j + k}{2}\right) + t_1 i + t_2 j + t_3 k\right) \\ &\quad - nx_0\left(\frac{1 + i + j + k}{2}\right) - nx_1 i - nx_2 j - nx_3 k \\ &= \tfrac{1}{2}(t_0 - nx_0) + \tfrac{1}{2}(t_0 + 2t_1 - n(t_0 + 2x_1))i \\ &\quad + \tfrac{1}{2}(t_0 + 2t_1 - n(t_0 + 2x_2))j + \tfrac{1}{2}(t_0 + 2t_3 - n(t_0 + 2x_3))k. \end{aligned}$$

If we could choose the integers x_0, x_1, x_2, x_3 in such a way as to make $|t_0 - nx_0| \leq \frac{1}{2}n$, $|t_0 + 2t_1 - n(t_0 + 2x_1)| \leq n$, $|t_0 + 2t_2 - n(t_0 + 2x_2)| \leq n$ and $|t_0 + 2t_3 - n(t_0 + 2x_3)| \leq n$ then we would have

$$\begin{aligned} N(a - cn) &= \frac{(t_0 - nx_0)^2}{4} + \frac{(t_0 + 2t_1 - n(t_0 + 2x_1))^2}{4} + \cdots \\ &\leq \tfrac{1}{16}n^2 + \tfrac{1}{4}n^2 + \tfrac{1}{4}n^2 + \tfrac{1}{4}n^2 < n^2 = N(n), \end{aligned}$$

which is the desired result. But now we claim this can always be done:

1. There is an integer x_0 such that $t_0 = x_0 n + r$ where $-\frac{1}{2}n \leq r \leq \frac{1}{2}n$; for this x_0, $|t_0 - x_0 n| = |r| \leq \frac{1}{2}n$.
2. There is an integer k such that $t_0 + 2t_1 = kn + r$ and $0 \leq r \leq n$. If $k - t_0$ is even, put $2x_1 = k - t_0$; then $t_0 + 2t_1 = (2x_1 + t_0)n + r$ and $|t_0 + 2t_1 - (2x_1 + t_0)n| = r < n$. If, on the other hand, $k - t_0$ is odd, put $2x_1 = k - t_0 + 1$; thus $t_0 + 2t_1 = (2x_1 + t_0 - 1)n + r = (2x_1 + t_0)n + r - n$, whence $|t_0 + 2t_1 - (2x_1 + t_0)n| = |r - n| \leq n$ since $0 \leq r < n$. Therefore we can find an integer x_1 satisfying $|t_0 + 2t_1 - (2x_1 + t_0)n| \leq n$.
3. As in part 2, we can find integers x_2 and x_3 which satisfy $|t_0 + 2t_2 - (2x_2 + t_0)n| \leq n$ and $|t_0 + 2t_3 - (2x_3 + t_0)n| \leq n$, respectively.

In the special case in which a is an arbitrary element of H and b is a positive integer we have now shown the lemma to be true.

We go to the general case wherein a and b are arbitrary elements of H and $b \neq 0$. By Lemma 7.4.4, $n = bb^*$ is a positive integer; thus there exists a $c \in H$ such that $ab^* = cn + d_1$ where $N(d_1) < N(n)$. Thus $N(ab^* - cn) < N(n)$; but $n = bb^*$ whence we get $N(ab^* - cbb^*) < N(n)$, and so $N((a - cb)b^*) < N(n) = N(bb^*)$. By Lemma 7.4.2 this reduces to $N(a - cb)N(b^*) < N(b)N(b^*)$; since $N(b^*) > 0$ we get $N(a - cb) < N(b)$. Putting $d = a - cb$ we have $a = cb + d$ where $N(d) < N(b)$. This completely proves the lemma.

As in the commutative case we are able to deduce from Lemma 7.4.5

LEMMA 7.4.6 *Let L be a left-ideal of H. Then there exists an element $u \in L$ such that every element in L is a left-multiple of u; in other words, there exists $u \in L$ such that every $x \in L$ is of the form $x = ru$ where $r \in H$.*

Proof. If $L = (0)$ there is nothing to prove, merely put $u = 0$.

Therefore we may assume that L has nonzero elements. The norms of the nonzero elements are positive integers (Lemma 7.4.4) whence there is an element $u \neq 0$ in L whose norm is minimal over the nonzero elements of L. If $x \in L$, by Lemma 7.4.5, $x = cu + d$ where $N(d) < N(u)$. However d is in L because both x and u, and so cu, are in L which is a left-ideal. Thus $N(d) = 0$ and so $d = 0$. From this $x = cu$ is a consequence.

Before we can prove the four-square theorem, which is the goal of this section, we need one more lemma, namely

LEMMA 7.4.7 *If $a \in H$ then $a^{-1} \in H$ if and only if $N(a) = 1$.*

Proof. If both a and a^{-1} are in H, then by Lemma 7.4.4 both $N(a)$ and $N(a^{-1})$ are positive integers. However, $aa^{-1} = 1$, hence, by Lemma 7.4.2, $N(a)N(a^{-1}) = N(aa^{-1}) = N(1) = 1$. This forces $N(a) = 1$.

On the other hand, if $a \in H$ and $N(a) = 1$, then $aa^* = N(a) = 1$ and so $a^{-1} = a^*$. But, by Lemma 7.4.4, since $a \in H$ we have that $a^* \in H$, and so $a^{-1} = a^*$ is also in H.

We now have determined enough of the structure of H to use it effectively to study properties of the integers. We prove the famous classical theorem of Lagrange,

THEOREM 7.4.1 *Every positive integer can be expressed as the sum of squares of four integers.*

Proof. Given a positive integer n we claim in the theorem that $n = x_0^2 + x_1^2 + x_2^2 + x_3^2$ for four integers x_0, x_1, x_2, x_3. Since every integer factors into a product of prime numbers, if every prime number were

realizable as a sum of four squares, in view of Lagrange's identity (Lemma 7.4.3) every integer would be expressible as a sum of four squares. We have reduced the problem to consider only prime numbers n. Certainly the prime number 2 can be written as $1^2 + 1^2 + 0^2 + 0^2$ as a sum of four squares.

Thus, without loss of generality, we may assume that n is an *odd prime number*. As is customary we denote it by p.

Consider the quaternions W_p over J_p, the integers mod p; $W_p = \{\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k \mid \alpha_0, \alpha_1, \alpha_2, \alpha_3 \in J_p\}$. W_p is a finite ring; moreover, since $p \neq 2$ it is not commutative for $ij = -ji \neq ji$. Thus, by Wedderburn's theorem it cannot be a division ring, hence by Problem 1 at the end of Section 3.5, it must have a left-ideal which is neither (0) nor W_p.

But then the two-sided ideal V in H defined by $V = \{x_0\zeta + x_1 i + x_2 j + x_3 k \mid p \text{ divides all of } x_0, x_1, x_2, x_3\}$ cannot be a maximal left-ideal of H, since H/V is isomorphic to W_p. (Prove!) (If V were a maximal left-ideal in H, H/V, and so W_p, would have no left-ideals other than (0) and H/V).

Thus there is a left-ideal L of H satisfying: $L \neq H$, $L \neq V$, and $L \supset V$. By Lemma 7.4.6, there is an element $u \in L$ such that every element in L is a left-multiple of u. Since $p \in V$, $p \in L$, whence $p = cu$ for some $c \in H$. Since $u \notin V$, c cannot have an inverse in H, otherwise $u = c^{-1}p$ would be in V. Thus $N(c) > 1$ by Lemma 7.4.7. Since $L \neq H$, u cannot have an inverse in H, whence $N(u) > 1$. Since $p = cu$, $p^2 = N(p) = N(cu) = N(c)N(u)$. But $N(c)$ and $N(u)$ are integers, since both c and u are in H, both are larger than 1 and both divide p^2. The only way this is possible is that $N(c) = N(u) = p$.

Since $u \in H$, $u = m_0\zeta + m_1 i + m_2 j + m_3 k$ where m_0, m_1, m_2, m_3 are integers; thus $2u = 2m_0\zeta + 2m_1 i + 2m_2 j + 2m_3 k = (m_0 + m_0 i + m_0 j + m_0 k) + 2m_1 i + 2m_2 j + 2m_3 k = m_0 + (2m_1 + m_0)i + (2m_2 + m_0)j + (2m_3 + m_0)k$. Therefore $N(2u) = m_0^2 + (2m_1 + m_0)^2 + (2m_2 + m_0)^2 + (2m_3 + m_0)^2$. But $N(2u) = N(2)N(u) = 4p$ since $N(2) = 4$ and $N(u) = p$. We have shown that $4p = m_0^2 + (2m_1 + m_0)^2 + (2m_2 + m_0)^2 + (2m_3 + m_0)^2$. We are almost done.

To finish the proof we introduce an old trick of Euler's: If $2a = x_0^2 + x_1^2 + x_2^2 + x_3^2$ where a, x_0, x_1, x_2 and x_3 are integers, then $a = y_0^2 + y_1^2 + y_2^2 + y_3^2$ for some integers y_0, y_1, y_2, y_3. To see this note that, since $2a$ is even, the x's are all even, all odd or two are even and two are odd. At any rate in all three cases we can renumber the x's and pair them in such a way that

$$y_0 = \frac{x_0 + x_1}{2}, \quad y_1 = \frac{x_0 - x_1}{2}, \quad y_2 = \frac{x_2 + x_3}{2}, \quad \text{and} \quad y_3 = \frac{x_2 - x_3}{2}$$

are all integers. But

$$\begin{aligned} y_0^2 + y_1^2 + y_2^2 + y_3^2 &= \left(\frac{x_0 + x_1}{2}\right)^2 + \left(\frac{x_0 - x_1}{2}\right)^2 + \left(\frac{x_2 + x_3}{2}\right)^2 + \left(\frac{x_2 - x_3}{2}\right)^2 \\ &= \tfrac{1}{2}(x_0^2 + x_1^2 + x_2^2 + x_3^2) \\ &= \tfrac{1}{2}(2a) \\ &= a. \end{aligned}$$

Since $4p$ is a sum of four squares, by the remark just made $2p$ also is; since $2p$ is a sum of four squares, p also must be such a sum. Thus $p = a_0^2 + a_1^2 + a_2^2 + a_3^2$ for some integers a_0, a_1, a_2, a_3 and Lagrange's theorem is established.

This theorem itself is the starting point of a large research area in number theory, the so-called *Waring problem.* This asks if every integer can be written as a sum of a fixed number of kth powers. For instance it can be shown that every integer is a sum of nine cubes, nineteen fourth powers, etc. The Waring problem was shown to have an affirmative answer, in this century, by the great mathematician Hilbert.

Problems

1. Prove Lemma 7.4.4.
2. Find all the elements a in Q_0 such that a^{-1} is also in Q_0.
3. Prove that there are exactly 24 elements a in H such that a^{-1} is also in H. Determine all of them.
4. Give an example of an a and b, $b \neq 0$, in Q_0 such that it is impossible to find c and d in Q_0 satisfying $a = cb + d$ where $N(d) < N(b)$.
5. Prove that if $a \in H$ then there exist integers α, β such that $a^2 + \alpha a + \beta = 0$.
6. Prove that there is a positive integer which cannot be written as the sum of three squares.

*7. Exhibit an infinite number of positive integers which cannot be written as the sum of three squares.

Supplementary Reading

For a deeper discussion of finite fields: ALBERT, A. A., *Fundamental Concepts of Higher Algebra.* Chicago: University of Chicago Press, 1956.

For many proofs of the four-square theorem and a discussion of the Waring problem: HARDY, G. H., and WRIGHT, E. M., *An Introduction to the Theory of Numbers*, 4th ed. New York: Oxford University Press, 1960.

For another proof of the Wedderburn theorem: ARTIN, E., "Über einen Satz von Herrn J. H. M. Wedderburn," *Abhandlungen, Hamburg Mathematisches Seminar*, Vol. 5 (1928), pages 245–50.