VIII. Commutative Rings and Their Modules, 370-451

DOI: 10.3792/euclid/9781429799980-8

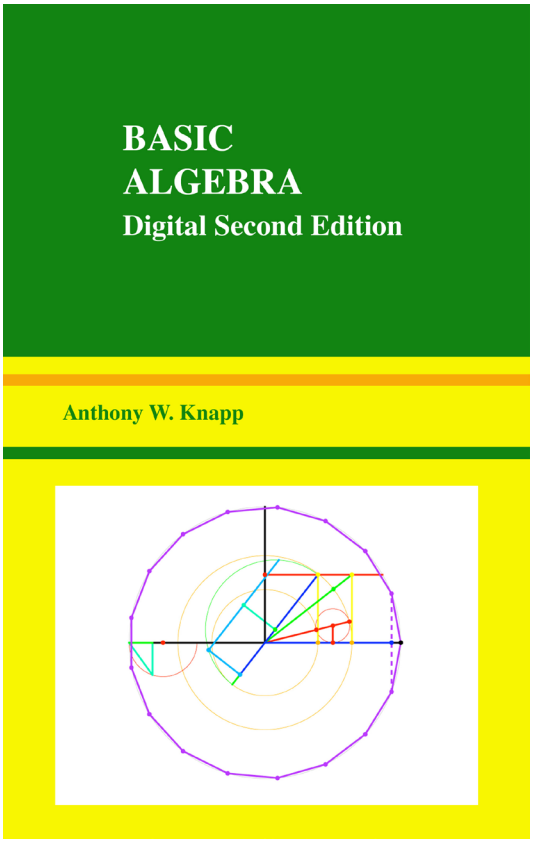
from

Basic Algebra

Digital Second Edition

Anthony W. Knapp

Full Book DOI: 10.3792/euclid/9781429799980
ISBN: 978-1-4297-9998-0



Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, N.Y. 11733–1729, U.S.A.
Email to: aknapp@math.stonybrook.edu
Homepage: www.math.stonybrook.edu/~aknapp

Title: Basic Algebra
Cover: Construction of a regular heptadecagon, the steps shown in color sequence; see page 505.

Mathematics Subject Classification (2010): 15–01, 20–01, 13–01, 12–01, 16–01, 08–01, 18A05, 68P30.

First Edition, ISBN-13 978-0-8176-3248-9

Published by Birkhäuser Boston

Digital Second Edition, not to be sold, no ISBN

Published by the Author

CHAPTER VIII

Commutative Rings and Their Modules

Abstract. This chapter amplifies the theory of commutative rings that was begun in Chapter IV, and it introduces modules for any ring. Emphasis is on the topic of unique factorization.

Section 1 gives many examples of rings, some commutative and some noncommutative, and introduces the notion of a module for a ring.

Sections 2–4 discuss some of the tools related to questions of factorization in integral domains. Section 2 defines the field of fractions for an integral domain and gives its universal mapping property. Section 3 defines prime and maximal ideals and relates quotients of them to integral domains and fields. Section 4 introduces principal ideal domains, which are shown to have unique factorization, and it defines Euclidean domains as a special kind of principal ideal domain for which greatest common divisors can be obtained constructively.

Section 5 proves that if R is an integral domain with unique factorization, then so is the polynomial ring $R[X]$. This result is a consequence of Gauss's Lemma, which addresses what happens to the greatest common divisor of the coefficients when one multiplies two members of $R[X]$. Gauss's Lemma has several other consequences that relate factorization in $R[X]$ to factorization in $F[X]$, where F is the field of fractions of R. Still another consequence is Eisenstein's irreducibility criterion, which gives a sufficient condition for a member of $R[X]$ to be irreducible.

Section 6 contains the theorem that every finitely generated unital module over a principal ideal domain is a direct sum of cyclic modules. The cyclic modules may be assumed to be primary in a suitable sense, and then the isomorphism types of the modules appearing in the direct-sum decomposition, together with their multiplicities, are uniquely determined. The main results transparently generalize the Fundamental Theorem for Finitely Generated Abelian Groups, and less transparently they generalize the existence and uniqueness of Jordan canonical form for square matrices with entries in an algebraically closed field.

Sections 7–11 contain foundational material related to factorization for the two subjects of algebraic number theory and algebraic geometry. Both these subjects rely heavily on the theory of commutative rings. Section 7 is a section of motivation, showing the analogy between a situation in algebraic number theory and a situation in algebraic geometry. Sections 8–10 introduce Noetherian rings, integral closures, and localizations. Section 11 uses this material to establish unique factorization of ideals for Dedekind domains, as well as some other properties.

1. Examples of Rings and Modules

Sections 4–5 of Chapter IV introduced rings and fields, giving a small number of examples of each. In the present section we begin by recalling those examples and giving further ones. Although Chapters VI and VII are not prerequisite for

the present chapter, our list of examples will include some rings and fields that arose in those two chapters.

The theory to be developed in this chapter is intended to apply to commutative rings, especially to questions related to unique factorization in such rings. Despite this limitation it seems wise to include examples of noncommutative rings in the list below.

In the conventions of this book, a ring need not have an identity. Many rings that arise only in the subject of algebra have an identity, but there are important rings in the subject of real analysis that do not. From the point of view of category theory, one therefore distinguishes between the category of all rings, with ring homomorphisms as morphisms, and the category of all rings with identity, with ring homomorphisms carrying 1 to 1 as morphisms. In the latter case one may want to exclude the zero ring from being an object in the category under certain circumstances.

EXAMPLES OF RINGS.

(1) Basic commutative rings from Chapter IV. All of the structures $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}/m\mathbb{Z}$, and $2\mathbb{Z}$ are commutative rings. All but the last have an identity. Of these, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields, and so is $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ if p is a prime number. The others are not fields.

(2) Polynomial rings. Let R be a nonzero commutative ring with identity. In Section IV.5 we defined the commutative ring $R[X_1, \dots, X_n]$ of polynomials over R in n indeterminates. It has a universal mapping property with respect to substitution for the indeterminates and use of a homomorphism on the coefficients. Making substitutions from R itself and mapping the coefficients by the identity homomorphism, we are led to the ring of all functions $(r_1, \dots, r_n) \mapsto f(r_1, \dots, r_n)$ for $r_1, \dots, r_n$ in R and $f(X_1, \dots, X_n)$ in $R[X_1, \dots, X_n]$; this is called the ring of all polynomial functions in n variables on R. Polynomials may be considered also in infinitely many variables, but we did not treat this case in any detail.

(3) Matrix rings over commutative rings. Let R be a nonzero commutative ring with identity. The set $M_n(R)$ of all n-by-n matrices with entries in R is a ring under entry-by-entry addition and the usual definition of matrix multiplication: $(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}$. It has an identity, namely the identity matrix I with $I_{ij} = \delta_{ij}$. In this setting, Section V.2 introduced a theory of determinants, and it was proved that a matrix has a one-sided inverse if and only if it has a two-sided inverse, if and only if its determinant is a member of the group $R^\times$ of units in R, i.e., elements of R invertible under multiplication. The matrix ring $M_n(R)$ is always noncommutative if $n > 1$.

(4) Matrix rings over noncommutative rings. If R is any ring, we can still make the set $M_n(R)$ of all n-by-n matrices with entries in R into a ring. However, if

R has no identity, $M_n(R)$ will have no identity. The theory of determinants does not directly apply if R is noncommutative or if R fails to have an identity,[1] and as a consequence, questions about the invertibility of matrices are more subtle than with the previous example.

(5) Spaces of linear maps from a vector space into itself. Let V be a vector space over a field $\mathbb{K}$. The vector space $\mathrm{End}_{\mathbb{K}}(V) = \mathrm{Hom}_{\mathbb{K}}(V, V)$ of all $\mathbb{K}$ linear maps from V to itself is initially a vector space over $\mathbb{K}$. Composition provides a multiplication that makes $\mathrm{End}_{\mathbb{K}}(V)$ into a ring with identity. In fact, associativity of multiplication is automatic for any kind of function, and so is the distributive law $(L_1+L_2)L_3 = L_1L_3+L_2L_3$. The distributive law $L_1(L_2+L_3) = L_1L_2+L_1L_3$ follows from the fact that L_1 is linear. This ring is isomorphic as a ring to $M_n(\mathbb{K})$ if V is n-dimensional, an isomorphism being determined by specifying an ordered basis of V.

(6) Associative algebras over fields. These were defined in Section VI.7, knowledge of which is not being assumed now. Thus we repeat the definition. If $\mathbb{K}$ is a field, then an associative algebra over $\mathbb{K}$, or associative $\mathbb{K}$ algebra, is a ring A that is also a vector space over $\mathbb{K}$ such that the multiplication $A \times A \to A$ is $\mathbb{K}$-linear in each variable. The conditions of linearity concerning multiplication have two parts to them: an additive part saying that the usual distributive laws are valid and a scalar-multiplication part saying that

$$(ka)b = k(ab) = a(kb) \qquad \text{for all } k \text{ in } \mathbb{K} \text{ and } a, b \text{ in } A.$$

If A has an identity, the displayed condition says that all scalar multiples of the identity lie in the **center** of A, i.e., commute with every element of A. In Examples 2 and 3, when R is a field $\mathbb{K}$, the polynomial rings and matrix rings over $\mathbb{K}$ provide examples of associative algebras over $\mathbb{K}$; scalar multiplication is to be done in entry-by-entry fashion. Example 5 is an associative algebra as well. If $\mathbb{L}$ is any field such that $\mathbb{K}$ is a subfield, then $\mathbb{L}$ may be regarded as an associative algebra over $\mathbb{K}$. An interesting commutative associative algebra over $\mathbb{C}$ without identity is the algebra $C_{\mathrm{com}}(\mathbb{R})$ of all continuous complex-valued functions on $\mathbb{R}$ that vanish outside a bounded interval; the vector-space operations are the usual pointwise operations, and the operation of multiplication is given by **convolution**

$$(f * g)(x) = \int_{\mathbb{R}} f(x-y)g(y)\,dy.$$

Section VII.4 worked with an analog $C(G, \mathbb{C})$ of this algebra in the context that $\mathbb{R}$ is replaced by a finite group G.

[1] A limited theory of determinants applies in the noncommutative case, but it will not be helpful for our purposes.

(7) Division rings. A division ring is a nonzero ring with identity such that every element has a two-sided inverse under multiplication. A commutative division ring is just a field. The ring $\mathbb{H}$ of quaternions is the only explicit noncommutative division ring that we have encountered so far. It is an associative algebra over $\mathbb{R}$. More generally, if A is a division ring, then we can easily check that the center $\mathbb{K}$ of A is a field and that A is an associative algebra over $\mathbb{K}$.[2]

(8) Tensor, symmetric, and exterior algebras. If E is a vector space over a field $\mathbb{K}$, Chapter VI defined the tensor, symmetric, and exterior algebras of E over $\mathbb{K}$, as well as the polynomial algebra on E in the case that E is finite-dimensional. These are all associative algebras with identity. Symmetric algebras and polynomial algebras are commutative. None of these algebras will be discussed further in this chapter.

(9) A field of 4 elements. This was constructed in Section IV.4. Further finite fields beyond the field of 4 elements and the fields $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ with p prime will be constructed in Chapter IX.

(10) Algebraic number fields $\mathbb{Q}[\theta]$. These were discussed in Sections IV.1 and IV.4. In defining $\mathbb{Q}[\theta]$, we assume that θ is a complex number and that there exists an integer $n > 0$ such that the complex numbers $1, \theta, \theta^2, \dots, \theta^n$ are linearly dependent over $\mathbb{Q}$. The set $\mathbb{Q}[\theta]$ is defined to be the subset of $\mathbb{C}$ obtained by substitution of θ into all members of $\mathbb{Q}[X]$. It coincides with the linear span over $\mathbb{Q}$ of $1, \theta, \theta^2, \dots, \theta^{n-1}$. Proposition 4.1 shows that it is closed under the arithmetic operations, including passage to multiplicative inverses of nonzero elements, and it is therefore a subfield of $\mathbb{C}$. This example ties in with the notion of minimal polynomial in Chapter V because the members of $\mathbb{Q}[X]$ with θ as a root are all multiples of one nonzero such polynomial that exhibits the linear dependence. We return to this example occasionally later in this chapter, particularly in Sections 7–11, and then we treat it in more detail in Chapter IX.

(11) Algebraic integers in a number field $\mathbb{Q}[\theta]$. Algebraic integers were defined in Section VII.4 as the roots in $\mathbb{C}$ of monic polynomials in $\mathbb{Z}[X]$, and they were shown to form a commutative ring with identity. The set of algebraic integers in $\mathbb{Q}[\theta]$ is therefore a commutative ring with identity, and it plays somewhat the same role for $\mathbb{Q}[\theta]$ that $\mathbb{Z}$ plays for $\mathbb{Q}$. We discuss this example further in Sections 7–11.

(12) Integral group rings. If G is a group, then we can make the free abelian group $\mathbb{Z}G$ on the elements of G into a ring by defining multiplication to be $\big(\sum_i m_i g_i\big)\big(\sum_j n_j h_j\big) = \sum_{i,j} (m_i n_j)(g_i h_j)$ when the m_i and n_j are in $\mathbb{Z}$ and the g_i and h_j are in G. It is immediate that the result is a ring with identity, and $\mathbb{Z}G$

[2]Use of the term "division algebra" requires some care. Some mathematicians understand division algebras to be associative, and others do not. The real algebra $\mathbb{O}$ of octonions, as defined in Problems 52–56 at the end of Chapter VI, is not associative, but it does have division.

is called the **integral group ring** of G. The group G is embedded as a subgroup of the group $(\mathbb{Z}G)^\times$ of units of $\mathbb{Z}G$, each element of g being identified with a sum $\iota(g) = \sum m_i g_i$ in which the only nonzero term is $1g$. The ring $\mathbb{Z}G$ has the universal mapping property illustrated in Figure 8.1 and described as follows: whenever $\varphi : G \to R$ is a group homomorphism of G into the group $R^\times$ of units of a ring R, then there exists a unique ring homomorphism $\Phi : \mathbb{Z}G \to R$ such that $\Phi\,\iota = \varphi$. The existence of Φ as a homomorphism of additive groups follows from the universal mapping property of free abelian groups, and then one readily checks that Φ respects multiplication.[3]

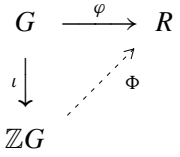


FIGURE 8.1. Universal mapping property of the integral group ring of G.

(13) Quotient rings. If R is a ring and I is a two-sided ideal, then we saw in Section IV.4 that the additive quotient R/I has a natural multiplication that makes it into a ring called a **quotient ring** of R. This in effect was the construction that obtained the ring $\mathbb{Z}/m\mathbb{Z}$ from the ring $\mathbb{Z}$.

(14) Direct product of rings. If $\{R_s \mid s \in S\}$ is a nonempty set of rings, then a **direct product** $\prod_{s\in S} R_s$ is a ring whose additive group is any direct product of the underlying additive groups and whose ring operations are given in entry-by-entry fashion. The resulting ring and the associated ring homomorphisms $p_{s_0} : \prod_{s\in S} R_s \to R_{s_0}$ amount to the product functor for the category of rings; if each R_s has an identity, the result amounts also to the product functor for the category of rings with identity.

We give further examples of rings near the end of this section after we have defined modules and given some examples.

Informally *a module is a vector space over a ring*. But let us be more precise. If R is a ring, then a **left R module**[4] M is an abelian group with the additional structure of a "scalar multiplication" $R \times M \to M$ such that

(i) $r(r'm) = (rr')m$ for r and r' in R and m in M,

[3]Universal mapping properties are discussed systematically in Problems 18–22 at the end of Chapter VI. The subject of such a property, here the pair $(\mathbb{Z}G, \iota)$, is always unique up to canonical isomorphism in a given category, but its existence has to be proved.

[4]Many algebra books write "R-module," using a hyphen. However, when R is replaced by an expression, particularly in applications of the theory, the hyphen is often dropped. For an example, see "module" in Hall's *The Theory of Groups*. The present book omits the hyphen in all cases in order to be consistent.

(ii) $(r + r')m = rm + r'm$ and $r(m + m') = rm + rm'$ if r and r' are in R and m and m' are in M.

In addition, if R has an identity, we say that M is **unital** if

(iii) $1m = m$ for all m in M.

One may also speak of **right R modules**. For these the scalar multiplication is usually written as mr with m in M and r in R, and the expected analogs of (i) and (ii) are to hold.

When R is commutative, it is immaterial which side is used for the scalar multiplication, and one speaks simply of an R **module**.

Let R be a ring, and let M and N be two left R modules. A **homomorphism of left R modules**, or more briefly an R **homomorphism**, is an additive group homomorphism $\varphi : M \to N$ such that $\varphi(rm) = r\varphi(m)$ for all r in R. Then we can form a category for fixed R in which the objects are the left R modules and the morphisms are the R homomorphisms from one left R module to another. Similarly the right R modules, along with the corresponding kind of R homomorphisms, form a category. If R has an identity, then the unital R modules form a subcategory in each case. These categories are fundamental to the subject of homological algebra, which we take up in Chapter IV of *Advanced Algebra*.

EXAMPLES OF MODULES.

(1) Vector spaces. If R is a field, *the unital R modules are exactly the vector spaces over R*.

(2) Abelian groups. *The unital $\mathbb{Z}$ modules are exactly the abelian groups*. Scalar multiplication is given in the expected way: If n is a positive integer, the product nx is the n-fold sum of x with itself. If $n = 0$, the product nx is 0. If $n < 0$, the product nx is $-((-n)x)$.

(3) Vector spaces as unital modules for the polynomial ring $\mathbb{K}[X]$. Let V be a finite-dimensional vector space over the field $\mathbb{K}$, and fix L be in $\mathrm{End}_{\mathbb{K}}(V)$. Then V becomes a unital $\mathbb{K}[X]$ module under the definition $A(X)v = A(L)(v)$ whenever $A(X)$ is a polynomial in $\mathbb{K}[X]$; here $A(L)$ is the member of $\mathrm{End}_{\mathbb{K}}(V)$ defined as in Section V.3. In Section 6 in this chapter we shall see that some of the deeper results in the theory of a single linear transformation, as developed in Chapter V, follow from the theory of unital $\mathbb{K}[X]$ modules that will emerge from the present chapter.

(4) Modules in the context of algebraic number fields. Let $\mathbb{Q}[\theta]$ be a subfield of $\mathbb{C}$ as in Example 10 of rings earlier in this section. It is assumed that the $\mathbb{Q}$ vector space $\mathbb{Q}[\theta]$ is finite-dimensional. Let L be the member of $\mathrm{End}_{\mathbb{Q}}(\mathbb{Q}[\theta])$ given as left multiplication by θ on $\mathbb{Q}[\theta]$. As in the previous example, $\mathbb{Q}[\theta]$ becomes a unital $\mathbb{Q}[X]$ module. Chapter V defines a minimal polynomial for

L, as well as a characteristic polynomial. These objects play a role in the study to be carried out in Chapter IX of fields like $\mathbb{Q}[\theta]$. If θ is an algebraic integer as in Example 11 of rings earlier in this section, then we can get more refined information by replacing $\mathbb{Q}$ by $\mathbb{Z}$ in the above analysis; this technique plays a role in the theory to be developed in Sections 7–11.

(5) Rings and their quotients. If R is a ring, then R is a left R module and also a right R module. If I is a two-sided ideal in I, then the quotient ring R/I, as defined in Proposition 4.20, is a left R module and also a right R module. These modules are automatically unital if R has an identity. Later in this section we shall consider quotients of R by "one-sided ideals."

(6) Spaces of rectangular matrices. If R is a ring, then the space $M_{mn}(R)$ of m-by-n matrices with entries in R is an abelian group under addition and becomes a left R module when multiplication by the scalar r is defined as left multiplication by r in each entry. Also, if we put $S = M_m(R)$, then $M_{mn}(R)$ is a left S module under the usual definition of matrix multiplication: $(sv)_{ij} = \sum_{k=1}^{n} s_{ik} v_{kj}$, where s is in S and v is in $M_{mn}(R)$.

(7) Direct product of R modules. If S is a nonempty set and $\{M_s\}_{s\in S}$ is a corresponding system of left R modules, then a **direct product** $\prod_{s\in S} M_s$ is obtained as an additive group by forming any direct product of the underlying additive groups of the M_s's and defining scalar multiplication by members of R to be scalar multiplication in each coordinate. The associated abelian-group homomorphisms $p_{s_0} : \prod_{s\in S} M_s \to M_{s_0}$ become R homomorphisms under this definition of scalar multiplication on the direct product. Direct product amounts to the product functor for the category of left R modules; we omit the easy verification, which makes use of the corresponding fact about abelian groups. As in the case of abelian groups, we can speak of an **external** direct product as the result of a construction that starts with the product of the sets M_s, and we can speak of recognizing a direct product as **internal** when the M_s's are contained in the direct product and the restriction of each p_s to M_s is the identity function.

(8) Direct sum of R modules. If S is a nonempty set and $\{M_s\}_{s\in S}$ is a corresponding system of left R modules, then a **direct sum** $\bigoplus_{s\in S} M_s$ is obtained as an additive group by forming any direct sum of the underlying additive groups of the M_s's and defining scalar multiplication by members of R to be scalar multiplication in each coordinate. The associated abelian-group homomorphisms $i_{s_0} : M_{s_0} \to \bigoplus_{s\in S} M_s$ become R homomorphisms under this definition of scalar multiplication on the direct sum. Direct sum amounts to the coproduct functor for the category of left R modules; we omit the easy verification, which makes use of the corresponding fact about abelian groups. As in the case of abelian groups, we can speak of an **external** direct sum as the result of a construction that starts with a subset of the product of the sets M_s, and we can speak of recognizing a

direct sum as **internal** when the M_s's are contained in the direct sum and each i_s is the inclusion mapping.

(9) Free R modules. Let R be a nonzero ring with identity, and let S be a nonempty set. As in Example 5, let us regard R as a unital left R module. Then the left R module given as the direct sum $F(S) = \bigoplus_{s \in S} R$ is called a **free** R **module**, or free left R module. We define $\iota : S \to F(S)$ by $\iota(s) = i_s(1)$, where i_s is the usual embedding map for the direct sum of R modules. The left R module $F(S)$ has a universal mapping property similar to the corresponding property of free abelian groups. This is illustrated in Figure 8.2 and is described as follows: whenever M is a unital left R module and $\varphi : S \to M$ is a function, then there exists a unique R homomorphism $\Phi : F(S) \to M$ such that $\Phi \iota = \varphi$. The existence of Φ as an R homomorphism follows from the universal mapping property of direct sums (Example 8) as soon as the property is demonstrated for S equal to a singleton set. Thus let A be any left R module, and let $a \in A$ be given; then it is evident that $r \mapsto ra$ is the unique R homomorphism of the left R module R into A carrying 1 to a.

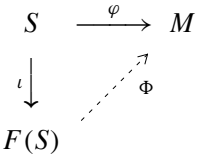


FIGURE 8.2. Universal mapping property of a free left R module.

If R is a ring and M is a left R module, then an R **submodule** N of M is an additive subgroup of M that is closed under scalar multiplication, i.e., has rm in N when r is in R and m is in N. In situations in which there is no ambiguity, the use of "left" in connection with R submodules is not necessary.

EXAMPLES OF SUBMODULES. If V is a vector space over a field $\mathbb{K}$, then a $\mathbb{K}$ submodule of V is a vector subspace of V. If M is an abelian group, then a $\mathbb{Z}$ submodule of M is a subgroup. In Example 6 of modules, in which $S = M_m(R)$, then an example of a left S submodule of $M_{mn}(R)$ is all matrices with 0 in every entry of a specified subset of the n columns.

If the ring R has an identity and M is a unital left R module, then the R submodule of M **generated** by $m \in M$, i.e., the smallest R submodule containing m, is Rm, the set of products rm with r in R. In fact, the set of all rm is an abelian group since $(r \pm s)m = rm \pm sm$, it is closed under scalar multiplication since $s(rm) = (sr)m$, and it contains m since $1m = m$. However, if the left R module M is not unital, then the R submodule generated by m may not equal Rm, and it was for that reason that R modules were assumed to be unital in the construction of free R modules in Example 9 of modules above. More generally the R submodule

of M **generated** by a finite set $\{m_1, \dots, m_n\}$ in M is $Rm_1 + \cdots + Rm_n$ if the left R module M is unital.

Example 5 of modules treated R as a left R module. In this setting the left R submodules are called **left ideals** in R. That is, a left ideal I is an additive subgroup of R such that ri is in I whenever r is in R and i is in I. As a special case of what was said in the previous paragraph, if the ring R has an identity, then the left R module R is automatically unital, and the left ideal of R generated by an element a is Ra, the set of all products ra with r in R.

Similarly a **right ideal** in R is an additive subgroup I such that ir is in I whenever r is in R and i is in I. The right ideals are the right R submodules of the right R module R. If R is commutative, then left ideals, right ideals, and two-sided ideals are all the same.

Suppose that $\varphi : M \to N$ is an R homomorphism of left R modules. In this situation we readily verify that the kernel of φ, denoted by $\ker\varphi$ as usual, is an R submodule of M, and the image of φ, denoted by $\operatorname{image}\varphi$ as usual, is an R submodule of N. The R homomorphism φ is one-one if and only if $\ker\varphi = 0$, as a consequence of properties of homomorphisms of abelian groups. A one-one R homomorphism of one left R module onto another is called an R **isomorphism**; its inverse is automatically an R isomorphism, and "is R isomorphic to" is an equivalence relation.

Still with R as a ring, suppose that M is a left R module and N is an R submodule. Then we can form the quotient M/N of abelian groups. This becomes a left R module under the definition $r(m + N) = rm + N$, as we readily check. We call M/N a **quotient module**. The quotient mapping $m \mapsto m + N$ of M to M/N is an R homomorphism onto. A particular example of a quotient module is R/I, where I is a left ideal in R.

We can now go over the results on quotients of abelian groups in Section IV.2, specifically Proposition 4.11 through Theorem 4.14, and check that they extend immediately to results about left R modules. The statements appear below. The arguments are all routine, and there is no point in repeating them. In the special case that R is a field and the R modules are vector spaces, these results specialize to results proved in Sections II.5 and II.6.

Proposition 8.1. Let R be a ring, let $\varphi : M_1 \to M_2$ be an R homomorphism between left R modules, let $N_0 = \ker\varphi$, let N be an R submodule of M_1 contained in N_0, and define $q : M_1 \to M_1/N$ to be the R module quotient map. Then there exists an R homomorphism $\overline{\varphi} : M_1/N \to M_2$ such that $\varphi = \overline{\varphi}q$, i.e, $\overline{\varphi}(m_1 + N) = \varphi(m_1)$. It has the same image as φ, and $\ker\overline{\varphi} = \{h_0N \mid h_0 \in N_0\}$.

REMARK. As with groups, one says that φ **factors through** M_1/N or **descends to** M_1/N. Figure 8.3 illustrates matters.

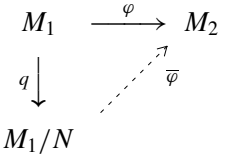


FIGURE 8.3. Factorization of R homomorphisms via a quotient of R modules.

Corollary 8.2. Let R be a ring, let $\varphi : M_1 \to M_2$ be an R homomorphism between left R modules, and suppose that φ is onto M_2 and has kernel N. Then φ exhibits the left R module M_1/N as canonically R isomorphic to M_2.

Theorem 8.3 (First Isomorphism Theorem). Let R be a ring, let $\varphi : M_1 \to M_2$ be an R homomorphism between left R modules, and suppose that φ is onto M_2 and has kernel K. Then the map $N_1 \mapsto \varphi(N_1)$ gives a one-one correspondence between

(a) the R submodules N_1 of M_1 containing K and
(b) the R submodules of M_2.

Under this correspondence the mapping $m + N_1 \mapsto \varphi(m) + \varphi(N_1)$ is an R isomorphism of M_1/N_1 onto $M_2/\varphi(N_1)$.

REMARK. In the special case of the last statement that $\varphi : M_1 \to M_2$ is an R module quotient map $q : M \to M/K$ and N is an R submodule of M containing K, the last statement of the theorem asserts the R isomorphism $M/N \cong (M/K)\big/(N/K)$.

Theorem 8.4 (Second Isomorphism Theorem). Let R be a ring, let M be a left R module, and let N_1 and N_2 be R submodules of M. Then $N_1 \cap N_2$ is an R submodule of N_1, the set $N_1 + N_2$ of sums is an R submodule of M, and the map $n_1 + (N_1 \cap N_2) \mapsto n_1 + N_2$ is a well-defined canonical R isomorphism

$$N_1/(N_1 \cap N_2) \cong (N_1 + N_2)/N_2.$$

A quotient of a direct sum of R modules by the direct sum of R submodules is the direct sum of the quotients, according to the following proposition. The result generalizes Lemma 4.58, which treats the special case of abelian groups (unital $\mathbb{Z}$ modules).

Proposition 8.5. Let R be a ring, let $M = \bigoplus_{s\in S} M_s$ be a direct sum of left R modules, and for each s in S, let N_s be a left R submodule of M_s. Then the natural map of $\bigoplus_{s\in S} M_s$ to the direct sum of quotients descends to an R isomorphism

$$\bigoplus_{s\in S} M_s \Big/ \bigoplus_{s\in S} N_s \cong \bigoplus_{s\in S} (M_s/N_s).$$

PROOF. Let $\varphi : \bigoplus_{s\in S} M_s \to \bigoplus_{s\in S} (M_s/N_s)$ be the R homomorphism defined by $\varphi(\{m_s\}_{s\in S}) = \{m_s + N_s\}_{s\in S}$. The mapping φ is onto $\bigoplus_{s\in S} (M_s/N_s)$, and the kernel is $\bigoplus_{s\in S} N_s$. Then Corollary 8.2 shows that φ descends to the required R isomorphism. □

EXAMPLES OF RINGS, CONTINUED.

(15) Associative algebras over commutative rings with identity. These directly generalize Example 6 of rings. Let R be a nonzero commutative ring with identity. An **associative algebra over** R, or **associative** R **algebra**, is a ring A that is also a left R module such that multiplication $A \times A \to A$ is R linear in each variable. The conditions of R linearity in each variable mean that addition satisfies the usual distributive laws for a ring and that the following condition is to be satisfied relating multiplication and scalar multiplication:

$$(ra)b = r(ab) = a(rb) \qquad \text{for all } r \text{ in } R \text{ and } a, b \in A.$$

If A has an identity, the displayed condition says that all scalar multiples of the identity lie in the **center** of A, i.e., commute with every element of A. Examples 2 and 3, treating polynomial rings and matrix rings whose scalars lie in a commutative ring with identity, furnish examples. Every ring R is an associative $\mathbb{Z}$ algebra when the $\mathbb{Z}$ action is defined so as to make the abelian group underlying the additive structure of R into a $\mathbb{Z}$ module. All that needs to be checked is the displayed formula. For $n = 1$, we have $(1a)b = 1(ab) = a(1b)$ since the $\mathbb{Z}$ module R is unital. If we also have $(na)b = n(ab) = a(nb)$ for a positive integer n, then we can add and use the appropriate distributive laws to obtain $((n + 1)a)b = (n + 1)(ab) = a((n + 1)b)$. Induction therefore gives $(na)b = n(ab) = a(nb)$ for all positive integers n, and this equality extends to all integers n by using additive inverses. The associative R algebras form a category in which the morphisms from one such algebra to another are the ring homomorphisms that are also R homomorphisms. The product functor for this category is the direct product as in Example 14 with an overlay of scalar multiplication as in Example 7 of modules. The coproduct functor in the category of commutative associative R algebras with identity is more subtle and involves a tensor product over R, a notion we postpone introducing until Chapter X.

(16) Group algebra RG over R. If G is a group and R is a commutative ring with identity, then we can introduce a multiplication in the free R module RG on the elements of G by the definition $\big(\sum_i r_i g_i\big)\big(\sum_j s_j h_j\big) = \sum_{i,j} (r_i s_j)(g_i h_j)$ when the r_i and s_j are in R and the g_i and h_j are in G. It is immediate that this multiplication makes the free R module into an associative R algebra with identity, and RG is called the **group algebra** of G over R. The special case $R = \mathbb{Z}$ leads to the integral group ring as in Example 12. The group G is embedded as a

subgroup of the group $(RG)^\times$ of units of RG, each element of g being identified with a sum $\iota(g) = \sum r_i g_i$ in which the only nonzero term is $1g$. The associative R algebra RG has a universal mapping property similar to that in Figure 8.1 and given in Figure 8.4 as follows: whenever $\varphi : G \to A$ is a group homomorphism of G into the group $A^\times$ of units of an associative R algebra A, then there exists a unique associative R algebra homomorphism $\Phi : RG \to A$ such that $\Phi\,\iota = \varphi$.

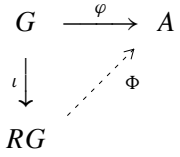


FIGURE 8.4. Universal mapping property of the group algebra RG.

(17) Scalar-valued functions of finite support on a group, with convolution as multiplication. If G is a group and R is a commutative ring with identity, denote by $C(G, R)$ the R module of all functions from G into R that are of **finite support** in the sense that each function is 0 except on a finite subset of G. This R module readily becomes an associative R algebra if ring multiplication is taken to be pointwise multiplication, but the interest here is in a different definition of multiplication. Instead, multiplication is defined to be **convolution** with

$$(f_1 * f_2)(x) = \sum_{y \in G} f_1(xy^{-1}) f_2(y) = \sum_{y \in G} f_1(y) f_2(y^{-1}x).$$

The sums in question are finite because of the finite support of f_1 and f_2, and the sums are equal by a change of variables. This multiplication was introduced in the special case $R = \mathbb{C}$ in Section VII.4, and the argument for associativity given there in the special case works in general. With convolution as multiplication, $C(G, R)$ becomes an associative R algebra with identity. Problem 14 at the end of the chapter asks for a verification that the mapping $g \mapsto f_g$ with

$$f_g(x) = \begin{cases} 1 & \text{for } x = g, \\ 0 & \text{for } x \neq g, \end{cases}$$

extends to an R algebra isomorphism of RG onto $C(G, R)$.

2. Integral Domains and Fields of Fractions

For the remainder of the chapter we work with commutative rings only. In several of the sections, including this one, the commutative ring will be an integral domain, i.e., a nonzero commutative ring with identity and with no zero divisors.

In this section we show how an integral domain can be embedded canonically in a field. This embedding is handy for recognizing certain facts about integral domains as consequences of facts about fields. For example Proposition 4.28b established that if R is a nonzero integral domain and if $A(X)$ is a polynomial in $R[X]$ of degree $n > 0$, then $A(X)$ has at most n roots. Since the coefficients of the polynomial can be considered to be members of the larger field that contains R, this result is an immediate consequence of the corresponding fact about fields (Corollary 1.14).

The prototype is the construction of the field $\mathbb{Q}$ of rationals from the integral domain $\mathbb{Z}$ of integers as in Section A3 of the appendix, in which one thinks of $\frac{a}{b}$ as a pair (a, b) with $b \neq 0$ and then identifies pairs by saying that $\frac{a}{b} = \frac{c}{d}$ if and only if $ad = bc$.

We proceed in the same way in the general case. Thus let R be an integral domain, form the set

$$\widetilde{F} = \{(a, b) \mid a \in R,\ b \in R,\ b \neq 0\},$$

and impose the equivalence relation $(a, b) \sim (c, d)$ if $ad = bc$. The relation $\sim$ is certainly reflexive and symmetric. To see that it is transitive, suppose that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. Then $ad = bc$ and $cf = de$, and these together force $adf = bcf = bde$. In turn, this implies $af = be$ since R is an integral domain and d is assumed $\neq 0$. Thus $\sim$ is transitive and is an equivalence relation. Let F be the set of equivalence classes.

The definition of addition in $\widetilde{F}$ is $(a, b)+(c, d) = (ad+bc, bd)$, the expression we get by naively clearing fractions, and we want to see that addition is consistent with the equivalence relation. In checking this, we need change only one of the pairs at a time. Thus suppose that $(a', b') \sim (a, b)$ and that (c, d) is given. We know that $a'b = ab'$, and we want to see that $(ad + bc, bd) \sim (a'd + b'c, b'd)$, i.e., that $(ad + bc)b'd = (a'd + b'c)bd$. In other words, we are to check that $adb'd = a'dbd$; we see immediately that this equality is valid since $ab' = a'b$. Consequently addition is consistent with the equivalence relation and descends to be defined on the set F of equivalence classes.

Taking into account the properties satisfied by members of an integral domain, we check directly that addition is commutative and associative on $\widetilde{F}$, and it follows that addition is commutative and associative on F.

The element $(0, 1)$ is a two-sided identity for addition in $\widetilde{F}$, and hence the class of $(0, 1)$ is a two-sided identity for addition in F. We denote this class by 0. Let us identify this class. A pair (a, b) is in the class of $(0, 1)$ if and only if $0 \cdot b = 1 \cdot a$, hence if and only if $a = 0$. In other words, the class of $(0, 1)$ consists of all $(0, b)$ with $b \neq 0$.

In $\widetilde{F}$, we have $(a, b) + (-a, b) = (ab + b(-a), bb) = (0, b^2) \sim (0, 1)$, and therefore the class of $(-a, b)$ is a two-sided inverse to the class of (a, b) under

addition. Consequently F is an abelian group under addition.

The definition of multiplication in $\widetilde{F}$ is $(a, b)(c, d) = (ac, bd)$, and it is routine to see that this definition is consistent with the equivalence relation. Therefore multiplication descends to be defined on F. We check by inspection that multiplication is commutative and associative on $\widetilde{F}$, and it follows that it is commutative and associative on F. The element $(1, 1)$ is a two-sided identity for multiplication in $\widetilde{F}$, and the class of $(1, 1)$ is therefore a two-sided identity for multiplication in F. We denote this class by 1.

If (a, b) is not in the class 0, then $a \neq 0$, as we saw above. Then $ab \neq 0$, and we have $(a, b)(b, a) = (ab, ab) \sim (1, 1) = 1$. Hence the class of (b, a) is a two-sided inverse of the class of (a, b) under multiplication. Consequently the nonzero elements of F form an abelian group under multiplication.

For one of the distributive laws, the computation

$$\begin{aligned}(a, b)((c, d) + (e, f)) &= (a, b)(cf + de, df) = (a(cf + de), bdf) \\ &= (acf + ade, bdf) \sim (acbf + bdae, b^2df) \\ &= (ac, bd) + (ae, bf) = (a, b)(c, d) + (a, b)(e, f)\end{aligned}$$

shows that the classes of $(a, b)((c, d) + (e, f))$ and of $(a, b)(c, d) + (a, b)(e, f)$ are equal. The other distributive law follows from this one since F is commutative under multiplication. Therefore F is a field.

The field F is called the **field of fractions** of the integral domain R. The function $\eta : R \to F$ defined by saying that $\eta(r)$ is the class of $(r, 1)$ is easily checked to be a homomorphism of rings sending 1 to 1. It is one-one. Let us call it the canonical embedding of R into F. The pair (F, η) has the universal mapping property stated in Proposition 8.6 and illustrated in Figure 8.5.

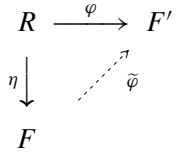


FIGURE 8.5. Universal mapping property of the field of fractions of R.

Proposition 8.6. Let R be an integral domain, let F be its field of fractions, and let η be the canonical embedding of R into F. Whenever φ is a one-one ring homomorphism of R into a field F' carrying 1 to 1, then there exists a unique ring homomorphism $\widetilde{\varphi} : F \to F'$ such that $\varphi = \widetilde{\varphi}\eta$, and $\widetilde{\varphi}$ is one-one as a homomorphism of fields.

REMARK. We say that $\widetilde{\varphi}$ is the extension of φ from R to F. Once this proposition has been proved, it is customary to drop η from the notation and regard R as a subring of its field of fractions.

PROOF. If (a, b) with $b \neq 0$ is a pair in $\widetilde{F}$, we define $\Phi(a, b) = \varphi(a)\varphi(b)^{-1}$. This is well defined since $b \neq 0$ and since φ, being one-one, cannot have $\varphi(b) = 0$. Let us see that Φ is consistent with the equivalence relation, i.e., that $(a, b) \sim (a', b')$ implies $\Phi(a, b) = \Phi(a', b')$. Since $(a, b) \sim (a', b')$, we have $ab' = a'b$ and therefore also $\varphi(a)\varphi(b') = \varphi(a')\varphi(b)$ and $\Phi(a, b) = \varphi(a)\varphi(b)^{-1} = \varphi(a')\varphi(b')^{-1} = \Phi(a', b')$, as required.

We can thus define $\widetilde{\varphi}$ of the class of (a, b) to be $\Phi(a, b)$, and $\widetilde{\varphi}$ is well defined as a function from F to F'. If r is in R, then $\widetilde{\varphi}(\eta(r)) = \widetilde{\varphi}(\text{class of } (r, 1)) = \Phi(r, 1) = \varphi(r)\varphi(1)^{-1}$, and this equals $\varphi(r)$ since φ is assumed to carry 1 into 1. Therefore $\widetilde{\varphi}\eta = \varphi$.

For uniqueness, let the class of (a, b) be given in F. Since b is nonzero, this class is the same as the class of $(a, 1)(b, 1)^{-1}$, which equals $\eta(a)\eta(b)^{-1}$. Since $(\widetilde{\varphi}\eta)(a) = \varphi(a)$ and $(\widetilde{\varphi}\eta)(b) = \varphi(b)$, we must have $\widetilde{\varphi}(\text{class of } (a, b)) = \widetilde{\varphi}(\eta(a))\widetilde{\varphi}(\eta(b))^{-1} = \varphi(a)\varphi(b)^{-1}$. Therefore φ uniquely determines $\widetilde{\varphi}$. □

If $\mathbb{K}$ is a field, then $R = \mathbb{K}[X]$ is an integral domain, and Proposition 8.6 applies to this R. The field of fractions consists in effect of formal rational expressions $P(X)Q(X)^{-1}$ in the indeterminate X, with the expected identifications made. We write $\mathbb{K}(X)$ for this field of fractions. More generally the field of fractions of the integral domain $\mathbb{K}[X_1, \ldots, X_n]$ consists of formal rational expressions in the indeterminates $X_1, \ldots, X_n$, with the expected identifications made, and is denoted by $\mathbb{K}(X_1, \ldots, X_n)$.

3. Prime and Maximal Ideals

In this section, R will denote a commutative ring, not necessarily having an identity. We shall introduce the notions of "prime ideal" and "maximal ideal," and we shall investigate relationships between these two notions.

A proper ideal I in R is **prime** if $ab \in I$ implies $a \in I$ or $b \in I$. The ideal $I = R$ is not prime, by convention.[5] We give three examples of prime ideals; a fourth example will be given in a proposition immediately afterward.

EXAMPLES.

(1) For $\mathbb{Z}$, it was shown in an example just before Proposition 4.21 that each ideal is of the form $m\mathbb{Z}$ for some integer m. We may assume that $m \geq 0$. The prime ideals are 0 and all $p\mathbb{Z}$ with p prime. To see this latter fact, consider $m\mathbb{Z}$ with $m \geq 2$. If $m = ab$ nontrivially, then neither a nor b is in I, but ab is in I; hence I is not prime. Conversely if m is prime, and if ab is in $I = m\mathbb{Z}$, then

[5]This convention is now standard. Books written before about 1960 usually regarded $I = R$ as a prime ideal. Correspondingly they usually treated the zero ring as an integral domain.

$ab = mc$ for some integer c. Since m is prime, Lemma 1.6 shows that m divides a or m divides b. Hence a is in I or b is in I. Therefore I is prime.

(2) If $\mathbb{K}$ is a field, then each ideal in $R = \mathbb{K}[X]$ is of the form $A(X)\mathbb{K}[X]$ with $A(X)$ in $\mathbb{K}[X]$, and $A(X)\mathbb{K}[X]$ is prime if and only if $A(X)$ is 0 or is a prime polynomial. In fact, each ideal is of the form $A(X)\mathbb{K}[X]$ by Proposition 5.8. If $A(X)$ is not a constant polynomial, then the argument that $A(X)\mathbb{K}[X]$ is prime if and only if the polynomial $A(X)$ is prime proceeds as in Example 1, using Lemma 1.16 in place of Lemma 1.6.

(3) In $R = \mathbb{Z}[X]$, the structure of the ideals is complicated, and we shall not attempt to list all ideals. Let us observe simply that the ideal $I = X\mathbb{Z}[X]$ is prime. In fact, if $A(X)B(X)$ is in $X\mathbb{Z}[X]$, then $A(X)B(X) = XC(X)$ for some $C(X)$ in $\mathbb{Z}[X]$. If the constant terms of $A(X)$ and $B(X)$ are a_0 and b_0, this equation says that $a_0b_0 = 0$. Therefore $a_0 = 0$ or $b_0 = 0$. In the first case, $A(X) = XP(X)$ for some $P(X)$, and then $A(X)$ is in I; in the second case, $B(X) = XQ(X)$ for some $Q(X)$, and then $B(X)$ is in I. We conclude that I is prime.

Proposition 8.7. An ideal I in the commutative ring R is prime if and only if R/I is an integral domain.

PROOF. If a proper ideal I fails to be prime, choose ab in I with $a \notin I$ and $b \notin I$. Then $a + I$ and $b + I$ are nonzero in R/I and have product $0 + I$. So R/I is nonzero and has a zero divisor; by definition, R/I fails to be an integral domain.

Conversely if R/I (is nonzero and) has a zero divisor, choose $a + I$ and $b + I$ nonzero with product $0 + I$. Then neither a nor b is in I but ab is in I. Since I is certainly proper, I is not prime. □

A proper ideal I in the commutative ring R is said to be **maximal** if R has no proper ideal J with $I \subsetneq J$. If the commutative ring R has an identity, a simple way of testing whether an ideal I is proper is to check whether 1 is in I; in fact, if 1 is in I, then $I \supseteq RI \supseteq R1 = R$ implies $I = R$. Maximal ideals exist in abundance when R is nonzero and has an identity, as a consequence of the following result.

Proposition 8.8. In a commutative ring R with identity, any proper ideal is contained in a maximal ideal.

PROOF. This follows from Zorn's Lemma (Section A5 of the appendix). Specifically let I be the given proper ideal, and form the set S of all proper ideals that contain I. This set is nonempty, containing I as a member, and we order it by inclusion upward. If we have a chain in S, then the union of the members of the chain is an ideal that contains all the ideals in the chain, and it is

proper since it does not contain 1. Therefore the union of the ideals in the chain is an upper bound for the chain. By Zorn's Lemma the set S has a maximal element, and any such maximal element is a maximal ideal containing I. □

Lemma 8.9. If R is a nonzero commutative ring with identity, then R is a field if and only if the only proper ideal in R is 0.

PROOF. If R is a field and I is a nonzero ideal in R, let $a \neq 0$ be in I. Then $1 = aa^{-1}$ is in I, and consequently $I = R$. Conversely if the only ideals in R are 0 and R, let $a \neq 0$ be given in R, and form the ideal $I = aR$. Since 1 is in R, a is in I. Thus $I \neq 0$. Then I must be R. So there exists some b in R with $1 = ba$, and a is exhibited as having the inverse b. □

Proposition 8.10. If R is a commutative ring with identity, then an ideal I is maximal if and only if R/I is a field.

REMARK. One can readily give a direct proof, but it seems instructive to give a proof reducing the result to Lemma 8.9.

PROOF. We consider R and R/I as unital R modules, the ideals for each of R and R/I being the R submodules. The quotient ring homomorphism $R \to R/I$ is an R homomorphism. By the First Isomorphism Theorem for modules (Theorem 8.3), there is a one-one correspondence between the ideals in R containing I and the ideals in R/I. Then the result follows immediately from Lemma 8.9. □

Corollary 8.11. If R is a commutative ring with identity, then every maximal ideal is prime.

PROOF. If I is maximal, then R/I is a field by Proposition 8.10. Hence R/I is an integral domain, and I must be prime by Proposition 8.7. □

In the converse direction nonzero prime ideals need not be maximal, as the following example shows. However, Proposition 8.12 will show that nonzero prime ideals *are* necessarily maximal in certain important rings.

EXAMPLE. In $R = \mathbb{Z}[X]$, we have seen that $I = X\mathbb{Z}[X]$ is a prime ideal. But I is not maximal since $X\mathbb{Z}[X] + 2\mathbb{Z}[X]$ is a proper ideal that strictly contains I.

Proposition 8.12. In $R = \mathbb{Z}$ or $R = \mathbb{K}[X]$ with $\mathbb{K}$ a field, every nonzero prime ideal is maximal.

PROOF. Examples 1 and 2 at the beginning of this section show that every nonzero prime ideal is of the form $I = pR$ with p prime. If such an I is given and if J is any ideal strictly containing I, choose a in J with a not in I. Since a

is not in $I = pR$, it is not true that p divides a. So p and a are relatively prime, and there exist elements x and y in R with $xp + ya = 1$, by Proposition 1.2c or 1.15d. Since p and a are in J, so is 1. Therefore $J = R$, and I is not strictly contained in any proper ideal. So I is maximal. □

EXAMPLE. Algebraic number fields $\mathbb{Q}[\theta]$. These were introduced briefly in Chapter IV and again in Section 1 as the $\mathbb{Q}$ linear span of all powers $1, \theta, \theta^2, \ldots$. Here θ is a nonzero complex number, and we make the assumption that $\mathbb{Q}[\theta]$ is a finite-dimensional vector space over $\mathbb{Q}$. Proposition 4.1 showed that $\mathbb{Q}[\theta]$ is then indeed a field. Let us see how this conclusion relates to the results of the present section. In fact, write a nontrivial linear dependence of $1, \theta, \theta^2, \ldots$ over $\mathbb{Q}$ in the form $c_0 + c_1\theta + c_2\theta^2 + \cdots + c_{n-1}\theta^{n-1} + \theta^n = 0$. Without loss of generality, suppose that this particular linear dependence has n as small as possible among all such relations. Then θ is a root of

$$P(X) = c_0 + c_1X + c_2X^2 + \cdots + c_{n-1}X^{n-1} + X^n.$$

Consider the substitution homomorphism $E : \mathbb{Q}[X] \to \mathbb{C}$ given by $E(A(X)) = A(\theta)$. This ring homomorphism carries $\mathbb{Q}[X]$ onto the ring $\mathbb{Q}[\theta]$, and the kernel is some ideal I. Specifically I consists of all polynomials $A(X)$ with $A(\theta) = 0$, and $P(X)$ is one of these of lowest possible degree. Proposition 5.8 shows that I consists of all multiples of some polynomial, and that polynomial may be taken to be $P(X)$ by minimality of the integer n. Proposition 8.1 therefore shows that $\mathbb{Q}[\theta] \cong \mathbb{Q}[X]/P(X)\mathbb{Q}[X]$ as a ring. If $P(X)$ were to have a nontrivial factorization as $P(X) = Q_1(X)Q_2(X)$, then $P(\theta) = 0$ would imply $Q_1(\theta) = 0$ or $Q_2(\theta) = 0$, and we would obtain a contradiction to the minimality of n. Therefore $P(X)$ is prime. By Example 2 earlier in the section, $I = P(X)\mathbb{Q}[X]$ is a nonzero prime ideal, and Proposition 8.12 shows that it is maximal. By Proposition 8.10 the quotient ring $\mathbb{Q}[\theta] = \mathbb{Q}[X]/P(X)\mathbb{Q}[X]$ is a field. These computations with $\mathbb{Q}[\theta]$ underlie the first part of the theory of fields that we shall develop in Chapter IX.

4. Unique Factorization

We have seen that the positive members of $\mathbb{Z}$ and the nonzero members of $\mathbb{K}[X]$, when $\mathbb{K}$ is a field, factor into the products of "primes" and that these factorizations are unique up to order and up to adjusting each of the prime factors in $\mathbb{K}[X]$ by a unit. In this section we shall investigate this idea of unique factorization more generally. Zero divisors are problematic from the point of view of factorization, and it will be convenient to exclude them. Therefore we work exclusively with integral domains.

The first observation is that unique factorization is not a completely general notion for integral domains. Let us consider an example in detail.

EXAMPLE. $R = \mathbb{Z}[\sqrt{-5}\,]$. This is the subring of $\mathbb{C}$ whose members are of the form $a + b\sqrt{-5}$ with a and b integers. Since $(a + b\sqrt{-5}\,)(c + d\sqrt{-5}\,) = (ac - 5cd) + (ad + bc)\sqrt{-5}$, R is closed under multiplication and is indeed a subring. Define $N(a + b\sqrt{-5}\,) = a^2 + 5b^2 = (a + b\sqrt{-5}\,)\overline{(a + b\sqrt{-5})}$. This is a nonnegative-integer-valued function on R and is 0 only on the 0 element of R. Since complex conjugation is an automorphism of $\mathbb{C}$, we check immediately that

$$N\big((a + b\sqrt{-5}\,)(c + d\sqrt{-5}\,)\big) = N(a + b\sqrt{-5}\,)N(c + d\sqrt{-5}\,).$$

The group of units of R, i.e., of elements with inverses under multiplication, is denoted by $R^\times$ as usual. If r is in $R^\times$, then $rr^{-1} = 1$, and so $N(r)N(r^{-1}) = N(1) = 1$. Consequently the units r of R all have $N(r) = 1$. Setting $a^2+5b^2 = 1$, we see that the units are ± 1. The product formula for N shows that if we start factoring a member of R, then factor its factors, and so on, and if we forbid factorizations into two factors when one is a unit, then the process of factorization has to stop at some point. So complete factorization makes sense. Now consider the equality

$$6 = (1 + \sqrt{-5}\,)(1 - \sqrt{-5}\,) = 2 \cdot 3.$$

The factors here have $N(1 + \sqrt{-5}\,) = N(1 - \sqrt{-5}\,) = 6$, $N(2) = 4$, and $N(3) = 9$. Considering the possible values of $a^2 + 5b^2$, we see that $N(\,\cdot\,)$ does not take on either of the values 2 and 3 on R. Consequently $1 + \sqrt{-5}$, $1 - \sqrt{-5}$, 2, and 3 do not have nontrivial factorizations. On the other hand, consideration of the values of $N(\,\cdot\,)$ shows that 2 and 3 are not products of either of $1 \pm \sqrt{-5}$ with units. We conclude that the displayed factorizations of 6 show that unique factorization has failed.

Thus unique factorization is not universal for integral domains. It is time to be careful about terminology. With $\mathbb{Z}$ and $\mathbb{K}[X]$, we have referred to the individual factors in a complete factorization as "primes." Their defining property in Chapter I was that they could not be factored further in nontrivial fashion. Primes in these rings were shown to have the additional property that if a prime divides a product then it divides one of the factors. It is customary to separate these two properties for general integral domains. Let us say that a nonzero element a **divides** b if $b = ac$ for some c. In this case we say also that a is a **factor** of b. In an integral domain R, a nonzero element r that is not a unit is said to be **irreducible** if every factorization $r = r_1r_2$ in R has the property that either r_1 or r_2 is a unit. Nonzero nonunits that are not irreducible are said

to be **reducible**. A nonzero element p that is not a unit is said to be **prime**[6] if the condition that p divides a product ab always implies that p divides a or p divides b.

Prime implies irreducible. In fact, if p is a prime that is reducible, let us write $p = r_1 r_2$ with neither r_1 nor r_2 equal to a unit. Since p is prime, p divides r_1 or r_2, say r_1. Then $r_1 = pc$ with c in R, and we obtain $p = r_1 r_2 = pcr_2$. Since R is an integral domain, $1 = cr_2$, and r_2 is exhibited as a unit with inverse c, in contradiction to the assumption that r_2 is not a unit.

On the other hand, irreducible does not imply prime. In fact, we saw in $\mathbb{Z}[\sqrt{-5}\,]$ that $1 + \sqrt{-5}$ is irreducible. But $1 + \sqrt{-5}$ divides $2 \cdot 3$, and $1 + \sqrt{-5}$ does not divide either of 2 or 3. Therefore $1 + \sqrt{-5}$ is not prime.

We shall see in a moment that the distinction between "irreducible" and "prime" lies at the heart of the question of unique factorization. Let us make a definition that helps identify our problem precisely. We say that an integral domain R is a **unique factorization domain** if R has the two properties

(UFD1) every nonzero nonunit of R is a finite product of irreducible elements,

(UFD2) the factorization in (UFD1) is always unique up to order and to multiplication of the factors by units.

The problem that arises for us for a given R is to decide whether R is a unique factorization domain. The following proposition shows the relevance of the distinction between "irreducible" and "prime."

Proposition 8.13. In an integral domain R in which (UFD1) holds, the condition (UFD2) is equivalent to the condition

(UFD2′) every irreducible element is prime.

REMARKS. In fact, showing that irreducible implies prime was the main step in Chapter I in proving unique factorization for positive integers and for $\mathbb{K}[X]$ when $\mathbb{K}$ is a field. The mechanism for carrying out the proof that irreducible implies prime for those settings will be abstracted in Theorems 8.15 and 8.17.

PROOF. Suppose that (UFD2) holds, that p is an irreducible element, and that p divides ab. We are to show that p divides a or p divides b. We may assume that $ab \neq 0$. Write $ab = pc$, and let $a = \prod_i p_i$, $b = \prod_j p'_j$, and $c = \prod_k q_k$ be factorizations via (UFD1) into products of irreducible elements.

[6]This definition enlarges the definition of "prime" in $\mathbb{Z}$ to include the negatives of the usual prime numbers. Unique factorization immediately extends to nonzero integers of either sign, but the prime factors are now determined only up to factors of ± 1. In cases where confusion about the sign of an integer prime might arise, the text will henceforth refer to "primes of $\mathbb{Z}$" or "integer primes" when both signs are allowed, and to "positive primes" or "prime numbers" when the primes are understood to be as in Chapter I.

Then $\prod_{i,j} p_i p'_j = p \prod_k q_k$. By (UFD2) one of the factors on the left side is εp for some unit ε. Then p either is of the form $\varepsilon^{-1} p_i$ and then p divides a, or is of the form $\varepsilon^{-1} p'_j$ and then p divides b. Hence (UFD2′) holds.

Conversely suppose that (UFD2′) holds. Let the nonzero nonunit r have two factorizations into irreducible elements as $r = p_1 p_2 \cdots p_m = \varepsilon_0 q_1 q_2 \cdots q_n$ with $m \le n$ and with ε_0 a unit. We prove the uniqueness by induction on m, the case $m = 0$ following vacuously since r is not a unit and the case $m = 1$ following from the definition of "irreducible." Inductively from (UFD2′) we know that p_m divides q_k for some k. Since q_k is irreducible, $q_k = \varepsilon p_m$ for some unit ε. Thus we can cancel q_k and obtain $p_1 p_2 \cdots p_{m-1} = \varepsilon_0 \varepsilon q_1 q_2 \cdots \widehat{q_k} \cdots q_n$, the hat indicating an omitted factor. By induction the factors on the two sides here are the same except for order and units. Thus the same conclusion is valid when comparing the two sides of the equality $p_1 p_2 \cdots p_m = \varepsilon_0 q_1 q_2 \cdots q_n$. The induction is complete, and (UFD2) follows. □

It will be convenient to simplify our notation for ideals. In any commutative ring R with identity, if a is in R, we let (a) denote the ideal Ra generated by a. An ideal of this kind with a single generator is called a **principal ideal**. More generally, if $a_1, \ldots, a_n$ are members of R, then $(a_1, \ldots, a_n)$ denotes the ideal $Ra_1 + \cdots + Ra_n$ generated by $a_1, \ldots, a_n$. For example, in $\mathbb{Z}[X]$, $(2, X)$ denotes the ideal $2\mathbb{Z} + X\mathbb{Z}$ of all polynomials whose constant term is even. The following condition explains a bit the mystery of what it means for an element to be prime.

Proposition 8.14. A nonzero element p in an integral domain R is prime if and only if the ideal (p) in R is prime.

PROOF. Suppose that the element p is prime. Then the ideal (p) is not R; in fact, otherwise 1 would have to be of the form $1 = rp$ for some $r \in R$, r would be a multiplicative inverse of p, and p would be a unit. Now suppose that a product ab is in the ideal (p). Then $ab = pr$ for some r in R, and p divides ab. Since p is prime, p divides a or p divides b. Therefore the ideal (p) is prime.

Conversely suppose that (p) is a prime ideal with $p \neq 0$. Since $(p) \neq R$, p is not a unit. If p divides the product ab, then $ab = pc$ for some c in R. Hence ab is in (p). Since (p) is assumed prime, either a is in (p) or b is in (p). In the first case, p divides a, and in the second case, p divides b. Thus the element p is prime. □

An integral domain R is called a **principal ideal domain** if every ideal in R is principal. At the beginning of Section 3, we saw a reminder that $\mathbb{Z}$ is a principal ideal domain and that so is $\mathbb{K}[X]$ whenever $\mathbb{K}$ is a field. It turns out that unique factorization for these cases is a consequence of this fact.

Theorem 8.15. Every principal ideal domain is a unique factorization domain.

REMARKS. Let R be the given principal ideal domain. Proposition 8.13 shows that it is enough to show that (UFD1) and (UFD2′) hold in R.

PROOF OF (UFD1). Let a_1 be a nonzero nonunit of R. Then the ideal (a_1) in R is proper and nonzero, and Proposition 8.8 shows that it is contained in a maximal ideal. Since R is a principal ideal domain, this maximal ideal is of the form (c_1) for some c_1, and c_1 is a nonzero nonunit. Maximal ideals are prime by Corollary 8.11, and Proposition 8.14 thus shows that c_1 is a prime element, necessarily irreducible. Therefore the inclusion $(a_1) \subseteq (c_1)$ shows that some irreducible element, namely c_1, divides a_1.

Write $a_1 = c_1 a_2$, and repeat the above argument with a_2. Iterating this construction, we obtain $a_n = c_n a_{n+1}$ for each n with c_n irreducible. Thus $a_1 = c_1 c_2 \cdots c_n a_{n+1}$ with $c_1, \ldots, c_n$ irreducible. Let us see that this process cannot continue indefinitely. Assuming the contrary, we are led to the strict inclusions

$$(a_1) \subsetneqq (a_2) \subsetneqq (a_3) \subsetneqq \cdots .$$

Put $I = \bigcup_{n=1}^{\infty} (a_n)$. Then I is an ideal. Since R is a principal ideal domain, $I = (a)$ for some element a. This element a must be in (a_k) for some k, and then we have $(a_k) = (a_{k+1}) = \cdots = (a)$. Since $(a_k) = (a_{k+1})$, c_k has to be a unit, contradiction. Thus a_k has no nontrivial factorization, and $a_1 = c_1 \cdots c_{k-1} a_k$ is the desired factorization. This proves (UFD1). □

PROOF OF (UFD2′). If p is an irreducible element, we prove that the ideal (p) is maximal. Corollary 8.11 shows that (p) is prime, and Proposition 8.14 shows that p is prime. Thus (UFD2′) will follow.

The element p, being irreducible, is not a unit. Thus (p) is proper. Suppose that I is an ideal with $I \supsetneqq (p)$. Since R is a principal ideal domain, $I = (c)$ for some c. Then $p = rc$ for some r in R. Since $I \neq (p)$, r cannot be a unit. Therefore the irreducibility of p implies that c is a unit. Then $I = (c) = (1) = R$, and we conclude that (p) is maximal. □

Let us record what is essentially a corollary of the proof.

Corollary 8.16. In a principal ideal domain, every nonzero prime ideal is maximal.

PROOF. Let (p) be a nonzero prime ideal. Proposition 8.14 shows that p is prime, and prime elements are automatically irreducible. The argument for (UFD2′) in the proof of Theorem 8.15 then deduces in the context of a principal ideal domain that (p) is maximal. □

Principal ideal domains arise comparatively infrequently, and recognizing them is not necessarily easy. The technique that was used with $\mathbb{Z}$ and $\mathbb{K}[X]$ generalizes slightly, and we take up that generalization now. An integral domain R is called a **Euclidean domain** if there exists a function $\delta : R \to \{\text{integers } \geq 0\}$ such that whenever a and b are in R with $b \neq 0$, there exist q and r in R with $a = bq + r$ and $\delta(r) < \delta(b)$. The ring $\mathbb{Z}$ of integers is a Euclidean domain if we take $\delta(n) = |n|$, and the ring $\mathbb{K}[X]$ for $\mathbb{K}$ a field is a Euclidean domain if we take $\delta(P(X))$ to be $2^{\deg P}$ if $P(X) \neq 0$ and to be 0 if $P(X) = 0$.

Another example of a Euclidean domain is the ring $\mathbb{Z}[\sqrt{-1}\,] = \mathbb{Z} + \mathbb{Z}\sqrt{-1}$ of **Gaussian integers**. It has $\delta(a+b\sqrt{-1}\,) = (a+b\sqrt{-1}\,)(a-b\sqrt{-1}\,) = a^2+b^2$, a and b being integers. Let us abbreviate $\sqrt{-1}$ as i. To see that δ has the required property, we first extend δ to $\mathbb{Q}[i]$, writing $\delta(x+yi) = (x+yi)(x-yi) = x^2+y^2$ if x and y are rational. We use the fact that

$$\delta(zz') = \delta(z)\delta(z') \qquad \text{for } z \text{ and } z' \text{ in } \mathbb{Q}[i],$$

which follows from the computation $\delta(zz') = zz' \cdot \overline{zz'} = z\bar{z}z'\overline{z'} = \delta(z)\delta(z')$. For any real number u, let $[u]$ be the greatest integer $\leq u$. Every real u satisfies $\big|[u + \frac{1}{2}] - u\big| \leq \frac{1}{2}$. Given $a + ib$ and $c + di$ with $c + di \neq 0$, we write

$$\frac{a+bi}{c+di} = \frac{(a+bi)(c-di)}{c^2+d^2} = \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2}\,i.$$

Put $p = \left[\frac{ac+bd}{c^2+d^2} + \frac{1}{2}\right]$, $q = \left[\frac{bc-ad}{c^2+d^2} + \frac{1}{2}\right]$, and $r+si = (a+bi)-(c+di)(p+qi)$. Then

$$a + bi = (c + di)(p + qi) + (r + si),$$

and

$$\delta(r+si) = \delta\big((a+bi)-(c+di)(p+qi)\big) = \delta(c+di)\delta\Big(\frac{a+bi}{c+di} - (p+qi)\Big).$$

The complex number $x + yi = \frac{a+bi}{c+di} - (p+qi) = \big(\frac{ac+bd}{c^2+d^2} - p\big) + \big(\frac{bc-ad}{c^2+d^2} - q\big)i$ has $|x| \leq \frac{1}{2}$ and $|y| \leq \frac{1}{2}$, and therefore $\delta(x+yi) = x^2+y^2 \leq \frac{1}{4}+\frac{1}{4} = \frac{1}{2}$. Hence $\delta(r+si) < \delta(c+di)$, as required.

Some further examples of this kind appear in Problems 13 and 25–26 at the end of the chapter. The matter is a little delicate. The ring $\mathbb{Z}[\sqrt{-5}\,]$ may seem superficially similar to $\mathbb{Z}[\sqrt{-1}\,]$. But $\mathbb{Z}[\sqrt{-5}\,]$ does not have unique factorization, and the following theorem, in combination with Theorem 8.15, assures us that $\mathbb{Z}[\sqrt{-5}\,]$ cannot be a Euclidean domain.

Theorem 8.17. Every Euclidean domain is a principal ideal domain.

PROOF. Let I be an ideal in R. We are to show that I is principal. Without loss of generality, we may assume that $I \neq 0$. Choose $b \neq 0$ in I with $\delta(b)$ as small as possible. Certainly $I \supseteq (b)$. If $a \neq 0$ is in I, write $a = bq + r$ with $\delta(r) < \delta(b)$. Then $r = a - bq$ is in I with $\delta(r) < \delta(b)$. The minimality of b forces $r = 0$ and $a = bq$. Thus $I \subseteq (b)$, and we conclude that $I = (b)$. □

5. Gauss's Lemma

In the previous section we saw that every principal ideal domain has unique factorization. In the present section we shall establish that certain additional integral domains have unique factorization, namely any integral domain $R[X]$ for which R is a unique factorization domain. A prototype is $\mathbb{Z}[X]$, which will be seen to have unique factorization even though there exist nonprincipal ideals like $(2, X)$ in the ring. An important example for applications, particularly in algebraic geometry, is $\mathbb{K}[X_1, \ldots, X_n]$, where $\mathbb{K}$ is a field; in this case our result is to be applied inductively, making use of the isomorphism $\mathbb{K}[X_1, \ldots, X_n] \cong \mathbb{K}[X_1, \ldots, X_{n-1}][X_n]$ given in Corollary 4.31.

For the conclusion that $R[X]$ has unique factorization if R does, the heart of the proof is an application of a result known as Gauss's Lemma, which we shall prove in this section. Gauss's Lemma has additional consequences for $R[X]$ beyond unique factorization, and we give them as well.

Before coming to Gauss's Lemma, let us introduce some terminology and prove one preliminary result. In any integral domain R, we call two nonzero elements a and b **associates** if $a = b\varepsilon$ for some ε in the group $R^\times$ of units. The property of being associates is an equivalence relation because $R^\times$ is a group.

Still with the nonzero integral domain R, let us define a **greatest common divisor** of two nonzero elements a and b to be any element c of R such that c divides both a and b and such that any divisor of a and b divides c. Any associate of a greatest common divisor of a and b is another greatest common divisor of a and b. Conversely if a and b have a greatest common divisor, then any two greatest common divisors are associates. In fact, if c and c' are greatest common divisors, then each of them divides both a and b, and the definition forces each of them to divide the other. Thus $c' = c\varepsilon$ and $c = c'\varepsilon'$, and then $c' = c'\varepsilon'\varepsilon$ and $1 = \varepsilon'\varepsilon$. Consequently ε is a unit, and c and c' are associates.

If R is a unique factorization domain, then any two nonzero elements a and b have a greatest common divisor. In fact, we decompose a and b into the product of a unit by powers of nonassociate irreducible elements as $a = \varepsilon \prod_{i=1}^m p_i^{k_i}$ and $b = \varepsilon' \prod_{j=1}^n p_j'^{l_j}$. For each p_j' such that p_j' is associate to some p_i, we replace p_j' by p_i in the factorization of b, adjusting ε' as necessary, and then we reorder

the factors of a and b so that the common p_i's are the ones for $1 \le i \le r$. Then $c = \prod_{i=1}^{r} p_i^{\min(k_i, l_i)}$ is a greatest common divisor of a and b. We write $\mathrm{GCD}(a, b)$ for a greatest common divisor of a and b; as we saw above, this is well defined up to a factor of a unit.[7]

One should not read too much into the notation. In a principal ideal domain if a and b are nonzero, then, as we shall see momentarily, $\mathrm{GCD}(a, b)$ is defined by the condition on ideals that

$$(\mathrm{GCD}(a, b)) = (a, b).$$

This condition implies that there exist elements x and y in R such that

$$xa + yb = \mathrm{GCD}(a, b).$$

However, in the integral domain $\mathbb{Z}[X]$, in which $\mathrm{GCD}(2, X) = 1$, there do not exist polynomials $A(X)$ and $B(X)$ with $A(X)2 + B(X)X = 1$.

To prove that $(\mathrm{GCD}(a, b)) = (a, b)$ in a principal ideal domain, write (c) for the principal ideal (a, b); c satisfies $c = xa + yb$ for some x and y in R. Since a and b lie in (c), $a = rc$ and $b = r'c$. Hence c divides both a and b. In the reverse direction if d divides a and b, then $ds = a$ and $ds' = b$. Hence $c = xa + yb = (xs + ys')d$, and d divides c. So c is indeed a greatest common divisor of a and b.

In a unique factorization domain the definition of greatest common divisor immediately extends to apply to n nonzero elements, rather than just two. We readily check up to a unit that

$$\mathrm{GCD}(a_1, \ldots, a_n) = \mathrm{GCD}\big(\mathrm{GCD}(a_1, \ldots, a_{n-1}), a_n\big).$$

Moreover, we can allow any of $a_2, \ldots, a_n$ to be 0, and there is no difficulty. In addition, we have

$$\mathrm{GCD}(da_1, \ldots, da_n) = d\,\mathrm{GCD}(a_1, \ldots, a_n) \qquad \text{up to a unit}$$

if d and a_1 are not 0.

Let R be a unique factorization domain. If $A(X)$ is a nonzero element of $R[X]$, we say that $A(X)$ is **primitive** if the GCD of its coefficients is a unit. In this case no prime of R divides all the coefficients of $A(X)$.

[7] Greatest common divisors can exist for certain integral domains that fail to have unique factorization, but we shall not have occasion to work with any such domains.

Theorem 8.18 (Gauss's Lemma). If R is a unique factorization domain, then the product of primitive polynomials is primitive.

PROOF #1. Arguing by contradiction, let $A(X) = a_m X^m + \cdots + a_0$ and $B(X) = b_n X^n + \cdots + b_0$ be primitive polynomials such that every coefficient of $A(X)B(X)$ is divisible by some prime p. Since $A(X)$ and $B(X)$ are primitive, we may choose k and l as small as possible such that p does not divide a_k and does not divide b_l. The coefficient of X^{k+l} in $A(X)B(X)$ is

$$a_0 b_{k+l} + a_1 b_{k+l-1} + \cdots + a_k b_l + \cdots + a_{k+l} b_0$$

and is divisible by p. Then all the individual terms, and their sum, are divisible by p except possibly for $a_k b_l$, and we conclude that p divides $a_k b_l$. Since p is prime and p divides $a_k b_l$, p must divide a_k or b_l, contradiction. □

PROOF #2. Arguing by contradiction, let $A(X)$ and $B(X)$ be primitive polynomials such that every coefficient of $A(X)B(X)$ is divisible by some prime p. Proposition 8.14 shows that the ideal (p) is prime, and Proposition 8.7 shows that $R' = R/(p)$ is an integral domain. Let $\varphi : R \to R'[X]$ be the composition of the quotient homomorphism $R \to R'$ and the inclusion of R' into constant polynomials in $R'[X]$, and let $\Phi : R[X] \to R'[X]$ be the corresponding substitution homomorphism of Proposition 4.24 that carries X to X. Since $A(X)$ and $B(X)$ are primitive, $\Phi(A(X))$ and $\Phi(B(X))$ are not zero. Their product $\Phi(A(X))\Phi(B(X)) = \Phi(A(X)B(X))$ is 0 since p divides every coefficient of $A(X)B(X)$, and this conclusion contradicts the assertion of Proposition 4.29 that $R'[X]$ is an integral domain. □

Let F be the field of fractions of the unique factorization domain R. The consequences of Theorem 8.18 exploit a simple relationship between $R[X]$ and $F[X]$, which we state below as Proposition 8.19. Once that proposition is in hand, we can state the consequences of Theorem 8.18. If $A(X)$ is a nonzero polynomial in $R[X]$, let $c(A)$ to be the greatest common divisor of the coefficients, i.e.,

$$c(A) = \mathrm{GCD}(a_n, \ldots, a_1, a_0) \qquad \text{if } A(X) = a_n X^n + \cdots + a_1 X + a_0.$$

The element $c(A)$ is well defined up to a factor of a unit. In this notation the definition of "primitive" becomes, $A(X)$ is primitive if and only if $c(A)$ is a unit. We shall make computations with $c(A)$ as if it were a member of R, in order to keep the notation simple. To be completely rigorous, one should regard $c(A)$ as an orbit of the group $R^\times$ of units in R, using equality to refer to equality of orbits.

If $A(X)$ is not necessarily primitive, then at least $c(A)$ divides each coefficient of $A(X)$, and hence $c(A)^{-1}A(X)$ is in $R[X]$, say with coefficients $b_n, \ldots, b_1, b_0$. Then we have

$$\begin{aligned} c(A) &= \mathrm{GCD}(a_n, \ldots, a_1, a_0) = \mathrm{GCD}(c(A)b_n, \ldots, c(A)b_1, c(A)b_0) \\ &= c(A)\mathrm{GCD}(b_n, \ldots, b_1, b_0) = c(A)c\big(c(A)^{-1}A(X)\big) \end{aligned}$$

up to a unit factor, and hence $c\big(c(A)^{-1}A(X)\big)$ is a unit. We conclude that

$$A(X) \in R[X] \qquad \text{implies that} \qquad c(A)^{-1}A(X) \text{ is primitive.}$$

Proposition 8.19. Let R be a unique factorization domain, and let F be its field of fractions. If $A(X)$ is any nonzero polynomial in $F[X]$, then there exist α in F and $A_0(X)$ in $R[X]$ such that $A(X) = \alpha A_0(X)$ with $A_0(X)$ primitive. The scalar α and the polynomial $A_0(X)$ are unique up to multiplication by units in R.

REMARK. We call $A_0(X)$ the **associated primitive polynomial** to $A(X)$. According to the proposition, it is unique up to a unit factor in R.

PROOF. Let $A(X) = c_nX^n + \cdots + c_1X + c_0$ with each c_k in F. We can write each c_k as $a_kb_k^{-1}$ with a_k and b_k in R and $b_k \neq 0$. We clear fractions. That is, we let $\beta = \prod_{k=0}^{n} b_k$. Then the k^{th} coefficient of $\beta A(X)$ is $a_k \prod_{l\neq k} b_l$ and is in R. Hence $\beta A(X)$ is in $R[X]$. The observation just before the proposition shows that $c(\beta A)^{-1}\beta A$ is primitive. Thus $A(X) = \alpha A_0(X)$ with $\alpha = \beta^{-1}c(\beta A)$ and $A_0(X) = c(\beta A)^{-1}\beta A(X)$, $A_0(X)$ being primitive. This proves existence.

If $\alpha_1 A_1(X) = \alpha_2 A_2(X)$ with α_1 and α_2 in F and with $A_1(X)$ and $A_2(X)$ primitive, choose $r \neq 0$ in R such that $r\alpha_1$ and $r\alpha_2$ are in R. Up to unit factors in R, we then have $r\alpha_1 = r\alpha_1 c(A_1) = c(r\alpha_1 A_1) = c(r\alpha_2 A_2) = r\alpha_2 c(A_2) = r\alpha_2$. Hence, up to a unit factor in R, we have $\alpha_1 = \alpha_2$. This proves uniqueness. □

Corollary 8.20. Let R be a unique factorization domain, and let F be its field of fractions.

(a) Let $A(X)$ and $B(X)$ be nonzero polynomials in $R[X]$, and suppose that $B(X)$ is primitive. If $B(X)$ divides $A(X)$ in $F[X]$, then it divides $A(X)$ in $R[X]$.

(b) If $A(X)$ is an irreducible polynomial in $R[X]$ of degree > 0, then $A(X)$ is irreducible in $F[X]$.

(c) If $A(X)$ is a monic polynomial in $R[X]$ and if $B(X)$ is a monic factor of $A(X)$ within $F[X]$, then $B(X)$ is in $R[X]$.

(d) If $A(X)$, $B(X)$, and $C(X)$ are in $R[X]$ with $A(X)$ primitive and with $A(X) = B(X)C(X)$, then $B(X)$ and $C(X)$ are primitive.

PROOF. In (a), write $A(X) = B(X)Q(X)$ in $F(X)$, and let $Q(X) = \rho Q_0(X)$ be a decomposition of $Q(X)$ as in Proposition 8.19. Since $c(A)^{-1}A(X)$ is primitive, the corresponding decomposition of $A(X)$ is $A(X) = c(A)\big(c(A)^{-1}A(X)\big)$. The equality $A(X) = \rho B(X)Q_0(X)$ then reads $c(A)(c(A)^{-1}A(X)) = \rho B(X)Q_0(X)$. Since $B(X)Q_0(X)$ is primitive according to Theorem 8.18, the uniqueness in Proposition 8.19 shows that $c(A)^{-1}A(X) = B(X)Q_0(X)$ except possibly for a unit factor in R. Then $B(X)$ divides $A(X)$ with quotient $c(A)Q_0(X)$, apart from a unit factor in R. Since $c(A)Q_0(X)$ is in $R[X]$, (a) is proved.

In (b), the condition that $\deg A(X) > 0$ implies that $A(X)$ is not a unit in $F[X]$. Arguing by contradiction, suppose that $A(X) = B(X)Q(X)$ in $F[X]$ with neither of $B(X)$ and $Q(X)$ of degree 0. Let $B(X) = \beta B_0(X)$ be a decomposition of $B(X)$ as in Proposition 8.19. Then we have $A(X) = B_0(X)(\beta Q(X))$, and (a) shows that $\beta Q(X)$ is in $R[X]$, in contradiction to the assumed irreducibility of $A(X)$ in $R[X]$.

In (c), write $A(X) = B(X)Q(X)$, and let $B(X) = \beta B_0(X)$ be a decomposition of $B(X)$ as in Proposition 8.19. Then we have $A(X) = B_0(X)(\beta Q(X))$ with $\beta Q(X)$ in $F[X]$. Conclusion (a) shows that $\beta Q(X)$ is in $R[X]$. If $b \in R$ is the leading coefficient of $B_0(X)$ and if $q \in R$ is the leading coefficient of $\beta Q(X)$, then we have $1 = bq$, and consequently b and q are units in R. Since $B(X) = \beta B_0(X)$ and $B(X)$ is monic, $1 = \beta b$, and therefore $\beta = b^{-1}$ is a unit in R. Hence $B(X)$ is in $R[X]$.

In (d), we argue along the same lines as in (a). We may take $B(X) = c(B)(c(B)^{-1}B(X))$ and $C(X) = c(C)(c(C)^{-1}C(X))$ as decompositions of $B(X)$ and $C(X)$ according to Proposition 8.19. Then we have $A(X) = (c(B)c(C))\big[c(B)^{-1}B(X)c(C)^{-1}C(X)\big]$. Theorem 8.18 says that the factor in brackets is primitive, and the uniqueness in Proposition 8.19 shows that $1 = c(B)c(C)$, up to unit factors. Therefore $c(B)$ and $c(C)$ are units in R, and $B(X)$ and $C(X)$ are primitive. □

Corollary 8.21. If R is a unique factorization domain, then the ring $R[X]$ is a unique factorization domain.

REMARK. As was mentioned at the beginning of the section, $\mathbb{Z}[X]$ and $\mathbb{K}[X_1, \ldots, X_n]$, when $\mathbb{K}$ is a field, are unique factorization domains as a consequence of this result.

PROOF. We begin with the proof of (UFD1). Suppose that $A(X)$ is a nonzero member of $R[X]$. We may take its decomposition according to Proposition 8.19 to be $A(X) = c(A)(c(A)^{-1}A(X))$. Consider divisors of $c(A)^{-1}A(X)$ in $R[X]$. These are all primitive, according to (d). Hence those of degree 0 are units in R. Thus any nontrivial factorization of $c(A)^{-1}A(X)$ is into two factors of strictly lower degree, both primitive. In a finite number of steps, this process of factorization with primitive factors has to stop. We can then factor $c(A)$ within R. Combining the factorizations of $c(A)$ and $c(A)^{-1}A(X)$, we obtain a factorization of $A(X)$.

For (UFD2′), let $P(X)$ be irreducible in $R[X]$. Since the factorization $P(X) = c(P)(c(P)^{-1}P(X))$ has to be trivial, either $c(P)$ is a unit, in which case $P(X)$ is primitive, or $c(P)^{-1}P(X)$ is a unit, in which case $P(X)$ has degree 0. In either case, suppose that $P(X)$ divides a product $A(X)B(X)$.

In the first case, $P(X)$ is primitive. Since $F[X]$ is a principal ideal domain, hence a unique factorization domain, either $P(X)$ divides $A(X)$ in $F[X]$ or $P(X)$

divides $B(X)$ in $F[X]$. By symmetry we may assume that $P(X)$ divides $A(X)$ in $F[X]$. Then (a) shows that $P(X)$ divides $A(X)$ in $R[X]$.

In the second case, $P(X) = P$ has degree 0 and is prime in R. Put $R' = R(P)$ as in Proof #2 of Theorem 8.18. Then $A(X)B(X)$ maps to zero in the integral domain $R'[X]$, and hence $A(X)$ or $B(X)$ is in $P\,R[X]$. □

The final application, Eisenstein's irreducibility criterion, is proved somewhat in the style of Gauss's Lemma (Theorem 8.18). We shall give only the analog of Proof #1 of Gauss's Lemma, leaving the analog of Proof #2 to Problem 21 at the end of the chapter.

Corollary 8.22 (Eisenstein's irreducibility criterion). Let R be a unique factorization domain, let F be its field of fractions, and let p be a prime in R. If $A(X) = a_N X^N + \cdots + a_1 X + a_0$ is a polynomial of degree ≥ 1 in $R[X]$ such that p divides $a_{N-1}, \ldots, a_0$ but not a_N and such that p^2 does not divide a_0, then $A(X)$ is irreducible in $F[X]$.

REMARK. The polynomial $A(X)$ will be irreducible in $R[X]$ also unless all its coefficients are divisible by some nonunit of R.

PROOF. Without loss of generality, we may replace $A(X)$ by $c(A)^{-1}A(X)$ and thereby assume that $A(X)$ is primitive; this adjustment makes use of the hypothesis that p does not divide a_N. Corollary 8.20b shows that it is enough to prove irreducibility in $R[X]$. Assuming the contrary, suppose that $A(X)$ factors in $R[X]$ as $A(X) = B(X)C(X)$ with $B(X) = b_m X^m + \cdots + b_1 X + b_0$, $C(X) = c_n X^n + \cdots + c_1 X + c_0$, and neither of $B(X)$ and $C(X)$ equal to a unit. Corollary 8.20d shows that $B(X)$ and $C(X)$ are primitive. In particular, $B(X)$ and $C(X)$ have to be nonconstant polynomials. Define $a_k = 0$ for $k > N$, $b_k = 0$ for $k > m$, and $c_k = 0$ for $k > n$. Since p divides $a_0 = b_0 c_0$ and p is prime, p divides either b_0 or c_0. Without loss of generality, suppose that p divides b_0. Since p^2 does not divide a_0, p does not divide c_0.

We show, by induction on k, that p divides b_k for every $k < N$. The case $k = 0$ is the base case of the induction. If p divides b_j for $j < k$, then we have

$$a_k = b_0 c_k + b_1 c_{k-1} + \cdots + b_{k-1} c_1 + b_k c_0.$$

Since $k < N$, the left side is divisible by p. The inductive hypothesis shows that p divides every term on the right side except possibly the last. Consequently p divides $b_k c_0$. Since p does not divide c_0, p divides b_k. This completes the induction.

Since $C(X)$ is nonconstant, the degree of $B(X)$ is $< N$, and therefore we have shown that every coefficient of $B(X)$ is divisible by p. Then $c(B)$ is divisible by p, in contradiction to the fact that $B(X)$ is primitive. □

EXAMPLES.

(1) Cyclotomic polynomials in $\mathbb{Q}[X]$. Let us see for each prime number p that the polynomial $\Phi(X) = X^{p-1} + X^{p-2} + \cdots + X + 1$ is irreducible in $\mathbb{Q}[X]$. We have $X^p - 1 = (X-1)\Phi(X)$. Replacing $X-1$ by Y gives $(Y+1)^p - 1 = Y\Phi(Y+1)$. The left side, by the Binomial Theorem, is $\sum_{k=1}^{p} \binom{p}{k} Y^k$. Hence $\Phi(Y+1) = \sum_{k=1}^{p} \binom{p}{k} Y^{k-1}$. The binomial coefficient $\binom{p}{k}$ is divisible by p for $1 \le k \le p-1$ since p is prime, and therefore the polynomial $\Psi(Y) = \Phi(Y+1)$ satisfies the condition of Corollary 8.22 for the ring $\mathbb{Z}$. Hence $\Psi(Y)$ is irreducible over $\mathbb{Q}[Y]$. A nontrivial factorization of $\Phi(X)$ would yield a nontrivial factorization of $\Psi(Y)$, and hence $\Phi(X)$ is irreducible over $\mathbb{Q}[X]$.

(2) Certain polynomials in $\mathbb{K}[X, Y]$ when $\mathbb{K}$ is a field. Since $\mathbb{K}[X, Y] \cong \mathbb{K}[X][Y]$, it follows that $\mathbb{K}[X, Y]$ is a unique factorization domain, and any member of $\mathbb{K}[X, Y]$ can be written as $A(X, Y) = a_n(X)Y^n + \cdots + a_1(X)Y + a_0(X)$. The polynomial X is prime in $\mathbb{K}[X, Y]$, and Corollary 8.22 therefore says that $A(X, Y)$ is irreducible in $\mathbb{K}(X)[Y]$ if X does not divide $a_n(X)$ in $\mathbb{K}[X]$, X divides $a_{n-1}(X), \ldots, a_0(X)$ in $\mathbb{K}[X]$, and X^2 does not divide $a_0(X)$ in $\mathbb{K}[X]$. The remark with the corollary points out that $A(X, Y)$ is irreducible in $\mathbb{K}[X, Y]$ if also there is no nonconstant polynomial in $\mathbb{K}[X]$ that divides every $a_k(X)$. For example, $Y^5 + XY^2 + XY + X$ is irreducible in $\mathbb{K}[X, Y]$.

6. Finitely Generated Modules

The Fundamental Theorem of Finitely Generated Abelian Groups (Theorem 4.56) says that every finitely generated abelian group is a direct sum of cyclic groups. If we think of abelian groups as $\mathbb{Z}$ modules, we can ask whether this theorem has some analog in the context of R modules. The answer is yes—the theorem readily extends to the case that $\mathbb{Z}$ is replaced by an arbitrary principal ideal domain. The surprising addendum to the answer is that we have already treated a second special case of the generalized theorem. That case arises when the principal ideal domain is $\mathbb{K}[X]$ for some field $\mathbb{K}$. If V is a finite-dimensional vector space over $\mathbb{K}$ and $L : V \to V$ is a $\mathbb{K}$ linear map, then V becomes a $\mathbb{K}[X]$ module under the definition $Xv = L(v)$. This module is finitely generated even without the X present because V is finite-dimensional, and the generalized theorem that we prove in this section recovers the analysis of L that we carried out in Chapter V. When $\mathbb{K}$ is algebraically closed, we obtain the Jordan canonical form; for general $\mathbb{K}$, we obtain a different canonical form involving cyclic subspaces that was worked out in Problems 32–40 at the end of Chapter V.

The definitions for the generalization of Theorem 4.56 are as follows. Let R be a principal ideal domain. A subset S of an R module M is called a set of **generators** of M if M is the smallest R submodule of M containing all the

members of S. If $\{m_s \mid s \in S\}$ is a subset of M, then the set of all finite sums $\sum_{s\in S} r_s m_s$ is an R submodule, but it need not contain the elements m_s and therefore need not be the R submodule generated by all the m_s. However, if M is unital, then taking $r_{s_0} = 1$ and all other r_s equal to 0 exhibits m_{s_0} as in the R submodule of all finite sums $\sum_{s\in S} r_s m_s$. For this reason we shall insist that all the R submodules in this section be unital.

We say that the R module M is **finitely generated** if it has a finite set of generators. The main theorem gives the structure of unital finitely generated R modules when R is a principal ideal domain. We need to take a small preliminary step that eliminates technical complications from the discussion, the same step that was carried out in Lemma 4.51 and Proposition 4.52 in the case of $\mathbb{Z}$ modules, i.e., abelian groups.

Lemma 8.23. Let R be a commutative ring with identity, and let $\varphi : M \to N$ be a homomorphism of unital R modules. If $\ker\varphi$ and $\text{image}\,\varphi$ are finitely generated, then M is finitely generated.

PROOF. Let $\{x_1, \dots, x_m\}$ and $\{y_1, \dots, y_n\}$ be respective finite sets of generators for $\ker\varphi$ and $\text{image}\,\varphi$. For $1 \le j \le n$, choose x'_j in M with $\varphi(x'_j) = y_j$. We shall prove that $\{x_1, \dots, x_m, x'_1, \dots, x'_n\}$ is a set of generators for M. Thus let x be in M. Since $\varphi(x)$ is in $\text{image}\,\varphi$, there exist $r_1, \dots, r_n$ in R with $\varphi(x) = r_1y_1 + \cdots + r_ny_n$. The element $x' = r_1x'_1 + \cdots + r_nx'_n$ of M has $\varphi(x') = r_1y_1 + \cdots + r_ny_n = \varphi(x)$. Therefore $\varphi(x - x') = 0$, and there exist $s_1, \dots, s_m$ in R such that $x - x' = s_1x_1 + \cdots + s_mx_m$. Consequently

$$x = s_1x_1 + \cdots + s_mx_m + x' = s_1x_1 + \cdots + s_mx_m + r_1x'_1 + \cdots + r_nx'_n. \quad \square$$

Proposition 8.24. If R is a principal ideal domain, then any R submodule of a finitely generated unital R module is finitely generated. Moreover, any R submodule of a singly generated unital R module is singly generated.

REMARK. The proof will show that if M can be generated by n elements, then so can the unital R submodule.

PROOF. Let M be unital and finitely generated with a set $\{m_1, \dots, m_n\}$ of n generators, and define $M_k = Rm_1 + \cdots + Rm_k$ for $1 \le k \le n$. Then $M_n = M$ since M is unital. We shall prove by induction on k that every R submodule of M_k is finitely generated. The case $k = n$ then gives the proposition. For $k = 1$, suppose that S is an R submodule of $M_1 = Rm_1$. Since S is an R submodule and every member of S lies in Rm_1, the subset I of all r in R with rm_1 in S is an ideal with $Im_1 = S$. Since every ideal in R is singly generated, we can write $I = (r_0)$. Then $S = Im_1 = Rr_0m_1$, and the single element r_0m_1 generates S.

Assume inductively that every R submodule of M_k is known to be finitely generated, and let N_{k+1} be an R submodule of M_{k+1}. Let $q : M_{k+1} \to M_{k+1}/M_k$

be the quotient R homomorphism, and let φ be the restriction $q\big|_{N_{k+1}}$, mapping N_{k+1} into M_{k+1}/M_k. Then $\ker\varphi = N_{k+1} \cap M_k$ is an R submodule of M_k and is finitely generated by the inductive hypothesis. Also, image φ is an R submodule of M_{k+1}/M_k, which is singly generated with generator equal to the coset of m_{k+1}. Since an R submodule of a singly generated unital R module was shown in the previous paragraph to be singly generated, image φ is finitely generated. Applying Lemma 8.23 to φ, we see that N_{k+1} is finitely generated. This completes the induction and the proof. □

According to the definition in Example 9 of modules in Section 1, a free R module is a direct sum, finite or infinite, of copies of the R module R. A free R module is said to have **finite rank** if *some* direct sum is a finite direct sum. A unital R module M is said to be **cyclic** if it is singly generated, i.e., if $M = Rm_0$ for some m_0 in M. In this case, we have an R isomorphism $M \cong R/I$, where I is the ideal $\{r \in R \mid rm_0 = 0\}$.

Before coming to the statement of the theorem and the proof, let us discuss the heart of the matter, which is related to row reduction of matrices. We regard the space $M_{1n}(R)$ of all 1-row matrices with n entries in R as a free R module. Suppose that R is a principal ideal domain, and suppose that we have a particular 2-by-n matrix with entries in R and with the property that the two rows have nonzero elements a and b, respectively, in the first column. We can regard the set of R linear combinations of the two rows of our particular matrix as an R submodule of the free R module $M_{1n}(R)$. Let $c = \mathrm{GCD}(a, b)$. This member of R is defined only up to multiplication by a unit, but we make a definite choice of it. The idea is that we can do a kind of invertible row-reduction step that simultaneously replaces the two rows of our 2-by-n matrix by a first row whose first entry is c and a second row whose first entry is 0; in the process the corresponding R submodule of $M_{1n}(R)$ will be unchanged. In fact, we saw in the previous section that the hypothesis on R implies that there exist members x and y of R with $xa + yb = c$. Since c divides a and b, we can rewrite this equality as $x(ac^{-1}) + y(bc^{-1}) = 1$. Then the 2-by-2 matrix $M = \left(\begin{smallmatrix} x & y \\ -bc^{-1} & ac^{-1} \end{smallmatrix}\right)$ with entries in R has the property that

$$\begin{pmatrix} x & y \\ -bc^{-1} & ac^{-1} \end{pmatrix} \begin{pmatrix} a & * \\ b & * \end{pmatrix} = \begin{pmatrix} c & * \\ 0 & * \end{pmatrix}.$$

This equation shows explicitly that the rows of $\left(\begin{smallmatrix} c & * \\ 0 & * \end{smallmatrix}\right)$ lie in the R linear span of the rows of $\left(\begin{smallmatrix} a & * \\ b & * \end{smallmatrix}\right)$. The key fact about M is that its determinant $x(ac^{-1}) + y(bc^{-1})$ is 1 and that M is therefore invertible with entries in R: the inverse is just $M^{-1} = \left(\begin{smallmatrix} ac^{-1} & -y \\ bc^{-1} & x \end{smallmatrix}\right)$. This invertibility shows that the rows of $\left(\begin{smallmatrix} a & * \\ b & * \end{smallmatrix}\right)$ lie in the R linear span of $\left(\begin{smallmatrix} c & * \\ 0 & * \end{smallmatrix}\right)$. Consequently the R linear span of the rows of our given 2-by-n matrix is preserved under left multiplication by M.

In effect we can do the same kind of row reduction of matrices over R as we did with matrices over $\mathbb{Z}$ in the proof of Theorem 4.56. The only difference is that this time we do not see constructively how to find the x and y that relate a, b, and c. Thus we would lack some information if we actually wanted to follow through and calculate a particular example. We were able to make calculations to imitate the proof of Theorem 4.56 because we were able to use the Euclidean algorithm to arrive at what x and y are. In the present context we would be able to make explicit calculations if R were a Euclidean domain.

Theorem 8.25 (Fundamental Theorem of Finitely Generated Modules). If R is a principal ideal domain, then

(a) the number of R summands in a free R module of finite rank is independent of the direct-sum decomposition,
(b) any R submodule of a free R module of finite rank n is a free R module of rank $\le n$,
(c) any finitely generated unital R module is the finite direct sum of cyclic modules.

REMARK. Because of (a), it is meaningful to speak of the **rank** of a free R module of finite rank; it is the number of R summands. By convention the 0 module is a free R module of rank 0. Then the statement of (b) makes sense. Statement (c) will be amplified in Corollary 8.29 below. Some people use the name "Fundamental Theorem of Finitely Generated Modules" to refer to Corollary 8.29 rather than to Theorem 8.25.

PROOF. Let F be a free R module of the form $Rx_1 \oplus \cdots \oplus Rx_n$, and suppose that $y_1, \dots, y_m$ are elements of F such that no nontrivial combination $r_1y_1 + \cdots + r_my_m$ is 0. We argue as in the proof of Proposition 2.2. Define an m-by-n matrix C with entries in R by $y_i = \sum_{j=1}^n C_{ij}x_j$ for $1 \le i \le m$. If Q is the field of fractions of R, then we can regard C as a matrix with entries in Q. As such, the matrix has rank $\le n$. If $m > n$, then the rows are linearly dependent, and we can find members $q_1, \dots, q_m$ of Q, not all 0, such that $\sum_{i=1}^m q_iC_{ij} = 0$ for $1 \le j \le n$. Clearing fractions, we obtain members $r_1, \dots, r_m$ of R, not all 0, such that $\sum_{i=1}^m r_iC_{ij} = 0$ for $1 \le j \le n$. Then

$$\sum_{i=1}^m r_iy_i = \sum_{i=1}^m r_i\Big(\sum_{j=1}^n C_{ij}x_j\Big) = \sum_{j=1}^n \Big(\sum_{i=1}^m r_iC_{ij}\Big)x_j = \sum_{j=1}^n 0x_j = 0,$$

in contradiction to the assumed independence property of $y_1, \dots, y_m$. Therefore we must have $m \le n$.

If we apply this conclusion to a set $x_1, \dots, x_n$ that exhibits F as free and to another set, possibly infinite, that does the same thing, we find that the second

set has $\le n$ members. Reversing the roles of the two sets, we find that they both have n members. This proves (a).

For (b) and (c), we shall reduce the result to a lemma saying that a certain kind of result can be achieved by row and column reduction of matrices with entries in R. Let F be a free R module of rank n, defined by a subset $x_1, \dots, x_n$ of F, and let M be an R submodule of F. Proposition 8.24 shows that M is finitely generated. We let $y_1, \dots, y_m$ be generators, not necessarily with any independence property. Define an m-by-n matrix C with entries in R by $y_i = \sum_{j=1}^n C_{ij} x_j$. We can recover F as the set of R linear combinations of $x_1, \dots, x_n$, and we can recover M as the set of R linear combinations of $y_1, \dots, y_m$.

If B is an n-by-n matrix with entries in R and with determinant in the group $R^\times$ of units, then Corollary 5.5 shows that B^{-1} exists and has entries in R. If we define $x_i' = \sum_{j=1}^n B_{ij} x_j$, then any R linear combination of $x_1', \dots, x_n'$ is an R linear combination of $x_1, \dots, x_n$. Also, the computation $\sum_{i=1}^n (B^{-1})_{ki} x_i' = \sum_{i,j} (B^{-1})_{ki} B_{ij} x_j = \sum_j \delta_{kj} x_j = x_k$ shows that any R linear combination of $x_1, \dots, x_n$ is an R linear combination of $x_1', \dots, x_n'$. Thus we can recover the same F and M if we replace C by CB. Arguing in the same way with $y_1, \dots, y_m$ and $y_1', \dots, y_m'$, we see that we can recover the same F and M if we replace CB by ACB, where A is an m-by-m matrix with entries in R and with determinant in $R^\times$.

Lemma 8.26 below will say that we can find A and B such that the nonzero entries of $D = ACB$ are exactly the diagonal ones D_{kk} for $1 \le k \le l$, where l is a certain integer with $0 \le l \le \min(m, n)$.

That is, the resulting equations restricting $y_1', \dots, y_m'$ in terms of $x_1', \dots, x_n'$ will be of the form

$$y_k' = \begin{cases} D_{kk} x_k' & \text{for } 1 \le k \le l, \\ 0 & \text{for } l+1 \le k \le m. \end{cases} \tag{$*$}$$

Now let us turn to (b) and (c). For (b), the claim is that the elements y_k' with $1 \le k \le l$ exhibit M as a free R module. We know that $y_1', \dots, y_m'$ generate M and hence that $y_1', \dots, y_l'$ generate M. For the independence, suppose we can find members $r_1, \dots, r_l$ not all 0 in R such that $\sum_{k=1}^l r_k y_k' = 0$. Then substitution gives $\sum_{k=1}^l r_k D_{kk} x_k' = 0$, and the independence of $x_1', \dots, x_l'$ forces $r_k D_{kk} = 0$ for $1 \le k \le l$. Since R is an integral domain, $r_k = 0$ for such k. Thus indeed the elements y_k' with $1 \le k \le l$ exhibit M as a free R module. Since $l \le \min(m, n)$, the rank of M is at most the rank of F.

For (c), let S be a finitely generated unital R module, say with n generators. By the universal mapping property of free R modules (Example 9 in Section 1), there exists a free R module F of rank n with S as quotient. Let $x_1, \dots, x_n$ be generators of F that exhibit F as free, and let M be the kernel of the quotient R

homomorphism $M \to S$, so that $S \cong F/M$. Then (b) shows that M is a free R module of rank $m \le n$. Let $y_1, \dots, y_m$ be generators of M that exhibit M as free, and define an m-by-n matrix C with entries in R by $y_i = \sum_{j=1}^n C_{ij} x_j$ for $1 \le i \le m$. The result is that we are reduced to the situation we have just considered, and we can obtain equations of the form $(*)$ relating their respective generators, namely $y'_1, \dots, y'_m$ for M and $x'_1, \dots, x'_n$ for F.

For $1 \le k \le n$, define $F_k = Rx'_k$ and

$$M_k = \begin{cases} Ry'_k = RD_{kk}x'_k & \text{for } 1 \le k \le l, \\ 0 & \text{for } l+1 \le k \le n, \end{cases}$$

so that $M \cong M_1 \oplus \cdots \oplus M_n$. Then F_k/M_k is R isomorphic to the cyclic R module $R/(D_{kk})$ if $1 \le k \le l$, while $F_k/M_k = F_k$ is isomorphic to the cyclic R module R if $l+1 \le k \le n$. Applying Proposition 8.5, we obtain

$$F/M \cong (F_1 \oplus \cdots \oplus F_n)/(M_1 \oplus \cdots \oplus M_n) \cong (F_1/M_1) \oplus \cdots \oplus (F_n/M_n).$$

Thus F/M is exhibited as a direct sum of cyclic R modules. □

To complete the proof of Theorem 8.25, we are left with proving the following lemma, which is where row and column reduction take place.

Lemma 8.26. Let R be a principal ideal domain. If C is an m-by-n matrix with entries in R, then there exist an m-by-m matrix A with entries in R and with determinant in $R^\times$ and an n-by-n matrix B with entries in R and with determinant in $R^\times$ such that for some l with $0 \le l \le \min(m, n)$, the nonzero entries of $D = ACB$ are exactly the diagonal entries $D_{11}, D_{22}, \dots, D_{ll}$.

PROOF. The matrices A and B will be constructed as products of matrices of determinant ± 1, and then $\det A$ and $\det B$ equal ± 1 by Proposition 5.1a. The matrix A will correspond to row operations on C, and B will correspond to column operations. Each factor will be the identity except in some 2-by-2 block. Among the row and column operations of interest are the interchange of two rows or two columns, in which the 2-by-2 block is $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Another row operation of interest replaces two rows having respective j^{th} entries a and b by R linear combinations of them in which a and b are replaced by $c = \text{GCD}(a, b)$ and 0. If $x(ac^{-1}) + y(bc^{-1}) = 1$, then the 2-by-2 block is $\left(\begin{smallmatrix} x & y \\ -bc^{-1} & ac^{-1} \end{smallmatrix}\right)$. A similar operation is possible with columns.

The reduction involves an induction that successively constructs the entries $D_{11}, D_{22}, \dots, D_{ll}$, stopping when the part of C involving rows and columns numbered $\ge l+1$ has been replaced by 0. We start by interchanging rows and columns to move a nonzero entry into position $(1, 1)$. By a succession of row

operations as in the previous paragraph, we can reduce the entry in position (1, 1) to the greatest common divisor of the entries of C in the first column, while reducing the remaining entries of the first column to 0. Next we do the same thing with column operations, reducing the entry in position (1, 1) to the greatest common divisor of the members of the first row, while reducing the remaining entries of the first row to 0. Then we go back and repeat the process with row operations and with column operations as many times as necessary until all the entries of the first row and column other than the one in position (1, 1) are 0. We need to check that this process indeed terminates at some point. If the entries that appear in position (1, 1) as the iterations proceed are $c_1, c_2, c_3, \dots$, then we have $(c_1) \subseteq (c_2) \subseteq (c_3) \subseteq \cdots$. The union of these ideals is an ideal, necessarily a principal ideal of the form (c), and c occurs in one of the ideals in the union; the chain of ideals must be constant after that stage. Once the corner entry becomes constant, the matrices $\left(\begin{smallmatrix} x & y \\ -bc^{-1} & ac^{-1} \end{smallmatrix}\right)$ for the row operations can be chosen to be of the form $\left(\begin{smallmatrix} 1 & 0 \\ -ba^{-1} & 1 \end{smallmatrix}\right)$, and the result is that the row operations do not change the entries of the first row. Similar remarks apply to the matrices for the column operations. The upshot is that we can reduce C in this way so that all entries of the first row and column are 0 except the one in position (1, 1). This handles the inductive step, and we can proceed until at some l^{th} stage we have only the 0 matrix to process. □

This completes the proof of Theorem 8.25. In Theorem 4.56, in which we considered the special case of abelian groups, we obtained a better conclusion than in Theorem 8.25c: we showed that the direct sum of cyclic groups could be written as the direct sum of copies of $\mathbb{Z}$ and of cyclic groups of prime-power order, and that in this case the decomposition was unique up to the order of the summands. We shall now obtain a corresponding better conclusion in the setting of Theorem 8.25.

The existence of the decomposition into cyclic modules of a special kind uses a very general form of the Chinese Remainder Theorem, whose classical statement appears as Corollary 1.9. The generalization below makes use of the following operations of addition and multiplication of ideals in a commutative ring with identity: if I and J are ideals, then $I + J$ denotes the set of sums $x + y$ with $x \in I$ and $y \in J$, and IJ denotes the set of all finite sums of products xy with $x \in I$ and $y \in J$; the sets $I + J$ and IJ are ideals.

Theorem 8.27 (Chinese Remainder Theorem). Let R be a commutative ring with identity, and let $I_1, \dots, I_n$ be ideals in R such that $I_i + I_j = R$ whenever $i \neq j$.

(a) If elements $x_1, \dots, x_n$ of R are given, then there exists x in R such that $x \equiv x_j \bmod I_j$, i.e., $x - x_j$ is in I_j, for all j. The element x is unique if

$I_1 \cap \cdots \cap I_n = 0$.

(b) The map $\varphi : R \to \prod_{j=1}^n R/I_j$ given by $\varphi(r) = (\dots, r + I_j, \dots)$ is an onto ring homomorphism, its kernel is $\bigcap_{j=1}^n I_j$, and the homomorphism descends to a ring isomorphism

$$R\Big/ \bigcap_{j=1}^n I_j \cong R/I_1 \times \cdots \times R/I_n.$$

(c) The intersection $\bigcap_{j=1}^n I_j$ and the product $I_1 \cdots I_n$ coincide.

PROOF. For existence in (a) when $n = 1$, we take $x = x_1$. For existence when $n = 2$, the assumption $I_1 + I_2 = R$ implies that there exist $a_1 \in I_1$ and $a_2 \in I_2$ with $a_1 + a_2 = 1$. Given x_1 and x_2, we put $x = x_1a_2 + x_2a_1$, and then $x \equiv x_1a_2 \equiv x_1 \bmod I_1$ and $x \equiv x_2a_1 \equiv x_2 \bmod I_2$.

For general n, the assumption $I_1 + I_j = R$ for $j \geq 2$ implies that there exist $a_j \in I_1$ and $b_j \in I_j$ with $a_j + b_j = 1$. If we expand out the product $1 = \prod_{j=2}^n (a_j + b_j)$, then all terms but one on the right side involve some a_j and are therefore in I_1. That one term is $b_2b_2 \cdots b_n$, and it is in $\bigcap_{j=2}^n I_j$. Thus $I_1 + \bigcap_{j=2}^n I_j = R$. The case $n = 2$, which was proved above, yields an element y_1 in R such that

$$y_1 \equiv 1 \bmod I_1 \qquad \text{and} \qquad y_1 \equiv 0 \bmod \textstyle\bigcap_{j\neq 1} I_j.$$

Repeating this process for index i and using the assumption $I_i + I_j = R$ for $j \neq i$, we obtain an element y_i in R such that

$$y_i \equiv 1 \bmod I_i \qquad \text{and} \qquad y_i \equiv 0 \bmod \textstyle\bigcap_{j\neq i} I_j.$$

If we put $x = x_1y_1 + \cdots + x_ny_n$, then we have $x \equiv x_iy_i \bmod I_i \equiv x_i \bmod I_i$ for each i, and the proof of existence is complete.

For uniqueness in (a), if we have two elements x and x' satisfying the congruences, then their difference $x - x'$ lies in I_j for every j, hence is 0 under the assumption that $I_1 \cap \cdots \cap I_n = 0$.

In (b), the map φ is certainly a ring homomorphism. The existence result in (a) shows that φ is onto, and the proof of the uniqueness result identifies the kernel. The isomorphism follows.

For (c), consider the special case that I and J are ideals with $I + J = R$. Certainly $IJ \subseteq I \cap J$. For the reverse inclusion, choose $x \in I$ and $y \in J$ with $x + y = 1$; this is possible since $I + J = R$. If z is in $I \cap J$, then $z = zx + zy$ with zx in JI and zy in IJ. Thus z is exhibited as in IJ.

Consequently $I_1I_2 = I_1 \cap I_2$. Suppose inductively that $I_1 \cdots I_k = I_1 \cap \cdots \cap I_k$. We saw in the proof of (a) that $I_{k+1} + \bigcap_{j\neq k+1} I_j = R$, and thus we certainly have $I_{k+1} + \bigcap_{j=1}^k I_j = R$. The special case in the previous paragraph, in combination with the inductive hypothesis, shows that $I_{k+1}I_1 \cdots I_k = I_{k+1} \cdot \big(\bigcap_{j=1}^k I_j\big) = \bigcap_{j=1}^{k+1} I_j$. This completes the induction and the proof. □

Corollary 8.28. Let R be a principal ideal domain, and let $a = \varepsilon p_1^{k_1} \cdots p_n^{k_n}$ be a factorization of a nonzero nonunit element a into the product of a unit and powers of nonassociate primes. Then there is a ring isomorphism

$$R/(a) \cong R/(p_1^{k_1}) \times \cdots \times R/(p_n^{k_n}).$$

PROOF. Let $I_j = (p_j^{k_j})$ in Theorem 8.27. For $i \neq j$, we have $\text{GCD}(p_i^{k_i}, p_j^{k_j}) = 1$. Since R is a principal ideal domain, there exist a and b in R with $ap_i^{k_i} + bp_j^{k_j} = 1$, and consequently $(p_i^{k_i}) + (p_j^{k_j}) = R$. The theorem applies, and the corollary follows. □

Corollary 8.29. If R is a principal ideal domain, then any finitely generated unital R module M is the direct sum of a nonunique free R submodule $\bigoplus_{i=1}^{s} R$ of a well-defined finite rank $s \geq 0$ and the R submodule T of all members m of M such that $rm = 0$ for some $r \neq 0$ in R. In turn, the R submodule T is isomorphic to a direct sum

$$T \cong \bigoplus_{j=1}^{n} R/(p_j^{k_j}),$$

where the p_j are primes in R and the ideals $(p_j^{k_j})$ are not necessarily distinct. The number of summands (p^k) for each class of associate primes p and each positive integer k is uniquely determined by M.

REMARK. As mentioned with Theorem 8.25, some people use the name "Fundamental Theorem of Finitely Generated Modules" to refer to Corollary 8.29 rather than to Theorem 8.25.

PROOF. Theorem 8.25c gives $M = F \oplus \bigoplus_{j=1}^{n} Ra_j$, where F is a free R submodule of some finite rank s and the a_j's are nonzero members of M that are each annihilated by some nonzero member of R. The set T of all m with $rm = 0$ for some $r \neq 0$ in R is exactly $\bigoplus_{j=1}^{n} Ra_j$. Then F is R isomorphic to M/T, hence is isomorphic to the same free R module independently of what direct-sum decomposition of M is used. By Theorem 8.25a, s is well defined.

The cyclic R module Ra_j is isomorphic to $R/(b_j)$, where (b_j) is the ideal of all elements r in R with $ra_j = 0$. The ideal (b_j) is nonzero by assumption and is not all of R since the element $r = 1$ has $1a_j = a_j \neq 0$. Applying Corollary 8.28 for each j and adding the results, we obtain $T \cong \bigoplus_{i=1}^{n} R/(p_i^{k_i})$ for suitable primes p_i and powers k_i. The isomorphism in Corollary 8.28 is given as a ring isomorphism, and we are reinterpreting it as an R isomorphism. The primes p_i that arise for fixed (b_j) are distinct, but there may be repetitions in the pairs (p_i, k_i) as j varies. This proves existence of the decomposition.

If p is a prime in R, then the elements m of T such that $p^k m = 0$ for some k are the ones corresponding to the sum of the terms in $\bigoplus_{j=1}^{n} R/(p_j^{k_j})$ in which p_j is an associate of p. Thus, to complete the proof, it is enough to show that the R isomorphism class of the R module

$$N = R/(p^{l_1}) \oplus \cdots \oplus R/(p^{l_m})$$

with p fixed and with $0 < l_1 \le \cdots \le l_m$ completely determines the integers $l_1, \ldots, l_m$.

For any unital R module L, we can form the sequence of R submodules $p^j L$. The element p carries $p^j L$ into $p^{j+1}L$, and thus each $p^j L/p^{j+1}L$ is an R module on which p acts as 0. Consequently each $p^j L/p^{j+1}L$ is an $R/(p)$ module. Corollary 8.16 and Proposition 8.10 together show that $R/(p)$ is a field, and therefore we can regard each $p^j L/p^{j+1}L$ as an $R/(p)$ vector space.

We shall show that the dimensions $\dim_{R/(p)}(p^j N/p^{j+1}N)$ of these vector spaces determine the integers $l_1, \ldots, l_m$. We start from

$$p^j N = p^j R/(p^{l_1}) \oplus \cdots \oplus p^j R/(p^{l_m}).$$

The term $p^j R/(p^{l_k})$ is 0 if $j \ge l_k$. Thus

$$p^j N = \bigoplus_{j<l_k} p^j R/(p^{l_k}) = \bigoplus_{j<l_k} p^j R/p^{l_k} R.$$

Similarly

$$p^{j+1} N = \bigoplus_{j<l_k} p^{j+1} R/(p^{l_k}) = \bigoplus_{j<l_k} p^{j+1} R/p^{l_k} R.$$

Proposition 8.5 and Theorem 8.3 give us the R isomorphisms

$$p^j N/p^{j+1}N \cong \bigoplus_{j<l_k} \left(p^j R/p^{l_k} R\right)\Big/\left(p^{j+1} R/p^{l_k} R\right) \cong \bigoplus_{j<l_k} p^j R/p^{j+1} R,$$

and these must descend to $R/(p)$ isomorphisms. Consequently

$$\dim_{R/(p)}(p^j N/p^{j+1}N) = \#\{k \mid l_k > j\} \dim_{R/(p)}(p^j R/p^{j+1}R).$$

The coset $p^j + p^{j+1}R$ of $p^j R/p^{j+1}R$ has the property that multiplication by arbitrary elements of R yields all of $p^j R/p^{j+1}R$. Therefore $\dim_{R/(p)}(p^j R/p^{j+1}R) = 1$, and we obtain

$$\dim_{R/(p)}(p^j N/p^{j+1}N) = \#\{k \mid l_k > j\}.$$

Thus the R module N determines the integers on the right side, and these determine the number of l_k's equal to each positive integer j. This proves uniqueness. □

Let us apply Theorem 8.25 and Corollary 8.29 to the principal ideal domain $R = \mathbb{K}[X]$, where $\mathbb{K}$ is a field. The particular unital module of interest is a finite-dimensional vector space V over $\mathbb{K}$, and the scalar multiplication by $\mathbb{K}[X]$ is given by $A(X)v = A(L)(v)$ for each polynomial $A(X)$, where L is a fixed linear map $L : V \to V$. Let us see that the results of this section recover the structure theory of L as developed in Chapter V.

Since V is finite-dimensional over $\mathbb{K}$, V is certainly finitely generated over $R = \mathbb{K}[X]$. Theorem 8.25 gives

$$V \cong R/(A_1(X)) \oplus \cdots \oplus R/(A_n(X)) \oplus R \oplus \cdots \oplus R$$

as R modules and in particular as vector spaces over $\mathbb{K}$. Each summand R is infinite-dimensional as a vector space, and consequently no summand R can be present. Corollary 8.29 refines the decomposition to the form

$$V \cong R/(P_1(X)^{k_1}) \oplus \cdots \oplus R/(P_m(X)^{k_m})$$

as R modules, the polynomials $P_j(X)$ being prime but not necessarily distinct. Since the R isomorphism is in particular an isomorphism of $\mathbb{K}$ vector spaces, each $R/(P_j(X)^{k_j})$ corresponds to a vector subspace V_j, and $V = V_1 \oplus \cdots \oplus V_m$. Since the R isomorphism respects the action by X, we have $L(V_j) \subseteq V_j$ for each j. Thus the direct sum decompositions of Theorem 8.25 and Corollary 8.29 are yielding a decomposition of V into a direct sum of vector subspaces invariant under L. Since the j^{th} summand is of the form $R/(P_j(X)^{k_j})$, L acts on V_j in a particular way, which we have to analyze.

Let us carry out this analysis in the case that $\mathbb{K}$ is algebraically closed (as for example when $\mathbb{K} = \mathbb{C}$), seeing that each V_j yields a Jordan block of the Jordan canonical form (Theorem 5.20a) of L. For the case of general $\mathbb{K}$, the analysis can be seen to lead to the corresponding more general results that were obtained in Problems 32–40 at the end of Chapter V.

Since $\mathbb{K}$ is algebraically closed, any polynomial in $\mathbb{K}[X]$ of degree ≥ 1 has a root in $\mathbb{K}$ and therefore has a first-degree factor $X - c$. Consequently all primes in $\mathbb{K}[X]$ are of the form $X - c$, up to a scalar factor, with c in $\mathbb{K}$. To understand the action of L on V_j, we are to investigate $\mathbb{K}[X]/((X - c)^k)$.

Suppose that $A(X)$ is in $\mathbb{K}[X]$ and is of degree $n \geq 1$. Expanding the monomials of $A(X)$ by the Binomial Theorem as

$$X^j = ((X - c) + c)^j = \textstyle\sum_{i=0}^{j} \binom{j}{i} c^{j-i}(X - c)^i,$$

we see that $A(X)$ has an expansion as

$$A(X) = a_0 + a_1(X - c) + \cdots + a_n(X - c)^n$$

for suitable coefficients $a_0, \ldots, a_n$ in $\mathbb{K}$. Let the invariant subspace that we are studying be $V_{j_0} \subseteq V$. Since V_{j_0} is isomorphic as an R module to $\mathbb{K}[X]/((X-c)^k)$, $(X-c)^k$ acts on V_{j_0} as 0. So does every higher power of $X-c$, and hence

$$A(X) \qquad \text{acts as} \qquad a_0 + a_1(X-c) + \cdots + a_{k-1}(X-c)^{k-1}.$$

The polynomials on the right, as their coefficients vary, represent distinct cosets of $\mathbb{K}[X]/((X-c)^k)$: in fact, if two were to be in the same coset, we could subtract and see that $(X-c)^k$ could not divide the difference unless it were 0. The distinct cosets match in one-one $\mathbb{K}$ linear fashion with the members of V_{j_0}, and thus $\dim V_{j_0} = k$. Let us write down this match. Let v_0 be the member of V_{j_0} that is to correspond to the coset 1 of $\mathbb{K}[X]/(X-c)^k$. On V_{j_0}, $\mathbb{K}[X]$ is acting with $Xv = L(v)$. We define recursively vectors $v_1, \ldots, v_{k-1}$ of V_{j_0} by

$$\begin{aligned}
v_1 = (L-cI)v_0 = (X-c)v_0 \qquad &\longleftrightarrow\ (X-c)\cdot 1 = X-c,\\
v_2 = (L-cI)v_1 = (X-c)v_1 \qquad &\longleftrightarrow\ (X-c)\cdot(X-c) = (X-c)^2,\\
&\ \ \vdots\\
v_{k-1} = (L-cI)v_{k-2} = (X-c)v_{k-2} \qquad &\longleftrightarrow\ (X-c)\cdot(X-c)^{k-2} = (X-c)^{k-1},\\
(L-cI)v_{k-1} = (X-c)v_{k-1} \qquad &\longleftrightarrow\ (X-c)\cdot(X-c)^{k-1} = (X-c)^k \equiv 0.
\end{aligned}$$

We conclude from this correspondence that the vectors $v_0, v_1, \ldots, v_{k-1}$ form a basis of V_{j_0} and that the matrix of $L-cI$ in the ordered basis $v_{k-1}, \ldots, v_1, v_0$ is

$$\begin{pmatrix}
0 & 1 & 0 & 0 & \cdots & 0 & 0\\
 & 0 & 1 & 0 & \cdots & 0 & 0\\
 & & 0 & 1 & \cdots & 0 & 0\\
 & & & \ddots & \ddots & \vdots & \vdots\\
 & & & & & 0 & 1\\
 & & & & & & 0
\end{pmatrix}.$$

Hence the matrix of L in the same ordered basis is

$$\begin{pmatrix}
c & 1 & 0 & 0 & \cdots & 0 & 0\\
 & c & 1 & 0 & \cdots & 0 & 0\\
 & & c & 1 & \cdots & 0 & 0\\
 & & & \ddots & \ddots & \vdots & \vdots\\
 & & & & & c & 1\\
 & & & & & & c
\end{pmatrix},$$

i.e., is a Jordan block. Thus Theorem 8.25 and Corollary 8.29 indeed establish the existence of Jordan canonical form (Theorem 5.20a) when $\mathbb{K}$ is algebraically closed. It is easy to check that Corollary 8.29 establishes also the uniqueness statement in Theorem 5.20a.

7. Orientation for Algebraic Number Theory and Algebraic Geometry

The remainder of the chapter introduces material on commutative rings with identity that is foundational for both algebraic number theory and algebraic geometry. Historically algebraic number theory grew out of Diophantine equations, particularly from two problems—from Fermat's Last Theorem and from representation of integers by binary quadratic forms. Algebraic geometry grew out of studying the geometry of solutions of equations and out of studying Riemann surfaces. Algebraic geometry and algebraic number theory are treated in more detail in *Advanced Algebra*.

These two subjects can be studied on their own, but they also have a great deal in common. The discovery that the plane could be coordinatized and that geometry could be approached through algebra was one of the great advances of all time for mathematics. Since then, fundamental connections between algebraic number theory and algebraic geometry have been discovered at a deeper level, and the distinction between the two subjects is more and more just a question of one's point of view. The emphasis in the remainder of this chapter will be on one aspect of this relationship, the theory that emerged from trying to salvage something in the way of unique factorization.

By way of illustration, let us examine an analogy between what happens with a certain ring of "algebraic integers" and what happens with a certain "algebraic curve." The ring of algebraic integers in question was introduced already in Section 4. It is $R = \mathbb{Z}[\sqrt{-5}] = \mathbb{Z} + \mathbb{Z}\sqrt{-5}$. The units are ± 1. Our investigation of unique factorization was aided by the function

$$N(a + b\sqrt{-5}) = (a + b\sqrt{-5})(a - b\sqrt{-5}) = a^2 + 5b^2,$$

which has the property that

$$N\big((a + b\sqrt{-5})(c + d\sqrt{-5})\big) = N(a + b\sqrt{-5})N(c + d\sqrt{-5}).$$

With this function we could determine candidates for factors of particular elements. In connection with the equality $2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$, we saw that the two factors on the left side and the two factors on the right side are all irreducible. Moreover, neither factor on the left is the product of a unit and a factor on the right. Therefore R is not a unique factorization domain. As a consequence it cannot be a principal ideal domain. In fact, $(2, 1 + \sqrt{-5})$ is an example of an ideal that is not principal. We shall return shortly to examine this ring further.

Now we introduce the algebraic curve. Consider $y^2 = (x - 1)x(x + 1)$ as an equation in two variables x and y. To fix the ideas, we think of a solution as a pair (x, y) of complex numbers. Although the variables in this discussion are

complex, it is convenient to be able to draw pictures of the solutions, and one does this by showing only the solutions (x, y) with x and y in $\mathbb{R}$. Figure 8.6 indicates the set of solutions in $\mathbb{R}^2$ for this particular curve. We can study these solutions for a while, looking for those pairs (x, y) with x and y rationals or integers, but a different level of understanding comes from studying functions on the locus of complex solutions. The functions of interest are polynomial functions in the pair (x, y), and we identify two of them if they agree on the locus. Thus we introduce the ring

$$R' = \mathbb{C}[x, y]/(y^2 - (x-1)x(x+1)).$$

There is a bit of a question whether this is indeed the space of restrictions, but that question is settled affirmatively by the "Nullstellensatz" in Section VII.1 of *Advanced Algebra* and a verification that the principal ideal $(y^2-(x-1)x(x+1))$ is prime.[8] The ring R' is called the "affine coordinate ring" of the curve, and the curve itself is an example of an "affine algebraic curve."

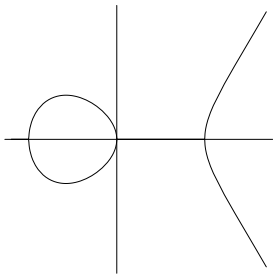


FIGURE 8.6. Real points of the curve $y^2 = (x-1)x(x+1)$.

We can recover the locus of the curve from the ring R' as follows. If (x_0, y_0) is a point of the curve, then it is meaningful to evaluate members of R' at (x_0, y_0), and we let $I_{(x_0,y_0)}$ be the ideal of all members of R' vanishing at (x_0, y_0). Evaluation at (x_0, y_0) exhibits the ring $R'/I_{(x_0,y_0)}$ as isomorphic to $\mathbb{C}$, which is a field. Thus $I_{(x_0,y_0)}$ is a maximal ideal and is in particular prime. It turns out for this example that all nonzero prime ideals are of this form.[9] We return to make use of this geometric interpretation of prime ideals in a moment.

[8]The polynomial $y^2 - (x-1)x(x+1)$ is prime since $(x-1)x(x+1)$ is not a square, or since Eisenstein's criterion applies. The principal ideal $(y^2 - (x-1)x(x+1))$ is therefore prime by Proposition 8.14. What the Nullstellensatz says when the underlying field is algebraically closed is that the only polynomials vanishing on the zero locus of a prime ideal are the members of the ideal.

[9]In Section 9, Example 3 of integral closures in combination with Proposition 8.45 shows that every nonzero prime ideal of R' is maximal. (In algebraic geometry one finds that this property of prime ideals is a reflection of the 1-dimensional nature of the curve.) The Nullstellensatz says that the maximal ideals are all of the form $I_{(x_0,y_0)}$.

Now let us consider factorization in R'. Every element of R' can be written uniquely as $A(x) + B(x)y$, where $A(x)$ and $B(x)$ are polynomials. The analog in R' of the quantity $N(a + b\sqrt{-5})$ in the ring R is the quantity

$$\begin{aligned} N\big(A(x) + B(x)y\big) &= (A(x) + B(x)y)(A(x) - B(x)y) \\ &= A(x)^2 - B(x)^2y^2 \\ &= A(x)^2 - B(x)^2(x^3 - x). \end{aligned}$$

Easy computation shows that

$$N\big((A(x) + B(x)y)(C(x) + D(x)y)\big) = N\big(A(x) + B(x)y\big)N\big(C(x) + D(x)y\big),$$

and hence $N(\cdot)$ gives us a device to use to check whether elements of R' are irreducible. We find in the equation

$$(x + y)(x - y) = x^2 - (x^3 - x) = -x(x - \tfrac{1}{2}(1 + \sqrt{5}))(x - \tfrac{1}{2}(1 - \sqrt{5}))$$

that the two elements on the left side and the three elements on the right side are irreducible. Therefore unique factorization fails in R'.

Although unique factorization fails for the elements of R', there is a notion of factorization for ideals in R' that behaves well algebraically and has a nice geometric interpretation. Recall that the nonzero prime ideals correspond to the points of the locus $y^2 = (x - 1)x(x + 1)$ via passage to the zero locus, the ideal corresponding to (x_0, y_0) being called $I_{(x_0, y_0)}$. For any two ideals I and J, we can form the product ideal IJ whose elements are the sums of products of a member of I and a member of J. Then $I_{(x_0, y_0)}^k$ may be interpreted as the ideal of all members of R' vanishing at (x_0, y_0) to order k or higher, and $I_{(x_1, y_1)}^{k_1} \cdots I_{(x_n, y_n)}^{k_n}$ becomes the ideal of all members of R' vanishing at each (x_j, y_j) to order at least k_j. We shall see in Section 11 that every nonzero proper ideal I in R' factors in this way. The points (x_j, y_j) and the integers k_j have a geometric interpretation in terms of I and are therefore uniquely determined: the (x_j, y_j)'s form the locus of common zeros of the members of I, and the integer k_j is the greatest integer such that the vanishing at (x_j, y_j) is always at least to order k_j. In a sense, factorization of elements was the wrong thing to consider; the right thing to consider is factorization of ideals, which is unique because of the associated geometric interpretation.

Returning to the ring $R = \mathbb{Z}[\sqrt{-5}]$, we can ask whether factorization of ideals is a useful notion in R. Again IJ is to be the set of all sums of products of an element in I and an element in J. For $I = (2, 1 + \sqrt{-5})$ and $J = (2, 1 - \sqrt{-5})$, we get all sums of expressions $(2a + b(1 + \sqrt{-5}))(2c + d(1 - \sqrt{-5}))$ in which a, b, c, d are in $\mathbb{Z}$, hence all sums of expressions

$$2(2ac + 3bd) + 2(bc + ad) + 2\sqrt{-5}(bc - ad).$$

All such elements are divisible by 2. Two examples come by taking $a = c = 1$ and $b = d = 0$ and by taking $a = c = 0$ and $b = d = 1$; these give 4 and 6. Subtracting, we see that 2 is a sum of products. Thus $IJ = (2)$. The element 2 is irreducible and not prime, and we know from Proposition 8.14 that the ideal (2) therefore cannot be prime. What we find is that the ideal (2) factors even though the element 2 does not factor. It turns out that R has unique factorization of ideals, just the way R' does.

The prime ideals of the ring R have a certain amount of structure in terms of the primes or prime ideals of $\mathbb{Z}$. To understand what to expect, let us digress for a moment to discuss what happens with the ring $R'' = \mathbb{Z}[i] = \mathbb{Z} + \mathbb{Z}\sqrt{-1}$ of Gaussian integers. This too was introduced in Section 4, and it is a Euclidean domain, hence a principal ideal domain. It has unique factorization. Its appropriate $N(\cdot)$ function is $N(a+ib) = a^2 + b^2$. Problems 27–31 at the end of the chapter ask one to verify that the primes of R'', up to multiplication by one of the units ± 1 and $\pm i$, are members of R'' of any of the three kinds

$p = 4n + 3$ that is prime in $\mathbb{Z}$ and has $n \geq 0$,
$p = a \pm ib$ with $a^2 + b^2$ prime in $\mathbb{Z}$ of the form $4n + 1$ with $n \geq 0$,
$p = 1 \pm i$ (these are associates).

These three kinds may be distinguished by what happens to the function $N(\cdot)$. In the first case $N(p) = p^2$ is the square of a prime of $\mathbb{Z}$ and is the square of a prime of R'', in the second case $N(p)$ is a prime of $\mathbb{Z}$ that is the product of two distinct primes of R'', and in the third case $N(p)$ is a prime of $\mathbb{Z}$ that is the square of a prime of R'', apart from a unit factor. The nonzero prime ideals of R'' are the principal ideals generated by the prime elements of R'', and they fall into three types as well. Each nonzero prime ideal P has a prime p of $\mathbb{Z}$ attached to it, namely the one with $(p) = \mathbb{Z} \cap P$, and the type of the ideal corresponds to the nature of the factorization of the ideal pR'' of R''. Specifically in the first case pR'' is a prime ideal in R'', in the second case pR'' is the product of two distinct prime ideals in R'', and in the third case pR'' is the square of a prime ideal in R''.

The structure of the prime ideals in R is of the same nature as with R''. Each nonzero prime ideal P has a prime p of $\mathbb{Z}$ attached to it, again given by $(p) = \mathbb{Z} \cap P$, and the three kinds correspond to the factorization of the ideal pR of R. Let us be content to give examples of the three possible behaviors:

$11R$ is prime in R,
$2R$ is the product of two distinct prime ideals in R,
$5R$ is the square of the prime ideal $(\sqrt{-5})$ in R.

We have already seen the decomposition of $2R$, and the decomposition of $5R$ is easy to check. With $11R$, the idea is to show that 11 is a prime element in R. Thus let 11 divide a product in R. Then $N(11) = 11^2$ divides the product of

the $N(\cdot)$'s, 11 divides the product of the $N(\cdot)$'s, and 11 must divide one of the $N(\cdot)$'s. Say that 11 divides $N(a+b\sqrt{-5})$, i.e., that $a^2+5b^2 \equiv 0 \bmod 11$. If 11 divides one of a or b, then this congruence shows that 11 divides the other of them; then 11 divides $a+b\sqrt{-5}$, as we wanted to show. The other possibility is that 11 divides neither a nor b. Then $(ab^{-1})^2 \equiv -5 \bmod 11$ says that -5 is a square modulo 11, and we readily check that it is not. The conclusion is that 11 is indeed prime in R.

This structure for the prime ideals of R has an analog with the curve and its ring R'. The analogs for the curve case of $\mathbb{Z}$ and $\sqrt{-5}$ for the number-theoretic case are $\mathbb{C}[x]$ and y. The primes of $\mathbb{C}[x]$ are nonzero scalars times polynomials $x-c$ with c complex, and the relevant question for R' is how the ideal $(x-c)R'$ decomposes into prime ideals. We can think about this problem algebraically or geometrically. Algebraically, the ideal of all polynomials vanishing at (x_0, y_0) is $I_{(x_0,y_0)} = (x-x_0, y-y_0)$, the set of all $(x-x_0)A(x)-y_0B(x)+yB(x)$ with $A(x)$ and $B(x)$ in $\mathbb{C}[x]$. The intersection with $\mathbb{C}[x]$ consists of all $(x-x_0)A(x)$ and is therefore the principal ideal $(x-x_0)$. We want to factor the ideal $(x-x_0)R'$.

If we pause for a moment and think about the problem geometrically, the answer is fairly clear. Ideals correspond to zero loci with multiplicities. The question is the factorization of the ideal of all polynomials vanishing when $x = x_0$. For most values of the complex number x_0, there are two choices of the complex y such that (x_0, y) is on the locus since y is given by a quadratic equation, namely $y^2 = (x_0-1)x_0(x_0+1)$. Thus for most values of x_0, $(x-x_0)R'$ is the product of two distinct prime ideals. The geometry thus suggests that

$$(x-x_0)R' = (x-x_0, y-y_0)(x-x_0, y+y_0),$$

where $y_0^2 = (x_0-1)x_0(x_0+1)$ and it is assumed that $y_0 \neq 0$. We can verify this algebraically: The members of the product ideal are the polynomials

$$\begin{aligned}&\big((x-x_0)A(x)+(y-y_0)B(x)\big)\big((x-x_0)C(x)+(y+y_0)D(x)\big)\\ &\quad=(x-x_0)^2A(x)C(x)+(x-x_0)\big(A(x)(y+y_0)D(x))+C(x)(y-y_0)B(x)\big)\\ &\qquad+(y^2-y_0^2)B(x)D(x).\end{aligned}$$

The last term on the right side is $\big((x^3-x)-(x_0^3-x_0)\big)B(x)D(x)$ and is divisible by $x-x_0$. Therefore every member of the product ideal lies in the principal ideal $(x-x_0)$. On the other hand, the product ideal contains $(x-x_0)(x-x_0)$ and also $(y^2-y_0^2) = (x^3-x_0^3)-(x-x_0) = (x-x_0)(x^2+xx_0+x_0^2)$. Since $\mathrm{GCD}\big((x-x_0), (x^2+xx_0+x_0^2)\big) = 1$, the product ideal contains $x-x_0$. Therefore the product ideal equals $(x-x_0)$.

The exceptional values of x_0 are $-1, 0, +1$, where the locus has $y_0 = 0$. The geometry of the factorization is not so clear in this case, but the algebraic

computation remains valid. Thus we have $(x - x_0)R' = (x - x_0, y)^2$ if x_0 equals -1, 0, or $+1$. The conclusion is that the nonzero prime ideals of R' are of two types, with $(x - x_0)R'$ equal to

the product of two distinct prime ideals in R' if x_0 is not in $\{-1, 0, +1\}$,

the square of a prime ideal in R' if x_0 is in $\{-1, 0, +1\}$.

The third type, with $(x - x_0)R'$ prime in R', does not arise. Toward the end of Chapter IX we shall see how we could have anticipated the absence of the third type.

That is enough of a comparison for now. Certain structural results useful in both algebraic number theory and algebraic geometry are needed even before we get started at factoring ideals, and those are some of the topics for the remainder of this chapter. In Section 11 we conclude by establishing unique factorization of ideals for a class of examples that includes the examples above. In the examples above, the rings we considered were $\mathbb{Z}[X]/(X^2+5) = \mathbb{Z}[\sqrt{-5}\,]$ and $\mathbb{C}[x, y]/(y^2 - (x-1)x(x+1)) \cong \mathbb{C}[x][\sqrt{(x-1)x(x+1)}\,]$. In each case the notation $[\,\cdot\,]$ refers to forming the ring generated by the coefficients and the expression or expressions in brackets.

First we establish a result saying that ideals in the rings of interest are not too wild. For example, in algebraic geometry, one wants to consider the set of restrictions of the members of $\mathbb{K}[X_1, \ldots, X_n]$, $\mathbb{K}$ being a field, to the locus of common zeros of a set of polynomials. The general tool will tell us that any ideal in $\mathbb{K}[X_1, \ldots, X_n]$ is finitely generated; thus a description of what polynomials vanish on the locus under study is not completely out of the question. The tool is the Hilbert Basis Theorem and is the main result of Section 8.

Second we need a way of understanding, in a more general setting, the relationship that we used in the above examples between $\mathbb{Z}$ and $\mathbb{Z}[\sqrt{-5}\,]$, and between $\mathbb{C}[x]$ and $\mathbb{C}[x][\sqrt{(x-1)x(x+1)}\,]$. The tool is the notion of integral closure and is the subject of Section 9.

Third we need a way of isolating the behavior of prime ideals, of eliminating the influence of algebraic or geometric factors that have nothing to do with the prime ideal under study. The tool is the notion of localization and is the subject of Section 10.

In Section 11 we make use of these three tools to establish unique factorization of ideals for a class of integral domains known as "Dedekind domains." It is easy to see that principal ideal domains are Dedekind domains, and we shall show that many other integral domains, including the examples above, are Dedekind domains. A refined theorem producing Dedekind domains will be obtained toward the end of Chapter IX once we have introduced the notion of a "separable" extension of fields.

8. Noetherian Rings and the Hilbert Basis Theorem

In this section, R will be a commutative ring with identity, and all R modules will be assumed unital. We begin by introducing three equivalent conditions on a unital R module.

Proposition 8.30. If R is a commutative ring with identity and M is a unital R module, then the following conditions on R submodules of M are equivalent:

(a) (**ascending chain condition**) every strictly ascending chain of R submodules $M_1 \subsetneqq M_2 \subsetneqq \cdots$ terminates in finitely many steps,
(b) (**maximum condition**) every nonempty collection of R submodules has a maximal element under inclusion,
(c) (**finite basis condition**) every R submodule is finitely generated.

PROOF. To see that (a) implies (b), let $\mathcal{C}$ be a nonempty collection of R submodules of M. Take M_1 in $\mathcal{C}$. If M_1 is not maximal, choose M_2 in $\mathcal{C}$ properly containing M_1. If M_2 is not maximal, choose M_3 in $\mathcal{C}$ properly containing M_2. Continue in this way. By (a), this process must terminate, and then we have found a maximal R submodule in $\mathcal{C}$.

To see that (b) implies (c), let N be an R submodule of M, and let $\mathcal{C}$ be the collection of all finitely generated R submodules of N. This collection is nonempty since 0 is in it. By (b), $\mathcal{C}$ has a maximal element, say N'. If x is in N but x is not in N', then $N' + Rx$ is a finitely generated R submodule of N that properly contains N' and therefore gives a contradiction. We conclude that $N' = N$, and therefore N is finitely generated.

To see that (c) implies (a), let $M_1 \subsetneqq M_2 \subsetneqq \cdots$ be given, and put $N = \bigcup_{n=1}^{\infty} M_n$. By (c), N is finitely generated. Since the M_n are increasing with n, we can find some M_{n_0} containing all the generators. Then the sequence stops no later than at M_{n_0}. □

Let us apply Proposition 8.30 with M taken to be the unital R module R. As always, the R submodules of R are the ideals of R.

Corollary 8.31. If R is a commutative ring with identity, then the following conditions on R are equivalent:

(a) ascending chain condition for ideals: every strictly ascending chain of ideals in R is finite,
(b) maximum condition for ideals of R: every nonempty collection of ideals in R has a maximal element under inclusion,
(c) finite basis condition for ideals: every ideal in R is finitely generated.

The corollary follows immediately from Proposition 8.30. A commutative ring with identity satisfying the equivalent conditions of Corollary 8.31 is said to be a **Noetherian** commutative ring.

EXAMPLES.

(1) Principal ideal domains, such as $\mathbb{Z}$ and $\mathbb{K}[X]$ when $\mathbb{K}$ is a field. The finite basis condition for ideals is satisfied since every ideal is singly generated. The fact that (c) implies (a) has already been proved manually for principal ideal domains twice in this chapter—once in the proof of (UFD1) for a principal ideal domain in Theorem 8.15 and once in the proof of Lemma 8.26.

(2) Any homomorphic image R' of a Noetherian commutative ring R, provided 1 maps to 1. In fact, if $I' \subseteq R'$ is an ideal, its inverse image I is an ideal in R; the image of a finite set of generators of I is a finite set of generators of I'.

(3) $\mathbb{K}[X_1, \ldots, X_n]$ when $\mathbb{K}$ is a field. This commutative ring is Noetherian by application of the Hilbert Basis Theorem (Theorem 8.32 below) and induction on n. This ring is also a unique factorization domain, as we saw in Section 5.

(4) $\mathbb{Z}[X]$. This commutative ring is Noetherian, also by the Hilbert Basis Theorem below. Example 2 shows therefore that the quotient $\mathbb{Z}[\sqrt{-5}] = \mathbb{Z}[X]/(X^2 + 5)$ is Noetherian. This ring is an integral domain, and we have seen that it is not a unique factorization domain.

Theorem 8.32 (Hilbert Basis Theorem). If R is a nonzero Noetherian commutative ring, then so is $R[X]$.

PROOF. If I is an ideal in $R[X]$ and if $k \geq 0$ is an integer, let $L_k(I)$ be the union of $\{0\}$ and the set of all nonzero elements of R that appear as the coefficient of X^k in some element of degree k in I. First let us see that $\{L_k(I)\}_{k\geq 0}$ is an increasing sequence of ideals in R. In fact, if $A(X)$ and $B(X)$ are polynomials of degree k in I with leading terms $a_k X^k$ and $b_k X^k$, then $A(X) + B(X)$ has degree k if $b_k \neq -a_k$, and hence $a_k + b_k$ is in $L_k(I)$ in every case. Similarly if r is in R and $ra_k \neq 0$, then $rA(X)$ has degree k, and hence ra_k is in $L_k(I)$ in every case. Consequently $L_k(I)$ is an ideal in R. Since I is closed under multiplication by X, $L_k(I) \subseteq L_{k+1}(I)$ for all $k \geq 0$.

Next let us prove that if J is any ideal in $R[X]$ such that $I \subseteq J$ and $L_k(I) = L_k(J)$ for all $k \geq 0$, then $I = J$. Let $B(X)$ be in J with $\deg B(X) = k$. Arguing by contradiction, we may suppose that $B(X)$ is not in I and that k is the smallest possible degree of a polynomial in J but not in I. Since $L_k(I) = L_k(J)$, we can find $A(X)$ in I whose leading term is the same as the leading term of $B(X)$. Since $B(X)$ is not in I, $B(X) - A(X)$ is not in I. Since $I \subseteq J$, $B(X) - A(X)$ is in J. Since $\deg(B(X) - A(X)) \leq k - 1$, we have arrived at a contradiction to the defining property of k. We conclude that $I = J$.

Now let $\{I_j\}_{j\geq 0}$ be an ascending chain of ideals in $R[X]$, and form $L_i(I_j)$ for each i. When i or j is fixed, these ideals are increasing as a function of the other index, j or i. By the maximum condition in R, $L_i(I_j) \subseteq L_p(I_q)$ for some p and q and all i and j. For $i \geq p$ and $j \geq q$, we have $L_i(I_j) \supseteq L_p(I_q)$ and thus $L_i(I_j) = L_p(I_q)$. The case $j = q$ gives $L_p(I_q) = L_i(I_q)$, and therefore $L_i(I_j) = L_i(I_q)$ for $i \geq p$ and $j \geq q$. For any fixed i, the ascending chain condition on ideals gives $L_i(I_j) = L_i(I_{n(i)})$ for $j \geq n(i)$, and the above argument shows that we may take $n(i) = q$ if $i \geq p$. Hence $n(i)$ may be taken to be bounded in i, say by n_0, and $L_i(I_j) = L_i(I_{n_0})$ for all $i \geq 0$ and $j \geq n_0$. By the result of the previous paragraph, $I_j = I_{n_0}$ for $j \geq n_0$, and hence the ascending chain condition has been verified for ideals in $R[X]$. □

Proposition 8.33. In a Noetherian integral domain R, every nonzero nonunit is a product of irreducible elements.

REMARK. The proof below gives an alternative argument for (UFD1) in Theorem 8.15, an argument that does not so explicitly use the full force of Zorn's Lemma.

PROOF. Let a_1 be a nonzero nonunit of R. If a_1 is not irreducible, then a_1 has a factorization $a_1 = a_2b_2$ in which neither a_2 nor b_2 is a unit. If a_2 is not irreducible, then a_2 has a factorization $a_2 = a_3b_3$ in which neither a_3 nor b_3 is a unit. We continue in this way as long as it is possible to do so. Let us see that this process cannot continue indefinitely. Assume the contrary. The equality $a_1 = a_2b_2$ with b_2 not a unit says that the inclusion of ideals $(a_1) \subseteq (a_1, a_2)$ is proper. Arguing in this way with a_2, a_3, and so on, we obtain

$$(a_1) \subsetneqq (a_1, a_2) \subsetneqq (a_1, a_2, a_3) \subsetneqq \cdots,$$

in contradiction to the ascending chain condition for ideals. Because of this contradiction we conclude that for some n, a_n does not have any decomposition $a_n = a_{n+1}b_{n+1}$ with b_{n+1} a nonunit. Hence a_n is irreducible. The upshot is that our original element a_1 has an irreducible factor, say c_1.

Write $a_1 = c_1d_2$. If d_2 is not a unit, repeat the process with it, obtaining $d_2 = c_2d_3$ with c_2 irreducible. If d_3 is not a unit, we can again repeat this process. This process cannot continue indefinitely because otherwise we would have a strictly increasing sequence of ideals

$$(c_1) \subsetneqq (c_1, c_2) \subsetneqq (c_1, c_2, c_3) \subsetneqq \cdots,$$

in contradiction to the ascending chain condition for ideals. Thus for some n, we have $a_1 = c_1c_2 \dots c_nd_{n+1}$ with $c_1, \dots, c_n$ irreducible and with d_{n+1} equal to a unit. Grouping c_n and d_{n+1} as a single irreducible factor, we obtain the desired factorization of the given element a_1. □

Proposition 8.34. If R is a Noetherian commutative ring, then any R submodule of a finitely generated unital R module is finitely generated.

REMARK. The proof follows the lines of the argument for Proposition 8.24.

PROOF. Let M be a unital finitely generated R module with a set $\{m_1, \dots, m_n\}$ of n generators, and define $M_k = Rm_1 + \cdots + Rm_k$ for $1 \le k \le n$. Then $M_n = M$ since M is unital. We shall prove by induction on k that every R submodule of M_k is finitely generated. The case $k = n$ then gives the proposition. For $k = 1$, suppose that S is an R submodule of $M_1 = Rm_1$. Let I be the subset of all r in R with rm_1 in S. Since S is an R submodule, I is an ideal in R, necessarily finitely generated since R is Noetherian. Let $I = (r_1, \dots, r_l)$. Then $S = Im_1 = Rr_1m_1 + Rr_2m_1 + \cdots + Rr_lm_1$, and the elements $r_1m_1, r_2m_1, \dots, r_lm_1$ form a finite set of generators of S.

Assume inductively that every R submodule of M_k is known to be finitely generated, and let N_{k+1} be an R submodule of M_{k+1}. Let $q : M_{k+1} \to M_{k+1}/M_k$ be the quotient R homomorphism, and let φ be the restriction $q\big|_{N_{k+1}}$, mapping N_{k+1} into M_{k+1}/M_k. Then $\ker \varphi = N_{k+1} \cap M_k$ is an R submodule of M_k and is finitely generated by the inductive hypothesis. Also, image φ is an R submodule of M_{k+1}/M_k, which is singly generated with generator equal to the coset of m_{k+1}. Since an R submodule of a singly generated unital R module was shown in the previous paragraph to be finitely generated, image φ is finitely generated. Applying Lemma 8.23 to φ, we see that N_{k+1} is finitely generated. This completes the induction and the proof. □

9. Integral Closure

In this section, we let R be an integral domain, F be its field of fractions, and K be a any field containing F. Sometimes we shall assume also that $\dim_F K$ is finite. The main cases of interest are as follows.

EXAMPLES OF GREATEST INTEREST.

(1) $R = \mathbb{Z}$, $F = \mathbb{Q}$, and $\dim_F K < \infty$. In Chapter IX we shall see in this case from the "Theorem of the Primitive Element" that K is necessarily of the form $\mathbb{Q}[\theta]$ as already described in Section 1 and in Chapter IV. This is the setting we used in Section 7 as orientation for certain problems in algebraic number theory.

(2) $R = \mathbb{K}[X]$ for a field $\mathbb{K}$, $F = \mathbb{K}(X)$ is the field of fractions of R, and K is a field containing F with $\dim_F K < \infty$. In the special case $\mathbb{K} = \mathbb{C}$, this is the setting we used in Section 7 as orientation for treating curves in algebraic geometry.

Proposition 8.35. Let R be an integral domain, F be its field of fractions, and K be any field containing F. Then the following conditions on an element x of K are equivalent:

(a) x is a root of a monic polynomial in $R[X]$,
(b) the subring $R[x]$ of K generated by R and x is a finitely generated R module,
(c) there exists a finitely generated nonzero unital R module $M \subseteq K$ such that $xM \subseteq M$.

REMARK. When the equivalent conditions of the proposition are satisfied, we say that x is **integral** over R or x is **integrally dependent** on R. In this terminology, in Section VII.5 and in Section 1 of the present chapter, we defined an **algebraic integer** to be any member of $\mathbb{C}$ that is integral over $\mathbb{Z}$. The equivalence of (a) and (c) in this setting allowed us to prove that the set of algebraic integers is a subring of $\mathbb{C}$.

PROOF. If (a) holds, we can write $x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$ for suitable coefficients in R. Solving for x^n and substituting, we see that the subring $R[x]$, which equals $R + Rx + Rx^2 + \cdots$, is actually given by $R[x] = R + Rx + \cdots + Rx^{n-1}$. Therefore $R[x]$ is a finitely generated R module, and (b) holds.

If (b) holds, then we can take $M = R[x]$ to see that (c) holds.

If (c) holds, let $m_1, \ldots, m_k$ be generators of M as an R module. Then we can find members a_{ij} of R for which

$$\begin{aligned} xm_1 &= a_{11}m_1 + \cdots + a_{1k}m_k, \\ &\vdots \\ xm_k &= a_{k1}m_1 + \cdots + a_{kk}m_k. \end{aligned}$$

This set of equations, regarded as a single matrix equation over K, becomes

$$\begin{pmatrix} x-a_{11} & -a_{12} & \cdots & -a_{1k} \\ -a_{21} & x-a_{22} & \cdots & -a_{2k} \\ & & \ddots & \\ -a_{k1} & -a_{k2} & \cdots & x-a_{kk} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The k-by-k matrix on the left is therefore not invertible, and its determinant, which is a member of the field K, must be 0. Expanding the determinant and replacing x by an indeterminate X, we obtain a monic polynomial of degree k in $R[X]$ for which x is a root. Thus (a) holds. □

If R, F, and K are as above, the **integral closure** of R in K is the set of all members of K that are integral over R. In Corollary 8.38 we shall prove that the integral closure of R in K is a subring of K.

EXAMPLES OF INTEGRAL CLOSURES.

(1) The integral closure of $\mathbb{Z}$ in $\mathbb{Q}$ is $\mathbb{Z}$ itself. This fact amounts to the statement that a rational root of a monic polynomial with integer coefficients is an integer; this was proved[10] in the course of Lemma 7.30. Recall the argument: If $x = p/q$ is a rational number in lowest terms that satisfies $x^n + a_{n-1}x^{n-1} + \cdots + a_1 a + a_0 = 0$, then we clear fractions and obtain $p^n + a_{n-1}p^{n-1}q + \cdots + a_1 pq^{n-1} + a_0 q^n = 0$. Examining divisibility by q, we see that q divides p^n. Hence any prime factor of q divides p and shows that p/q cannot be in lowest terms. Therefore q has no prime factors, and p/q is an integer.

(2) Let us determine the integral closure of $\mathbb{Z}$ in $\mathbb{Q}(\sqrt{m}\,)$, where m is a square-free integer other than 0 or 1. The result is going to be that the integral closure consists of all $a + b\sqrt{m}$ with

$$a \text{ and } b \begin{cases} \text{both in } \mathbb{Z} & \text{if } m \not\equiv 1 \bmod 4, \\ \text{both in } \mathbb{Z} \text{ or both in } \mathbb{Z} + \frac{1}{2} & \text{if } m \equiv 1 \bmod 4. \end{cases}$$

In other words, the integral closure is

$$\begin{cases} \mathbb{Z}[\sqrt{m}\,] & \text{if } m \not\equiv 1 \bmod 4, \\ \mathbb{Z}[\frac{1}{2}(1 + \sqrt{m}\,)] & \text{if } m \equiv 1 \bmod 4. \end{cases} \tag{$*$}$$

In fact, consider the polynomial

$$P(X) = X^2 - 2aX + (a^2 - mb^2),$$

whose roots are exactly $a \pm b\sqrt{m}$. If a and b are in $\mathbb{Z}$, then $P(X)$ has coefficients in $\mathbb{Z}$, and hence both of $a \pm b\sqrt{m}$ are in the integral closure. If $m \equiv 1 \bmod 4$ and a and b are both in $\mathbb{Z} + \frac{1}{2}$, write $a = c/2$ and $b = d/2$ with c and d in $2\mathbb{Z} + 1$. Since $a^2 - mb^2 = \frac{1}{4}(c^2 - md^2)$, we have

$$c^2 - md^2 \equiv c^2 - d^2 \bmod 4 \equiv 1 - 1 \bmod 4 \equiv 0 \bmod 4,$$

and therefore $\frac{1}{4}(c^2 - md^2) = a^2 - mb^2$ is in $\mathbb{Z}$. Consequently the polynomial $P(X)$ exhibits $a + b\sqrt{m}$ as in the integral closure.

For the reverse inclusion, suppose that $z = a + b\sqrt{m}$ is in the integral closure and is not in $\mathbb{Z}$. Then z is a root of some monic polynomial $A(X)$ in $\mathbb{Z}[X]$. In addition, z is a root of $P(X)$ above, and $P(X)$ is a monic prime polynomial in $\mathbb{Q}[X]$ because it has no rational first-degree factor. Writing $A(X) = B(X)P(X) + R(X)$ in $\mathbb{Q}[X]$ with $R(X) = 0$ or $\deg R(X) < \deg P(X) = 2$ and

[10] It is not assumed that the reader has looked at Chapter VII. A result that implies Lemma 7.30 will be obtained below as Corollary 8.38, which makes no use of material from Chapter VII.

substituting z for X, we see that $R(z) = 0$, and we conclude that $R(X) = 0$. Thus $P(X)$ divides $A(X)$. By Corollary 8.20c, $P(X)$ is in $\mathbb{Z}[X]$. Hence $2a$ and $a^2 - mb^2$ are in $\mathbb{Z}$. One case is that a is in $\mathbb{Z}$, and then mb^2 is in $\mathbb{Z}$; since m is square free, there are no candidates for primes dividing the denominator of b, and so b is in $\mathbb{Z}$. The other case is that a is in $\mathbb{Z} + \frac{1}{2}$, and then mb^2 is in $\mathbb{Z} + \frac{1}{4}$. So $m(2b)^2$ is in $4\mathbb{Z} + 1$. Since m is square free, there are no candidates for primes dividing the denominator of $2b$, and $2b$ is an integer. Since $m(2b)^2$ is in $4\mathbb{Z} + 1$, $m \equiv 1 \bmod 4$ and $2b \equiv 1 \bmod 2$ are forced. This completes the proof that the integral closure is given by $(*)$.

(3) Under the assumption that the characteristic of the field $\mathbb{K}$ is not 2, let us determine the integral closure T of $R = \mathbb{K}[x]$ in $K = \mathbb{K}(x)[\sqrt{P(x)}\,] = \mathbb{K}(x)[y]/(y^2 - P(x))$, where $P(X)$ is a square-free polynomial in $\mathbb{K}[x]$. Parenthetically we need to check that K is a field. Since $\mathbb{K}(x)$ is a field, $\mathbb{K}(x)[y]$ is a principal ideal domain, and the question is whether $(y^2 - P(x))$ is a prime (= maximal) ideal. We have only to observe that $y^2 - P(x)$ is irreducible because $P(x)$ is not a square, and then it follows that K is a field. Thus the situation for this example fits the setting of Proposition 8.35 with $R = \mathbb{K}[x]$, $F = \mathbb{K}(x)$, and $K = F(y)/(y^2 - P)$. We are going to show that the integral closure T of R in K consists of all $A(x) + B(x)\sqrt{P(x)}$ with $A(x)$ and $B(x)$ both in $R = \mathbb{K}[x]$. It follows that the integral closure will be

$$T = \mathbb{K}[x][\sqrt{P(x)}\,] = \mathbb{K}[x] + \mathbb{K}[x]\sqrt{P(x)}. \tag{$*$}$$

To see this, first let $A(x)$ and $B(x)$ be in $\mathbb{K}[x]$, and consider the monic polynomial

$$Q(y) = y^2 - 2Ay + (A^2 - PB^2) \tag{$**$}$$

in $\mathbb{K}[x][y]$. Its roots in K are exactly $A(x) \pm B(x)\sqrt{P(x)}$, and thus we see that both of $A(x) \pm B(X)P(x)$ are in T. Conversely let $z = A(x) + B(x)\sqrt{P(x)}$ be in T but not R. Here $A(x)$ and $B(x)$ are in $\mathbb{K}(x)$. Then z is a root in K of some monic polynomial $M(y)$ whose coefficients are in $\mathbb{K}[x]$. In addition, z is a root of the member $Q(y)$ of $\mathbb{K}(x)[y]$ defined in $(**)$. The division algorithm gives $M(y) = N(y)Q(y) + W(y)$ in $\mathbb{K}(x)[y]$ with $W = 0$ or $\deg W < \deg Q = 2$. Substituting $z \in T$ for y, we obtain

$$0 = M(z) = N(z)Q(z) + W(z) = N(z)0 + W(z).$$

Thus $W(z) = 0$. If $\deg W = 1$, then z is in F, and the same argument as in Example 1 shows that z is in R; since we are assuming that z is not in R, we conclude that $W = 0$. Therefore $Q(y)$ divides $M(y)$. By Corollary 8.20c, $M(y)$ is in $\mathbb{K}[x][y]$. Hence $2A$ and $A^2 - PB^2$ are in $\mathbb{K}[x]$. Since the characteristic of $\mathbb{K}$ is not 2, A is in $\mathbb{K}[x]$. Then PB^2 is in $\mathbb{K}[x]$, and B must be in $\mathbb{K}[x]$ since P is square free. Thus T is given as in $(*)$.

From these examples we can extract a rough description of the situation that will interest us. We start with a ring R such as $\mathbb{Z}$ or $\mathbb{K}[x]$, along with its field of fractions F. We assume that the integral closure of R in F is R itself, as is the case with $\mathbb{Z}$ in $\mathbb{Q}$ and as we shall see is the case with $\mathbb{K}[x]$ in $\mathbb{K}(x)$. Let K be a field containing F with $\dim_F K < \infty$. We are interested in an analog T of integral elements relative to K, and what works as T is the integral closure of R in K.

Lemma 8.36. If A, B, and C are integral domains with $A \subseteq B \subseteq C$ such that C is a finitely generated B module and B is a finitely generated A module, then C is a finitely generated A module.

PROOF. Let C be generated over B by $c_1, \dots, c_r$, and let B be generated over A by $b_1, \dots, b_s$. Then C is generated over A by the sr elements $b_j c_i$ for $1 \le i \le r$ and $1 \le j \le s$. □

Proposition 8.37. Let R be an integral domain, F be its field of fractions, and K be any field containing F. If $x_1, \dots, x_r$ are members of K integral over R, then the subring $R[x_1, \dots, x_r]$ of K generated by R and $x_1, \dots, x_r$ is a finitely generated R module.

REMARKS. The ring $R[x_1, \dots, x_r]$ is certainly finitely generated over R as a ring. The proposition asserts more—that it is finitely generated as an R module. This means that all products of powers of the x_j's are in the R linear span of finitely many of them.

PROOF. We induct on r. Since x_1 is assumed integral over R, the case $r = 1$ follows from Proposition 8.35b. For the inductive step, suppose that $R[x_1, \dots, x_s]$ is a finitely generated R module. Since x_{s+1} is integral over R, it is certainly integral over $R[x_1, \dots, x_s]$. Thus Proposition 8.35b shows that $R[x_1, \dots, x_{s+1}]$ is a finitely generated $R[x_1, \dots, x_s]$ module. Taking $A = R$, $B = R[x_1, \dots, x_s]$, and $C = R[x_1, \dots, x_{s+1}]$ in Lemma 8.36, we see that $R[x_1, \dots, x_{s+1}]$ is a finitely generated R module. □

Corollary 8.38. Let R be an integral domain, F be its field of fractions, and K be any field containing F. Then the integral closure of R in K is a subring of K.

REMARK. A special case of this corollary appears in somewhat different language as Lemma 7.30.

PROOF. Let x and y be integral over R. Then $R[x, y]$ is a finitely generated R module by Proposition 8.37. We have $(x \pm y)R[x, y] \subseteq R[x, y]$ and $(xy)R[x, y] \subseteq R[x, y]$. Taking $M = R[x, y]$ in Proposition 8.35c and using the

implication that (c) implies (a) in that proposition, we see that $x \pm y$ and xy are integral over R. □

Corollary 8.39. Let A, B, and C be integral domains with $A \subseteq B \subseteq C$. If every member of B is integral over A and if every member of C is integral over B, then every member of C is integral over A.

PROOF. Let K be the field of fractions of C, and regard C as a subring of K. If x is in C, then x is a root of some monic polynomial with coefficients in B, say $x^n + b_{n-1}x^{n-1} + \cdots + b_0 = 0$. By Proposition 8.37 the subring $D = A[b_{n-1}, \ldots, b_0]$ of C is a finitely generated A module. Since x is integral over D, $D[x]$ is a finitely generated D module, by a second application of Proposition 8.37. Lemma 8.36 shows that $D[x]$ is a finitely generated A module. By Proposition 8.35, x is integral over A. □

We say that the integral domain R is **integrally closed** if R equals its integral closure in its field of fractions. Example 1 above in essence observed that the ring $\mathbb{Z}$ of integers is integrally closed. Example 2 above showed, for the case $m = -3$, that the integral closure of $\mathbb{Z}$ in $\mathbb{Q}[\sqrt{-3}\,]$ is something other than the ring $\mathbb{Z}[\sqrt{-3}\,]$; consequently $\mathbb{Z}[\sqrt{-3}\,]$ cannot be integrally closed. A more direct argument is to observe that the element $x = \frac{1}{2}(-1 + \sqrt{-3}\,)$ of $\mathbb{Q}[\sqrt{-3}\,]$ satisfies $x^2 + x + 1 = 0$ but is not in $\mathbb{Z}[\sqrt{-3}\,]$.

Corollary 8.40. Let R be an integral domain, F be its field of fractions, and K be any field containing F. Then the integral closure T of R in K is integrally closed.

PROOF. Corollary 8.38 shows that T is a subring of K. Let C be the integral closure of T in K. We apply Corollary 8.39 to the integral domains $R \subseteq T \subseteq C$. The corollary says that every member of C is integral over R, and hence $C \subseteq T$. That is, $C = T$. Let $\eta : T \to L$ be the one-one homomorphism of T into its field of fractions, and let $\varphi : T \to K$ be the inclusion. By Proposition 8.6, there exists a unique ring homomorphism $\widetilde{\varphi} : L \to K$ such that $\varphi = \widetilde{\varphi}\eta$. Identifying L with $\widetilde{\varphi}(L) \subseteq K$, we can treat L as a subfield of K containing T. Since the only elements of K integral over T have been shown to be the members of T, the only elements of the subfield L integral over T are the members of T. Therefore T is integrally closed. □

Proposition 8.41. If R is a unique factorization domain, then R is integrally closed.

PROOF. Suppose that $y^{-1}x$ is a member of the field of fractions F of R, with x and y in R and $y \neq 0$, and suppose that $y^{-1}x$ satisfies the equation

$$(y^{-1}x)^n + a_{n-1}(y^{-1}x)^{n-1} + \cdots + a_1(y^{-1}x) + a_0 = 0$$

with coefficients in R. Clearing fractions and moving x^n over to one side by itself, we have

$$x^n = -y(a_{n-1}x^{n-1} + \cdots + a_1xy^{n-2} + a_0y^{n-1}).$$

If a prime p in R divides y, then it divides x^n and must divide x. If R is a unique factorization domain, this says that we cannot arrange for GCD(x, y) to equal 1 unless no prime divides y. In this case, y is a unit in R. Consequently $y^{-1}x$ is in R. □

Since $\mathbb{Z}$ is a unique factorization domain, Proposition 8.41 gives a new proof that $\mathbb{Z}$ is integrally closed. We see also that $\mathbb{K}[x]$ is integrally closed when $\mathbb{K}$ is a field.

We saw above that the ring $\mathbb{Z}[\sqrt{-3}\,]$ is not integrally closed; consequently it cannot be a unique factorization domain. Another way of drawing this conclusion is to verify in the equality $(1 + \sqrt{-3}\,)(1 - \sqrt{-3}\,) = 2 \cdot 2$ that the two elements on the left are irreducible and are not associates of the irreducible element 2 on the right.

A more significant example, taking advantage of the contrapositive of Proposition 8.41, is that any polynomial ring $\mathbb{K}[X_1, \ldots, X_n]$ over a field $\mathbb{K}$ is integrally closed. In fact, we know from Section 5 that $\mathbb{K}[X_1, \ldots, X_n]$ has unique factorization.

Proposition 8.42. Let R be an integral domain, F be its field of fractions, and K be any field containing F. If $\dim_F K < \infty$, then any x in K has the property that there is some $c \neq 0$ in R such that cx is integral over R.

REMARKS. Consequently K may be regarded as the field of fractions of the integral closure T of R in K. In fact, let $\{x_i\}$ be a basis of K over F, and choose $c_i \neq 0$ in R for each i such that $y_i = c_ix_i$ is integral over R. Then $\{y_i\}$ is a basis for K over F consisting of members of T, and it follows that every member of K is the quotient of a member of T by a member of R. Proposition 8.6 supplies a one-one ring homomorphism of the field of fractions for T into K, and the description just given for the elements of K shows that this homomorphism is onto K. Therefore K may be regarded as the field of fractions of T.

PROOF. Since $\dim_F K < \infty$, the elements $1, x, x^2, \ldots$ of K are linearly dependent over F. Therefore $a_nx^n + \cdots + a_1x + a_0 = 0$ for a suitable n and for suitable members of F with $a_n \neq 0$. Clearing fractions, we may assume that $a_n, \ldots, a_1, a_0$ are in R and that $a_n \neq 0$. Multiplying the equation by a_n^{n-1}, we obtain

$$(a_nx)^n + a_{n-1}(a_nx)^{n-1} + \cdots + a_1a_n^{n-2}(a_nx) + a_0a_n^{n-1} = 0.$$

Thus we can take $c = a_n$. □

In the base rings $\mathbb{Z}$ and $\mathbb{K}[x]$ of our examples, every nonzero prime ideal is maximal because the rings are principal ideal domains. In Section 7 we mentioned that every nonzero prime ideal in $\mathbb{Z}[\sqrt{-5}\,]$ is maximal even though $\mathbb{Z}[\sqrt{-5}\,]$ is not a principal ideal domain. The remainder of this section, particularly Proposition 8.45, shows that the feature that every nonzero prime ideal is maximal is always preserved in our passage from R to T.

Proposition 8.43. Let R be an integral domain, F be its field of fractions, K be any field containing F, and T be the integral closure of R in K. If Q is a nonzero prime ideal of T, then $P = R \cap Q$ is a nonzero prime ideal of R.

REMARKS. Corollary 8.38 shows that T is a ring. A construction for prime ideals that goes in the reverse direction, from R to T, appears below as Proposition 8.53.

PROOF. Let Q be a nonzero prime ideal of T, and put $P = R \cap Q$. The ideal P is proper since 1 is not in Q and cannot be in P. It is prime since $xy \in P$ implies that xy is in Q, x or y is in Q, and x or y is in $R \cap Q = P$. To see that P is nonzero, take $t \neq 0$ in Q. Since t is integral over R, t satisfies some monic polynomial equation $t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 = 0$ with coefficients in R. Without loss of generality, $a_0 \neq 0$ since otherwise we could divide the equation by a positive power of t. Then $a_0 = t(-t^{n-1} - a_{n-1}t^{n-2} - \cdots - a_1)$ exhibits a_0 as in Q as well as in R. Thus P is nonzero. □

Lemma 8.44. Let R and T be integral domains with $R \subseteq T$ and with every element of T integral over R. If T' is an integral domain and $\varphi : T \to T'$ is a homomorphism of rings onto T', then every member of T' is integral over $\varphi(R)$.

PROOF. If t is in T, then t satisfies some monic polynomial equation of the form $t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 = 0$ with coefficients in R. Applying φ to this equation, we see that $\varphi(t)$ satisfies a monic polynomial equation with coefficients in $\varphi(R)$. □

Proposition 8.45. Let R be an integral domain, F be its field of fractions, K be any field containing F, and T be the integral closure of R in K. If every nonzero prime ideal of R is maximal, then every nonzero prime ideal of T is maximal.

REMARK. As with Proposition 8.43, Corollary 8.38 shows that T is a ring.

PROOF. Let Q be a nonzero prime ideal in T, and let $P = R \cap Q$. Since P is a nonzero prime ideal of R by Proposition 8.43, the hypotheses say that P is maximal in R. We shall apply Lemma 8.44 to the quotient homomorphism $T \to T/Q$. The lemma says that every element of the integral domain T/Q is

integral over the subring $(R+Q)/Q$. Composing the inclusion homomorphism $R \to T$ with the homomorphism $T \to T/Q$ yields a ring homomorphism $R \to T/Q$ that carries P into the 0 coset. Since $P = R \cap Q$, this ring homomorphism descends to a one-one ring homomorphism $R/P \to T/Q$. The Second Isomorphism Theorem (for abelian groups) identifies the image of R/P with $(R+Q)/Q$. Since P is maximal as an ideal in R, R/P is a field. The ring isomorphism $R/P \cong (R+Q)/Q$ thus shows that every element of T/Q is integral over a field.

Let us write k for this field isomorphic to R/P, and let k' be the field of fractions of T/Q. We can now argue as in the proof of Proposition 4.1. If $x \neq 0$ is in T/Q, then x satisfies a monic polynomial equation $x^m + c_{m-1}x^{m-1} + \cdots + c_1 x + c_0 = 0$ with coefficients in k, and we may assume that $c_0 \neq 0$. Then the equality $x^{-1} = -c_0^{-1}(c_1 + \cdots + a_{m-1}x^{m-2} + x^{m-1})$ shows that the member x^{-1} of k' is in fact in T/Q. Therefore T/Q is a field, and the ideal Q is maximal in T. □

10. Localization and Local Rings

In this section, R denotes a commutative ring with identity. The objective is to enlarge or at least adjust R so as to make further elements of R become invertible under multiplication. The prototype is the construction of the field of fractions for an integral domain. A subset S of R is called a **multiplicative system** if 1 is in S and if the product of any two members of S is in S. The multiplicative system will be used as a set of new allowable denominators, and the new ring will be denoted[11] by $S^{-1}R$.

The construction proceeds along the same lines as in Section 2, except that some care is needed to take into account the possibility of zero divisors in R and even in S. We begin with an intermediate set

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

and impose the relation $(r, s) \sim (r', s')$ if $t(rs' - sr') = 0$ for some $t \in S$. To check transitivity, suppose that $(r, s) \sim (r', s')$ and $(r', s') \sim (r'', s'')$. Then we have $t(rs' - sr') = 0$ and $t'(r's'' - s'r'') = 0$ for some t and t' in S, and hence

$$s'tt'(rs'' - sr'') = s''t'\big(t(rs' - sr')\big) + st\big(t'(r's'' - s'r'')\big) = 0.$$

Since $s'tt'$ is in S, $(r, s) \sim (r'', s'')$. Thus $\sim$ is an equivalence relation.

[11] Some authors write R_S instead of $S^{-1}R$.

The set of equivalence classes is denoted by $S^{-1}R$ and is called the **localization**[12] **of R with respect to S**. Addition and multiplication are defined in $\widetilde{R}$ by $(r, s) + (r', s') = (rs' + sr', ss')$ and $(r, s)(r', s') = (rr', ss')$. Simple variants of the arguments in Section 2 show that these operations descend to operations on $S^{-1}R$. For example, with addition let (r, s), (r', s'), and (r'', s'') be in $\widetilde{R}$ with $(r', s') \sim (r'', s'')$, i.e., with $t'(r's'' - s'r'') = 0$ for some $t' \in S$. Then the equivalence

$$(r, s) + (r', s') = (rs' + sr', ss') \sim (rs'' + sr'', ss'') = (r, s) + (r'', s'')$$

holds because

$$t'\big((rs' + sr')ss'' - (rs'' + sr'')ss'\big) = s^2t'(r's'' - s'r'') = 0.$$

Similarly multiplication is well defined.

The result is that $S^{-1}R$ is a commutative ring with identity and that the mapping $r \mapsto r^*$, where r^* is the class of $(r, 1)$, is a ring homomorphism of R into $S^{-1}R$ carrying 1 to 1. Let us observe the following simple properties of $S^{-1}R$:

(i) $S^{-1}R = 0$ if and only if 0 is in S, since $S^{-1}R = 0$ if and only if $(1, 1) \sim (0, 1)$, if and only if $t(1 \cdot 1 - 1 \cdot 0) = 0$ for some $t \in S$.

(ii) $r \mapsto r^*$ is one-one if and only if S contains no zero divisors, since $r^* = 0$ if and only if $(r, 1) \sim (0, 1)$, if and only if $tr = 0$ for some $t \in S$.

(iii) s^* is a unit in $S^{-1}R$ for each $s \in S$, since the class of $(1, s)$ is a multiplicative inverse for s^*.

(iv) every member of $S^{-1}R$ is of the form $(s^*)^{-1}r^*$ for some $r \in R$ and $s \in S$, since $(r, s) = (r, 1)(1, s)$ is the class of $r^*(s^*)^{-1}$.

(v) $S^{-1}R$ is an integral domain if R is an integral domain and 0 is not in S.

In working with localizations, we shall normally drop the superscript * on the image r^* in $S^{-1}R$ of an element r of R.

Localizations arise in algebraic number theory and in algebraic geometry. In applications to algebraic number theory, the ring R typically is an integral domain, and therefore the map $r \mapsto r^*$ is one-one. In applications to algebraic geometry, S may have zero divisors.

EXAMPLES OF LOCALIZATIONS.

(1) R is arbitrary, and $S = \{1\}$. Then $S^{-1}R = R$.

[12]Some authors use a term like "ring of fractions" or "ring of quotients" in connection with localization in the general case or in some special cases. We shall not use these terms. In any event, "ring of quotients" is emphatically not to be confused with "quotient ring" as in Chapter IV, which is the coset space of a ring modulo an ideal.

(2) R is arbitrary, and $S = \{\text{nonzero elements that are not zero divisors in } R\}$. Then every nonzero element of $S^{-1}R$ is a zero divisor or is a unit. In this example when S consists of all members of R other than 0, then R is an integral domain and $S^{-1}R$ is the field of fractions of R.

(3) R is arbitrary, P is a prime ideal in R, and S is the set-theoretic complement of P. The identity is in S since P is proper. The prime nature of P is used in checking that S is a multiplicative system: if s and t are in S, then neither is in P, by definition, and their product st cannot be in P since P is prime; thus the product st is in S. With these definitions,

$$S^{-1}R \quad \text{is often denoted by } R_P$$

and is called the **localization of R at the prime P**. In practice this is the most important example of a localization,[13] directly generalizing the construction of the field of fractions of an integral domain as the localization at the prime ideal 0. Here are some special cases, $\mathbb{K}$ being a field in the cases in which it occurs:

(a) When $R = \mathbb{Z}$ and $P = (p)$ for a prime number p, the set S consists of nonzero integers not divisible by p, and R_P is the subset of all members of $\mathbb{Q}$ whose denominators are not divisible by p.

(b) When $R = \mathbb{K}[X]$ and $P = (X-c)$, the set S consists of all polynomials that are nonvanishing at c, and R_P is the set of formal rational expressions in X that are finite at c.

(c) When $R = \mathbb{K}[X, Y]$ and $P = (X - c, Y - d)$, the set S consists of all polynomials in X and Y that are nonvanishing at (c, d), and R_P is the set of formal rational expressions in X and Y that are finite at (c, d).

(d) When $R = \mathbb{K}[X, Y]$ and $P = (X)$, the set S consists of all polynomials in X and Y that are not divisible by X, and R_P is the set of formal rational expressions in X and Y that are meaningful as rational expressions in Y when X is set equal to 0. For example, $1/(X + Y)$ is in R_P, but $1/X$ is not.

(4) R is arbitrary, $\{P_\alpha\}$ is a nonempty collection of prime ideals, and S is the set of all elements of R that lie in none of the ideals P_α. Then $S^{-1}R$ may be regarded as the localization of R at the set of all primes P_α.

(5) R is arbitrary, u is an element of R, and $S = \{1, u, u^2, \dots\}$. For example, if $R = \mathbb{Z}/(p^2)$, where p is a prime, and if $u = p$, then 0 is in S, and observation (i) shows that $S^{-1}R = 0$.

(6) R is a Noetherian integral domain, E is an arbitrary set of nonzero elements of R, and S is the set of all finite products of members of E, including the element

[13] Beware of confusing R_P with R/P. The ring R_P is obtained by suitably enlarging R, at least in the case that R is an integral domain, whereas the ring R/P is obtained by suitably factoring something out from R.

1 as the empty product. Let us see that the same $S^{-1}R$ results when E is replaced by a certain set E' of units and irreducible elements of R, namely the union of $R^\times$ and the set of all irreducible elements x in R such that x^{-1} is in $S^{-1}R$. Define T to be the set of all finite products of members of E'. We show that $S^{-1}R = T^{-1}R$. If e is in E', then either e is a unit in R, in which case e^{-1} lies in R and therefore also $S^{-1}R$, or e is irreducible in R with e^{-1} in $S^{-1}R$. Passing to finite products of members of E', we see that $T^{-1} \subseteq S^{-1}R$. Hence $T^{-1}R \subseteq S^{-1}R$. Now let s be in S, and use Proposition 8.33 to write s as a product of irreducible elements $s = s_1 \cdots s_n$. Then $s_j^{-1} = s^{-1}(s_1 \cdots \widehat{s_j} \cdots s_n)$, with $\widehat{s_j}$ indicating a missing factor. By construction, each s_j is in E'. Therefore each s_j is in T, and s is in T. Consequently $S \subseteq T$, and $S^{-1}R \subseteq T^{-1}R$.

The localization of R at S is characterized up to canonical isomorphism by the same kind of universal mapping property that characterizes the field of fractions of an integral domain. To formulate a proposition, let us write η for the homomorphism $r \mapsto r^*$ of R into $S^{-1}R$. Then the pair $(S^{-1}R, \eta)$ has the universal mapping property stated in Proposition 8.46 and illustrated in Figure 8.7.

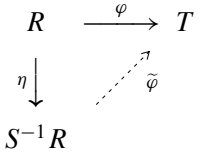


FIGURE 8.7. Universal mapping property of the localization of R at S.

Proposition 8.46. Let R be a commutative ring with identity, let S be a multiplicative system in R, let $S^{-1}R$ be the localization of R at S, and let η be the canonical homomorphism of R into $S^{-1}R$. Whenever φ is a ring homomorphism of R into a commutative ring T with identity such that $\varphi(1) = 1$ and such that $\varphi(s)$ is a unit in T for each $s \in S$, then there exists a unique ring homomorphism $\widetilde{\varphi} : S^{-1}R \to T$ such that $\varphi = \widetilde{\varphi}\eta$.

PROOF. If (r, s) with $s \in S$ is a pair in $\widetilde{R}$, we define $\Phi(r, s) = \varphi(r)\varphi(s)^{-1}$. This is well defined since $\varphi(s)$ is assumed to be a unit in T. Let us see that Φ is consistent with the equivalence relation, i.e., that $(r, s) \sim (r', s')$ implies $\Phi(r, s) = \Phi(r', s')$. Since $(r, s) \sim (r', s')$, we have $u(rs' - r's) = 0$ for some $u \in S$, and therefore also $\varphi(u)(\varphi(r)\varphi(s') - \varphi(r')\varphi(s)) = 0$. Since $\varphi(u)$ is a unit, $\varphi(r)\varphi(s') = \varphi(r')\varphi(s)$. Hence $\Phi(r, s) = \varphi(r)\varphi(s)^{-1} = \varphi(r')\varphi(s')^{-1} = \Phi(r', s')$, as required.

We can thus define $\widetilde{\varphi}$ of the class of (r, s) to be $\Phi(r, s)$, and $\widetilde{\varphi}$ is well defined as a function from $S^{-1}R$ to T. It is a routine matter to check that $\widetilde{\varphi}$ is a ring homomorphism. If r is in R, then $\widetilde{\varphi}(\eta(r)) = \widetilde{\varphi}(\text{class of } (r, 1)) = \Phi(r, 1) =$

$\varphi(r)\varphi(1)^{-1}$, and this equals $\varphi(r)$ since φ is assumed to carry 1 into 1. Therefore $\widetilde{\varphi}\eta = \varphi$.

For uniqueness, observation (iv) shows that the most general element of $S^{-1}R$ is of the form $\eta(r)\eta(s)^{-1}$ with $r \in R$ and $s \in S$. Since $(\widetilde{\varphi}\eta)(r) = \varphi(r)$ and $(\widetilde{\varphi}\eta)(s) = \varphi(s)$, we must have $\widetilde{\varphi}(\eta(r)\eta(s)^{-1}) = \widetilde{\varphi}(\eta(r))\widetilde{\varphi}(\eta(s))^{-1} = \varphi(r)\varphi(s)^{-1}$. Therefore φ uniquely determines $\widetilde{\varphi}$. □

We shall examine the relationship between ideals in R and ideals in the localization $S^{-1}R$. If I is an ideal in R, then $S^{-1}I = \{s^{-1}i \mid s \in S,\ i \in I\}$ is easily checked to be an ideal in $S^{-1}R$ and is called the **extension** of I to $S^{-1}R$. If J is an ideal in $S^{-1}R$, then $R \cap J$, i.e., the inverse image of J under the canonical homomorphism $\eta : R \to S^{-1}R$, is an ideal in R and is called the **contraction** of J.

Proposition 8.47. Let R be a commutative ring with identity, and let $S^{-1}R$ be a localization. If J is an ideal in $S^{-1}R$, then $S^{-1}(R \cap J) = J$. Consequently the mapping $I \mapsto S^{-1}I$ is a one-one mapping of the set of all ideals I in R of the form $I = R \cap J$ onto the set of all ideals in $S^{-1}R$, and this mapping respects intersections and inclusions.

REMARKS. As in the definition of contraction, $R \cap J$ means $\eta^{-1}(J)$, where $\eta : R \to S^{-1}R$ is the canonical homomorphism. The map $I \mapsto S^{-1}I$ that carries arbitrary ideals of R to ideals of $S^{-1}R$ need not be one-one; the localization could for example be the field of fractions of an integral domain and have only trivial ideals. The proposition says that the map $I \mapsto S^{-1}I$ is one-one, however, when restricted to ideals of the form $I = R \cap J$.

PROOF. From the facts that $R \cap J \subseteq J$ and J is an ideal in $S^{-1}R$, we obtain $S^{-1}(R \cap J) \subseteq S^{-1}J \subseteq J$. For the reverse inclusion let x be in J, and write $x = s^{-1}r$ with r in R and s in S. Then $sx = r$ is in $R \cap J$, and therefore x is in $S^{-1}(R \cap J)$.

For the conclusion about the mapping $I \mapsto S^{-1}I$, the mapping is one-one because $S^{-1}(R \cap J_1) = S^{-1}(R \cap J_2)$ implies $J_1 = J_2$ by what we have just shown; hence $R \cap J_1 = R \cap J_2$. The mapping is onto because if J is given, then $J = S^{-1}(R \cap J)$ by what has already been shown. To see that the mapping respects the intersection of ideals, let ideals $R \cap J_\alpha$ be given for α in some nonempty set. Then

$$S^{-1}\Big(\bigcap_\alpha (R \cap J_\alpha)\Big) = S^{-1}(R \cap \bigcap_\alpha J_\alpha) = \bigcap_\alpha J_\alpha = \bigcap_\alpha S^{-1}(R \cap J_\alpha).$$

Finally the fact that the mapping respects the intersection of two ideals implies that it respects inclusions. □

Corollary 8.48. Let R be a commutative ring with identity, and let $S^{-1}R$ be a localization.

(a) If R is Noetherian, then $S^{-1}R$ is Noetherian.

(b) If every nonzero prime ideal in R is maximal, then the same thing is true in $S^{-1}R$.

(c) If R is an integral domain that is integrally closed and if $S^{-1}R$ is not zero, then $S^{-1}R$ is integrally closed.

(d) If I is an ideal in R, then the ideal $S^{-1}I$ of $S^{-1}R$ is proper if and only if $I \cap S = \varnothing$.

PROOF. For (a), let $\{J_\alpha\}$ be a nonempty collection of ideals in $S^{-1}R$. Contraction of ideals is one-one by the first conclusion of Proposition 8.47, and it respects inclusions because it is given by the inverse image of a function. Since R is Noetherian, Corollary 8.31b produces a maximal element $R \cap J$ from among the ideals $R \cap J_\alpha$ of R. The first and second conclusions of Proposition 8.47 together show that $J = S^{-1}(R \cap J) \supseteq S^{-1}(R \cap J_\alpha) = J_\alpha$ for all α. Hence J is maximal among the J_α.

For (b), let J_1 be a nonzero prime ideal in $S^{-1}R$. Arguing by contradiction, suppose that J_2 is an ideal in $S^{-1}R$ with $J_1 \subsetneqq J_2 \subsetneqq S^{-1}R$. Then $R \cap J_1 \subseteq R \cap J_2 \subseteq R$. If either of these inclusions were an equality, then use of the second conclusion of Proposition 8.47 would give a corresponding equality for J_1, J_2, R, and there is no such equality. Hence $R \cap J_1 \subsetneqq R \cap J_2 \subsetneqq R$.

If J_1 is prime in $S^{-1}R$, then $R \cap J_1$ is prime in R: In fact, if a and b are members of R such that ab is in $R \cap J_1$, then ab is in J_1, and either a or b must be in J_1 since J_1 is prime. Since a and b are both in R, one of a and b is in $R \cap J_1$. Thus $R \cap J_1$ is prime.[14]

By assumption for (b), $R \cap J_1$ is then maximal in R, and this conclusion contradicts the fact that $R \cap J_1 \subsetneqq R \cap J_2 \subsetneqq R$. The assumption that J_2 exists has thus led us to a contradiction. Consequently there can be no such J_2, and J_1 is a maximal ideal in $S^{-1}R$.

For (c), let F be the field of fractions of R, so that $R \subseteq S^{-1}R \subseteq F$. The field of fractions of $S^{-1}R$ is the field F as a consequence of Proposition 8.6. If x is a member of F that is integral over $S^{-1}R$ and if x satisfies $x^n + b_{n-1}x^{n-1} + \cdots + b_0 = 0$ with coefficients in $S^{-1}R$, then we can find a common element s of S and rewrite this equation as

$$x^n + (s^{-1}a_{n-1})x^{n-1} + \cdots + (s^{-1}a_0) = 0$$

with $a_{n-1}, \ldots, a_0$ in R. Multiplying by s^n, we obtain

$$(sx)^n + a_{n-1}(sx)^{n-1} + \cdots + a_1 s^{n-2}(sx) + a_0 s^{n-1} = 0.$$

[14] Problem 9 at the end of the chapter puts this argument in a broader context.

Therefore sx is integral over R. Since R is integrally closed, sx is in R. Write $r = sx$. Then $x = s^{-1}r$ with r in R and s in S. Hence x is exhibited as in $S^{-1}R$, and we conclude that $S^{-1}R$ is integrally closed.

For (d), suppose that $I \cap S$ is nonempty. If s is in $I \cap S$, then $1 = s^{-1}s$ is in $S^{-1}I$ and the ideal $S^{-1}I$ equals $S^{-1}R$. Conversely if $S^{-1}I = S^{-1}R$, then 1 is in $S^{-1}I = \{s^{-1}i \mid s \in S,\ i \in I\}$, and hence $1 = s^{-1}i$ for some s and i; consequently $I \cap S$ contains the element $i = s$. □

A **local ring** is a commutative ring with identity having a unique maximal ideal. An equivalent definition is given in Proposition 8.49 below, and then it follows that the localization $S^{-1}R$ of Example 2 earlier in this section is a local ring. Corollary 8.50 below will produce a more useful example: localization with respect to a prime ideal, as in Example 3 earlier, always yields a local ring.[15]

Proposition 8.49. A nonzero commutative ring R with identity is a local ring if and only if the nonunits of R form an ideal.

REMARK. The zero ring is not local, having no proper ideals, and its set of nonunits is empty, hence is not an ideal.

PROOF. If the nonunits of R form an ideal, then that ideal is a unique maximal ideal since a proper ideal cannot contain a unit; hence R is local. Conversely suppose that R is local and that M is the unique maximal ideal. If x is any nonunit, then the principal ideal (x) is a proper ideal since 1 is not of the form xr. By Proposition 8.8, (x) is contained in some maximal ideal, and we must have $(x) \subseteq M$ since M is the unique maximal ideal. Then x is in M, and we conclude that every nonunit is contained in M. □

Corollary 8.50. Let R be an integral domain, let P be a prime ideal of R, let S be the set-theoretic complement of P, and let $R_P = S^{-1}R$ be the localization of R at P. Then R_P is a local ring, its unique maximal ideal is $M = S^{-1}P$, and P can be recovered from M as $P = R \cap M$. If Q is any prime ideal of R that is not contained in P, then $S^{-1}Q = S^{-1}R$.

PROOF. The subset $S^{-1}P$ of $S^{-1}R$ is an ideal by Proposition 8.47, and Corollary 8.48d shows that it is proper. Every member of $S^{-1}R$ that is not in $S^{-1}P$ is of the form $s'^{-1}s$ with s and s' in S and hence is a unit. Since no unit lies in any proper ideal, $S^{-1}R$ has $M = S^{-1}P$ as its unique maximal ideal, and $S^{-1}R$ is local by Proposition 8.49.

[15]For Example 3 with $R = \mathbb{K}[X]$ and $P = (X - c)$, the sense in which the ring R_P is "local" has a geometric interpretation: the only spot in $\mathbb{K}$ where we can regard members of R_P as $\mathbb{K}$-valued functions is "near" the point c, with "near" depending on the element of R_P. See the discussion after the proof of Corollary 8.50 below.

The contraction $R \cap M$ consists of all elements in R of the form $s^{-1}p$ with s in S and p in P. Let us see that the contraction equals P. Certainly $R \cap M \supseteq P$. For the reverse inclusion the equation $s^{-1}p = r$ says that $p = rs$. If r is not in P, then the facts that s is not in P and P is prime imply that $p = rs$ is not in P, contradiction. Thus r is in P, and we conclude that P can be recovered from M as $P = R \cap M$.

If Q is any prime ideal of R that is not contained in P, then $S^{-1}Q = S^{-1}R$. In fact, any element q of Q that is not in P is in S; therefore 1 is in the ideal $S^{-1}Q$, and $S^{-1}Q = S^{-1}R$. □

The construction of R_P in the corollary reduces to the construction of the field of fractions of R if $P = 0$. Other interesting and typical cases occur for suitable nonzero P's when $R = \mathbb{K}[X, Y]$, $\mathbb{K}$ being a field. One such prime ideal is $P = (X - c, Y - d)$; then, as was mentioned in connection with Example 3 above, the localization of R at P consists of the rational expressions $f(X, Y)$ that are well defined at (c, d). The maximal ideal in this case consists of all such rational expressions that are 0 at (c, d). Another example of a nonzero prime ideal in $R = \mathbb{K}[X, Y]$ is $P = (X)$; then the localization of R at P consists of the rational expressions $f(X, Y)$ whose denominators are not divisible by X, and the maximal ideal consists of all such rational expressions $f(X, Y)$ whose numerators are divisible by X if f is written in lowest terms.

A number-theoretic analog of the localizations of the previous paragraph is the localization of $R = \mathbb{Z}$ at (p), where p is a prime number. The discussion with Example 3 above mentioned that the localization consists of all members of $\mathbb{Q}$ with no factor of p in the denominator. In this case the maximal ideal consists of those rationals q whose numerators are divisible by p if q is written in lowest terms.

We conclude this section with introductory remarks about a product operation on ideals. Let R be a nonzero commutative ring with identity. If I and J are ideals in R, then once again IJ denotes[16] the set of all sums of products of a member of I by a member of J. Certainly IJ is closed under addition and negatives, and the fact that $r(IJ) = (rI)J \subseteq IJ$ for $r \in R$ shows that IJ is an ideal. Localization with respect to a prime ideal is a handy tool for extracting information about products of ideals. We illustrate with Propositions 8.52 and 8.53 below. The first of these will play an important role in Section 11.

[16]Sometimes, such as in the equality $S^{-1}S^{-1} = S^{-1}$, the product notation is meant to refer only to the set of all products, not to all sums of products. With ideals we are to allow sums of products. The applicable convention will normally be clear from the context, but we shall be explicit when there might be a possibility of confusion.

Lemma 8.51 (Nakayama's Lemma). Let R be a commutative ring with identity, let I be an ideal of R contained in all maximal ideals, and let M be a finitely generated unital R module. If $IM = M$, then $M = 0$.

REMARK. Here IM means the set of sums of products of a member of I by a member of M. The lemma applies to no ideals if $R = 0$.

PROOF. We induct on the number of generators of M. If M is singly generated, say by a generator m, then the hypothesis $IM = M$ implies that $rm = m$ for some r in I. Thus $(1 - r)m = 0$. If $1 - r$ is a unit, then we can multiply by its inverse and obtain $m = 0$; we conclude that $M = 0$. If $1 - r$ is not a unit, then it lies in some maximal ideal P, by application of Proposition 8.8 to the proper principal ideal $(1 - r)$. Since r lies in P by hypothesis, 1 lies in P, and we have a contradiction to the fact that P is proper.

Suppose that the lemma holds for $n - 1$ or fewer generators, and let M be generated by $m_1, \dots, m_n$. Since $IM = M$, we have $\sum_{j=1}^{n} r_j m_j = m_1$ for suitable $r_1, \dots, r_n$ in I. Then $(1 - r_1)m_1 = \sum_{j=2}^{n} r_j m_j$. If $1 - r_1$ is a unit, then we can multiply by its inverse and see that the generator m_1 is unnecessary; we conclude that $M = 0$ by induction. If $1 - r_1$ is not a unit, then it lies in some maximal ideal P. Since r_1 lies in P by hypothesis, 1 lies in P, and we have a contradiction. □

Proposition 8.52. Let R be a Noetherian commutative ring, and let I and P be ideals in R with P prime. If $IP = I$, then $I = 0$.

PROOF. Let us localize with respect to the prime ideal P. If we write S for the set-theoretic complement of P in R, then $R_P = S^{-1}R$ is a local ring by Corollary 8.50, and its unique maximal ideal is $S^{-1}P$. Since $(S^{-1}I)(S^{-1}R) = S^{-1}IR = S^{-1}I$, $S^{-1}I$ is an ideal in R_P. Also, $(S^{-1}I)(S^{-1}P) = S^{-1}IP = S^{-1}I$, and $S^{-1}I$ has to be proper. In Nakayama's Lemma (Lemma 8.51), let us take M to be the $S^{-1}R$ module $S^{-1}I$. Since $S^{-1}P$ is the only maximal ideal in $S^{-1}R$, M is contained in all maximal ideals of $S^{-1}R$. Since R is Noetherian, Corollary 8.48a shows $S^{-1}R$ to be Noetherian, and the ideal $S^{-1}I$ is a finitely generated $S^{-1}R$ module by Corollary 8.31c. The lemma applies since $(S^{-1}P)(S^{-1}I) = S^{-1}I$, and the conclusion is that $S^{-1}I = 0$. Then the subset I of $S^{-1}I$ must be 0. □

Proposition 8.53. Let R be an integral domain, F be its field of fractions, K be any field containing F, and T be the integral closure of R in K. If P is a maximal ideal in R, then $PT \neq T$, and there exists a maximal ideal Q of T with $P = R \cap Q$.

REMARKS. This result inverts the construction of Proposition 8.43, of course not necessarily uniquely. The examples in Section 7 illustrate what can happen

in simple cases. More detailed analysis of what can happen in general requires some field theory and is postponed to Chapter IX, specifically when we discuss "splitting of prime ideals in extensions."

PROOF. If $PT \neq T$, then Proposition 8.8 supplies a maximal ideal Q of T with $PT \subseteq Q$. Since 1 is not in Q, we then have $P \subseteq R \cap Q \subsetneqq R$. Consequently the maximality of P implies that $P = R \cap Q$.

Arguing by contradiction, we now assume that $PT = T$. Localizing, let S be the set-theoretic complement of P in R, so that $S^{-1}P$ is the unique maximal ideal of $S^{-1}R$ by Corollary 8.50. From $PT = T$, we can write

$$1 = a_1 t_1 + \cdots + a_n t_n \qquad (*)$$

with each a_i in P and each t_i in T. If we define T_0 to be the subring $R[t_1, \cdots, t_n]$ of T, then T_0 is a finitely generated R module by Proposition 8.37, and $S^{-1}T_0$ is therefore a finitely generated $S^{-1}R$ module. Equation $(*)$ shows that 1 lies in PT_0. Multiplying by an arbitrary element of T_0, we see that $PT_0 = T_0$. Since $S^{-1}S^{-1} = S^{-1}$, we obtain $(S^{-1}P)(S^{-1}T_0) = S^{-1}T_0$. Nakayama's Lemma (Lemma 8.51) allows us to conclude that $S^{-1}T_0 = 0$. Since 1 lies in T_0, we have arrived at a contradiction. □

11. Dedekind Domains

A **Dedekind domain** is an integral domain with the following three properties:

(i) it is Noetherian,
(ii) it is integrally closed,
(iii) every nonzero prime ideal is maximal.

Every principal ideal domain R is a Dedekind domain. In fact, (i) every ideal in R is singly generated, (ii) R is integrally closed by Proposition 8.41, and (iii) every nonzero prime ideal in R is maximal by Corollary 8.16.

We shall be interested in Dedekind domains that are obtained by enlarging a principal ideal domain suitably. The general theorem in this direction is that if R is a Dedekind domain with field of fractions F and if K is a field containing F with $\dim_F K$ finite, then the integral closure of R in K is a Dedekind domain. Let us state something less sweeping.

Theorem 8.54. If R is a Dedekind domain with field of fractions F and if K is a field containing F with $\dim_F K$ finite, then the integral closure T of R in K is a Dedekind domain if any of the following three conditions holds:

(a) T is Noetherian,
(b) T is finitely generated as an R module,
(c) the field extension $F \subseteq K$ is "separable."

REMARKS. The term "separable" will be defined in Chapter IX, and the fact that (c) implies (b) will be proved at that time. It will be proved also that characteristic 0 implies separable. For now, we shall be content with showing that (b) implies (a) and that (a) implies that T is a Dedekind domain.

PROOF. We are given that R satisfies conditions (i), (ii), (iii) above, and we are to verify the conditions for T. Condition (ii) holds for T by Corollary 8.40, and Proposition 8.45 shows that (iii) holds. If (a) holds, then T satisfies the defining conditions of a Dedekind domain.

Let us see that (b) implies (a). If (b) holds, then Proposition 8.34 shows that every R submodule of T is finitely generated. Since $T \supseteq R$, every T submodule of T is finitely generated. That is, every ideal of T is finitely generated, and T is Noetherian. Thus (a) holds, and the proof is complete. □

Example 2 of integral closures in Section 9 showed that the integral closure of $\mathbb{Z}$ in $\mathbb{Q}(\sqrt{m}\,)$ is doubly generated as a $\mathbb{Z}$ module, a set of generators being either $\{1, \sqrt{m}\,\}$ or $\{1, \frac{1}{2}(1 + \sqrt{m}\,)\}$, depending on the value of m. Example 3 showed, under the assumption that $\mathbb{K}$ has characteristic different from 2, that the integral closure of $\mathbb{K}[x]$ in $\mathbb{K}(x)[\sqrt{P(x)}\,]$ is doubly generated as a $\mathbb{K}[x]$ module, a set of generators being $\{1, \sqrt{P(x)}\,\}$. Since $\mathbb{Z}$ and $\mathbb{K}[x]$ are principal ideal domains and hence Dedekind domains, these examples give concrete cases in which hypothesis (b) in Theorem 8.54 is satisfied. Consequently in each case the theorem asserts that a certain explicit integral closure is a Dedekind domain.

Theorem 8.55 (unique factorization of ideals). If R is a Dedekind domain, then each nonzero proper ideal I in R decomposes as a finite product $\prod_{j=1}^{n} P_j^{k_j}$, where the P_j's are distinct nonzero prime ideals and the k_j's are positive integers. Moreover,

(a) the decomposition into positive powers of distinct nonzero prime ideals is unique up to the order of the factors,
(b) the power P^k of a nonzero prime ideal P appearing in the decomposition of I is characterized as the unique nonnegative integer such that P^k contains I and P^{k+1} does not contain I (with $k = 0$ interpreted as saying that P is not one of the P_j),
(c) whenever I, J_1, J_2 are nonzero ideals with $IJ_1 = IJ_2$, then $J_1 = J_2$,
(d) whenever I and J_1 are two nonzero proper ideals with $I \subseteq J_1$, then there exists a nonzero ideal J_2 with $I = J_1 J_2$.

Let us say that a nonzero ideal J_1 **divides** a nonzero ideal I if $I = J_1 J_2$ for some ideal J_2. We say also that J_1 is a **factor** of I. Conclusion (d), once it is established, is an important principle for working with ideals in a Dedekind domain: *to contain is to divide*.

Thinking along these lines leads us to expect that prime ideals play some special role with respect to containment. Such a role is captured by the following lemma.

Lemma 8.56. In an integral domain, if P is a prime ideal such that $P \supseteq I_1 \cdots I_n$ for the product of the ideals $I_1, \ldots, I_n$, then $P \supseteq I_j$ for some j.

PROOF. By induction it is enough to handle $n = 2$. Thus suppose $P \supseteq I_1 I_2$. We are to show that $P \supseteq I_1$ or $P \supseteq I_2$. Arguing by contradiction, suppose on the contrary that $x \in I_1$ and $y \in I_2$ are elements with $x \notin P$ and $y \notin P$. Then xy cannot be in P since P is prime, but xy is in $I_1 I_2 \subseteq P$, and we have a contradiction. □

Lemma 8.57. Let R be a Dedekind domain, and let I be a nonzero ideal of R. Then there exists a finite product $P_1 \cdots P_k$ of nonzero prime ideals, possibly empty and not necessarily having distinct factors, such that $P_1 \cdots P_k \subseteq I$.

PROOF. We argue by contradiction. Among all nonzero ideals for which there is no such finite product, choose one, say J, that is maximal under inclusion. This choice is possible since R is Noetherian. The ideal J cannot be prime since otherwise $J \subseteq J$ would be the containment asserted by the lemma. Thus we can choose elements a_1 and a_2 in R with $a_1 a_2 \in J$, $a_1 \notin J$, and $a_2 \notin J$. Define ideals I_1 and I_2 by $I_1 = J + Ra_1$ and $I_2 = J + Ra_2$. These strictly contain J, and their product manifestly has $I_1 I_2 \subseteq J$. By maximality of J, we can find products $P_1 \cdots P_k$ and $Q_1 \cdots Q_l$ of nonzero prime ideals with $P_1 \cdots P_k \subseteq I_1$ and $Q_1 \cdots Q_l \subseteq I_2$. Then $P_1 \cdots P_k Q_1 \cdots Q_l \subseteq I_1 I_2 \subseteq J$, contradiction. □

Lemma 8.58. Let R be a Dedekind domain, regard R as embedded in its field of fractions F, let P be a nonzero prime ideal in R, and define

$$P^{-1} = \{x \in F \mid xP \subseteq R\}.$$

Then the set PP^{-1} of sums of products equals R.

PROOF. By definition of P^{-1}, $P \subseteq PP^{-1} \subseteq R$. Since P is an ideal and PP^{-1} is closed under addition and negatives, PP^{-1} is an ideal. Property (iii) of Dedekind domains shows that P is a maximal ideal in R, and therefore $PP^{-1} = P$ or $PP^{-1} = R$. We are to rule out the first alternative.

Thus suppose that $PP^{-1} = P$. Since R is Noetherian by (i), P is a finitely generated R submodule of F. The equality $PP^{-1} = P$ implies that each member x of P^{-1} has $xP \subseteq P$, and Proposition 8.35c implies that each such x is integral over R. Since R is integrally closed by (ii), x is in R. Thus $P^{-1} \subseteq R$, and the definition of P^{-1} shows that $P^{-1} = R$.

Fix a nonzero element a of P. Applying Lemma 8.57, find a product of nonzero prime ideals such that $P_1 \cdots P_k \subseteq (a) \subseteq P$. Without loss of generality, we may assume that k is as small as possible among all such inclusions. Since P is prime and $P_1 \cdots P_k \subseteq P$, Lemma 8.56 shows that P contains some P_j, say P_1. By (iii), P_1 is maximal, and therefore $P = P_1$. Form the product $P_2 \cdots P_k$, taking this product to be R if $k = 1$. Then $P_2 \cdots P_k$ is not a subset of (a), by minimality of k, and there exists a member b of $P_2 \cdots P_k$ that is not in (a). On the other hand, $PP_2 \cdots P_k \subseteq (a)$ shows that $Pb \subseteq (a)$, hence that $a^{-1}bP \subseteq R$. Thus $a^{-1}b$ is in P^{-1}, which we are assuming is R. In other words, $a^{-1}b$ is in R, and b is in $aR = (a)$, contradiction. □

PROOF OF THEOREM 8.55. Arguing by contradiction, we may assume because R is Noetherian that I is maximal among the nonzero proper ideals that do not decompose as products of prime ideals. Then certainly I is not prime. Application of Proposition 8.8 produces a maximal ideal P containing I, and P is prime by Corollary 8.11. Multiplying $I \subseteq P$ by P^{-1} as in Lemma 8.58, we obtain $I \subseteq P^{-1}I \subseteq P^{-1}P = R$, the equality holding by Lemma 8.58. Hence $P^{-1}I$ is an ideal. An equality $I = P^{-1}I$ would imply that $PI = PP^{-1}I = I$ by Lemma 8.58, and then Proposition 8.52 would yield $I = 0$, a contradiction to the hypothesis that I is nonzero. An equality $P^{-1}I = R$ would imply $I = PP^{-1}I = PR = P$ by Lemma 8.58, in contradiction to the fact that I is not prime. We conclude that $I \subsetneqq P^{-1}I \subsetneqq R$. The maximal choice of I shows that $P^{-1}I$ decomposes as a product $P^{-1}I = P_1 \cdots P_r$ of prime ideals, not necessarily distinct. One more application of Lemma 8.58 yields $I = PP^{-1}I = PP_1 \cdots P_r$, and we have a contradiction. We conclude that every nonzero proper ideal decomposes as a product of prime ideals. Grouping equal factors, we can write the decomposition as in the statement of the theorem.

Next let us establish uniqueness as in (a). Suppose that we have two equal decompositions $P_1 \cdots P_r = Q_1 \cdots Q_s$ as the product of prime ideals, and suppose that $r \le s$. We show by induction on r that $r = s$ and that the factors on the two sides match, apart from their order. The base case of the induction is $r = 0$, and then it is evident that $s = 0$. Assume the uniqueness for $r - 1$. Since P_1 is prime and $P_1 \supseteq Q_1 \cdots Q_s$, $P_1 \supseteq Q_j$ for some j by Lemma 8.56. By (iii) for Dedekind domains, Q_j is a maximal ideal, and therefore $P_1 = Q_j$. Multiplying the equality $P_1 \cdots P_r = Q_1 \cdots Q_s$ by P_1^{-1} and applying Lemma 8.58 to each side, we obtain $P_2 \cdots P_r = Q_1 \cdots Q_{j-1}Q_{j+1} \cdots Q_s$. The inductive hypothesis implies that $r - 1 = s - 1$ and the factors on the two sides match, apart from their order. Then we can conclude about the equality $P_1 \cdots P_r = Q_1 \cdots Q_s$ that $r = s$ and that the factors on the two sides match, apart from their order. This proves (a).

Let us establish the formula in (b) for k_j. Suppose that P is a prime ideal.

By (a), we can write $I = P^n J$ for a certain integer $n \geq 0$ in such a way that P does not appear in the unique decomposition of J. Certainly $P^k \supseteq I$ for $k \leq n$ because $P^k \supseteq P^k P^{n-k} = P^n \supseteq P^n J = I$. Suppose $P^{n+1} \supseteq I$. Multiplying $P^{n+1} \supseteq I = P^n J$ by n factors of P^{-1} and using Lemma 8.58 repeatedly, we obtain $P \supseteq P^{-n} I = J$. Since P is prime, Lemma 8.56 shows that P must contain one of the factors when J is decomposed as the product of prime ideals, and we have a contradiction to the maximality of this factor unless this factor is P itself. In this case, P appears in the decomposition of J, and again we have a contradiction.

For (c), if $IJ_1 = IJ_2$, substitute the unique decompositions as products of prime ideals for I, J_1, and J_2, and use (a) to cancel the factors from I on each side, obtaining $J_1 = J_2$.

For (d), suppose that I and J_1 are two nonzero proper ideals with $I \subseteq J_1$. If $P_i^{k_i}$ is the largest power of a prime ideal P_i appearing in the decomposition of J_1, then $P_i^{k_i} \supseteq J_1 \supseteq I$, and (b) shows that $P_i^{k_i}$ appears in the decomposition of I. In other words, if l_i is the largest power of P_i appearing in the decomposition of I, then $l_i \geq k_i$. Let $J_2 = \prod_i P_i^{l_i - k_i}$. Then we obtain $I = J_1 J_2$, and (d) is proved. □

Corollary 8.59. Let R be a Dedekind domain, and let P be a nonzero prime ideal in R. Then there exists an element π in P such that π is not in P^2, and any such element has the property that π^k is not in P^{k+1} for any $k \geq 1$.

PROOF. Proposition 8.52 shows that P^2 is a proper subset of P, and therefore we can find an element π in P that is not in P^2. Since the principal ideal (π) has $(\pi) \subseteq P$ and $(\pi) \subsetneqq P^2$, the factorization of (π) involves P but not P^2. Thus we can use Theorem 8.55 to write $(\pi) = PQ_1 \cdots Q_n$ for prime ideals $Q_1, \ldots, Q_n$ different from P. Then $(\pi^k) = (\pi)^k = P^k Q_1^k \cdots Q_n^k$, and (b) of the theorem says that P^{k+1} does not contain (π^k). □

Corollary 8.60. Let R be a Dedekind domain, and let P be a nonzero prime ideal in R. For any integer $e \geq 1$, the natural action of R on powers of P makes P^{e-1}/P^e into a vector space over the field R/P, and this vector space is 1-dimensional.

REMARKS. This technical-sounding corollary will be used crucially late in Chapter IX of this volume and again in Chapter V of *Advanced Algebra*.

PROOF. Since $R(P^{e-1}) \subseteq P^{e-1}$ and $P(P^{e-1}) \subseteq P^e$, we obtain

$$(R/P)(P^{e-1}/P^e) \subseteq P^{e-1}/P^e.$$

Thus P^{e-1}/P^e is a unital R/P module, i.e., a vector space over the field R/P. We show that it has dimension 1. Corollary 8.59 shows that there exists a member

π of P not in P^2, and it shows that π^k is not in P^{k+1} for any k. This element π has the property that $(\pi) = PQ_1 \cdots Q_r$ for nonzero prime ideals $Q_1, \ldots, Q_r$ distinct from P, and thus

$$R\pi^{e-1} = (\pi^{e-1}) = (\pi)^{e-1} = P^{e-1}Q_1^{e-1} \cdots Q_r^{e-1}.$$

Hence

$$R\pi^{e-1} + P^e = P^{e-1}(Q_1^{e-1} \cdots Q_r^{e-1} + P).$$

The ideal in parentheses on the right side strictly contains P since the failure of P to divide $Q_1^{e-1} \cdots Q_r^{e-1}$ means that P does not contain $Q_1^{e-1} \cdots Q_r^{e-1}$ (by Theorem 8.55d). Since P is maximal, the ideal in parentheses is R, and we see that $R(\pi^{e-1} + P^e) = P^{e-1}/P^e$. Therefore $(R/P)(\pi^{e-1} + P^e) = P^{e-1}/P^e$. This formula says that P^{e-1}/P^e consists of all scalar multiples of a certain element, and it follows that P^{e-1}/P^e is 1-dimensional. □

Lemma 8.61. If P and Q are distinct maximal ideals in an integral domain R and if k and l are positive integers, then $P^k + Q^l = R$.

PROOF. We know that $P^k + Q^l$ is an ideal. Arguing by contradiction, assume that it is proper. Then we can find a maximal ideal M with $M \supseteq P^k + Q^l$. This M satisfies $M \supseteq P^k$ and $M \supseteq Q^l$. By Lemma 8.56, $M \supseteq P$ and $M \supseteq Q$. Since P and Q are distinct and maximal, we obtain $P = M = Q$, contradiction. □

Corollary 8.62. If R is a Dedekind domain with only finitely many prime ideals, then R is a principal ideal domain.

REMARKS. Corollary 8.48 may be used to produce examples to which Corollary 8.62 is applicable. All we have to do is to take one of our standard Dedekind domains R and localize with respect to a nonzero prime ideal P. The corollary says that the result R_P is a Dedekind domain, and it has a unique maximal ideal, hence a unique nonzero prime ideal. The conclusion is that R_P is a principal ideal domain.

PROOF. Let $P_1, \ldots, P_n$ be the distinct nonzero prime ideals. Theorem 8.55 shows that any nonzero ideal I in R factors uniquely as $I = P_1^{k_1} \cdots P_n^{k_n}$ with each $k_j \geq 0$. For $1 \leq i \leq n$, Corollary 8.59 produces π_i in P_i such that π_i is not in P_i^2, and it shows that π_i^m is not in P_i^{m+1}.

Lemma 8.61 gives $P_i^{k_i} + P_j^{k_j} = R$ if $i \neq j$. Applying the Chinese Remainder Theorem (Theorem 8.27a), we can find an element a in R with $a \equiv \pi_i^{k_i} \bmod P_i^{k_i+1}$ for $1 \leq i \leq n$. Using Theorem 8.55 again, let $(a) = P_1^{l_1} \cdots P_n^{l_n}$ be the unique factorization of the principal ideal (a). The defining property of a shows that a is in $P_i^{k_i}$ but not $P_i^{k_i+1}$ for each i. Thus (a) is contained in $P_i^{k_i}$ but not in $P_i^{k_i+1}$. By Theorem 8.55b, $l_i = k_i$ for each i. Hence the ideal $I = P_1^{k_1} \cdots P_n^{k_n} = (a)$ is exhibited as principal. □

Corollary 8.63. If R is a Dedekind domain and if $I = \prod_{j=1}^{n} P_j^{k_j}$ is the unique factorization of a nonzero proper ideal I as the product of positive powers of distinct prime ideals P_j, then the map $r \mapsto \prod_{j=1}^{n} P_j^{k_j}$ defined on R by $r \mapsto (\dots, r + P_j^{k_j}, \dots)$ descends to a ring isomorphism

$$R/I \cong R/P_1^{k_1} \times \cdots \times R/P_n^{k_n}.$$

PROOF. Lemma 8.61 shows that $P_i^{k_i} + P_j^{k_j} = R$ if $i \neq j$. Then the result follows immediately from the Chinese Remainder Theorem (Theorem 8.27). □

12. Problems

1. This problem examines ring homomorphisms of the field of real numbers into itself that carry 1 into 1. Let φ be such a homomorphism.
   (a) Prove that φ is the identity on $\mathbb{Q}$.
   (b) Prove that φ maps squares into squares.
   (c) Prove that φ respects the ordering of $\mathbb{R}$, i.e., that $a \leq b$ implies $\varphi(a) \leq \varphi(b)$.
   (d) Prove that φ is the identity on $\mathbb{R}$.
2. An element r in a commutative ring with identity is called **nilpotent** if $r^n = 0$ for some integer n. Prove that if r is nilpotent, then $1 + r$ is a unit.
3. If R is a field, prove that the embedding of R in its field of fractions exhibits R as isomorphic to its field of fractions.
4. Prove that X is prime in $R[X]$ if R is an integral domain.
5. Suppose that R is an integral domain that is not a field.
   (a) Prove that there is a nonzero prime ideal in $R[X]$ that is not maximal.
   (b) Prove that there is an ideal in $R[X]$ that is not principal.
6. This problem makes use of real-analysis facts concerning closed bounded intervals of the real line. Let R be the ring of all continuous functions from $[0, 1]$ into $\mathbb{R}$, with pointwise multiplication as the ring multiplication.
   (a) Prove for each x_0 in $[0, 1]$ that the set I_{x_0} of members of R that vanish at x_0 is a maximal ideal of R.
   (b) Prove that any maximal ideal I of R that is not some I_{x_0} contains finitely many members $f_1, \dots, f_n$ of R that have no common zero on $[0, 1]$.
   (c) By considering $f_1^2 + \cdots + f_n^2$ in (b), prove that every maximal ideal of R is of the form I_{x_0} for some x_0 in $[0, 1]$.
7. Let R be the ring of all bounded continuous functions from $\mathbb{R}$ into $\mathbb{R}$, with pointwise multiplication as the ring multiplication. Say that a member f of R vanishes at infinity if for each $\epsilon > 0$, there is some N such that $|f(x)| < \epsilon$ whenever $|x| \geq N$. Answer the following:

(a) Show that the subset I_∞ of all members of R that vanish at infinity is an ideal but not a maximal ideal.
(b) Why must R have at least one maximal ideal I that contains I_∞?
(c) Why can there be no x_0 in $\mathbb{R}$ such that the maximal ideal I of (b) consists of all members of R that vanish at x_0?

8. Let I be a nonzero ideal in $\mathbb{Z}[\sqrt{-5}\,]$.
(a) Prove that I contains some positive integer.
(b) Prove that I, as an abelian group under addition, is free abelian of rank 2.
(c) If n denotes the least positive integer in I, prove that I has a $\mathbb{Z}$ basis of the form $\{n, a + b\sqrt{-5}\,\}$ for a suitable member $a + b\sqrt{-5}$ of $\mathbb{Z}[\sqrt{-5}\,]$.

9. Let $\varphi : R \to R'$ be a homomorphism of commutative rings with identity such that $\varphi(1) = 1$. Prove that if P' is a prime ideal in R', then $P = \varphi^{-1}(P')$ is a prime ideal in R.

10. Determine the maximal ideals of each of the following rings:
(a) $\mathbb{R} \times \mathbb{R}$,
(b) $\mathbb{R}[X]/(X^2)$,
(c) $\mathbb{R}[X]/(X^2 - 3X + 2)$,
(d) $\mathbb{R}[X]/(X^2 + X + 1)$.

11. (a) Prove or disprove: If I is a nonzero prime ideal in $\mathbb{Q}[X]$, then $\mathbb{Q}[X]/I$ is a unique factorization domain.
(b) Prove or disprove: If I is a nonzero prime ideal in $\mathbb{Z}[X]$, then $\mathbb{Z}[X]/I$ is a unique factorization domain.

12. **(Partial fractions)** Let R be a principal ideal domain, and let F be its field of fractions.
(a) Let n be a nonzero member of R with a factorization $n = cd$ such that $\mathrm{GCD}(c, d) = 1$. Prove for each m in R that the member mn^{-1} of F has a decomposition as $mn^{-1} = ac^{-1} + bd^{-1}$ with a and b in R.
(b) Let n be a nonzero member of R with a factorization $n = p_1^{k_1} \cdots p_r^{k_r}$, the elements p_j being nonassociate primes in R. Prove for each m in R that the member mn^{-1} of F has a decomposition as $mn^{-1} = q_1 p_1^{-k_1} + \cdots + q_r p_r^{-k_r}$ with all q_j in R.

13. (a) By adapting the proof that the ring of Gaussian integers forms a Euclidean domain, prove that the function $\delta(a + b\sqrt{-2}) = a^2 + 2b^2$ satisfies $\delta(rr') = \delta(r)\delta(r')$ and exhibits $\mathbb{Z}[\sqrt{-2}]$ as a Euclidean domain.
(b) It was shown in Section 9 that $\mathbb{Z}[\sqrt{-3}\,]$ is not a unique factorization domain, hence cannot be a Euclidean domain. What goes wrong with continuing the adaptation in the previous problem so that it applies to $\mathbb{Z}[\sqrt{-3}\,]$?

14. Let G be a group, and let R be a commutative ring with identity. Examples 16 and 17 in Section 1 defined the group algebra RG and the R algebra $C(G, R)$ of functions from G into R, convolution being the multiplication in $C(G, R)$. Prove that the mapping $g \mapsto f_g$ described with Example 17 extends to an R algebra isomorphism of RG onto $C(R, G)$.

15. Let I be an ideal in $\mathbb{Z}[X]$, and suppose that the lowest degree of a nonzero polynomial in I is n and that I contains some monic polynomial of degree n. Prove that I is a principal ideal.

16. For each integer $n > 0$, exhibit an ideal I_n in $\mathbb{Z}[X]$ that cannot be written with fewer than n generators.

17. Let φ be the substitution homomorphism $\varphi : \mathbb{K}[x, y] \to \mathbb{K}[t]$ defined by $x \mapsto t^2$, $y \mapsto t^3$, and $\varphi(c) = c$ for $c \in \mathbb{K}$.
    (a) Prove that $\ker \varphi$ is the principal ideal $(y^2 - x^3)$.
    (b) What is image φ?

18. Let $R = \mathbb{Z}[i]$.
    (a) Show that each unital R module M may be regarded as an abelian group with an abelian-group homomorphism $\varphi : M \to M$ for which φ^2 is the mapping $m \mapsto -m$.
    (b) Show conversely that if M is an abelian group and there exists an abelian-group homomorphism $\varphi : M \to M$ for which φ^2 is the mapping $m \mapsto -m$, then M may be regarded as a unital R module.

19. Let R be a unique factorization domain, and let F be its field of fractions. Let $A(X)$ and $B(X)$ be nonzero polynomials in $F[X]$, let $A_0(X)$ and $B_0(X)$ be their associated primitive polynomials, and suppose that $B(X)$ divides $A(X)$ in $F[X]$. Prove that $B_0(X)$ divides $A_0(X)$ in $R[X]$.

20. Prove that an integral domain with finitely many elements is a field.

21. Two proofs of Theorem 8.18 were given, one using direct multiplication of polynomials and the other using polynomials with coefficients taken modulo (p), and it was stated that proofs in both these styles could be given for Corollary 8.22. A proof in the first style was supplied in the text. Supply a proof in the second style.

22. Let $\mathbb{K}$ be a field.
    (a) Prove that $\det \begin{pmatrix} W & X \\ Y & Z \end{pmatrix}$, when considered as a polynomial in $\mathbb{K}[W, X, Y, Z]$, is irreducible.
    (b) Let X_{ij} be indeterminates for i and j from 1 to n. Doing an induction, prove that the polynomial $\det[X_{ij}]$ is irreducible in $\mathbb{K}[X_{11}, X_{12}, \ldots, X_{nn}]$.

23. Prove that two members of $\mathbb{Z}[X]$ are relatively prime in $\mathbb{Q}[X]$ if and only if the ideal they generate in $\mathbb{Z}[X]$ contains a nonzero integer.

24. Let V be the $\mathbb{Z}[i]$ module with two generators u_1, u_2 related by the conditions $(1+i)u_1 + (2-i)u_2 = 0$ and $3u_1 + 5iu_2 = 0$. Express V as the direct sum of cyclic $\mathbb{Z}[i]$ modules.

Problems 25–26 concern the ring $R = \mathbb{Z}\left[\frac{1}{2}(1+\sqrt{-m})\right]$, where m is a square-free integer > 1 with $m \equiv 3 \bmod 4$. Let $F = \mathbb{Q}[\sqrt{-m}]$ be the field of fractions of R.

25. For $z = x + y\sqrt{-m}$ in F, define $\delta(z) = x^2 + my^2$.
    (a) Show that $\delta(zw) = \delta(z)\delta(w)$.
    (b) Show that if for each z in F there is some r in R with $\delta(z-r) < 1$, then δ exhibits R as a Euclidean domain.

26. Prove that the condition of part (b) of the previous problem is satisfied for $m = 3$, 7, and 11, and conclude that $\mathbb{Z}\left[\frac{1}{2}(1+\sqrt{-m})\right]$ is a Euclidean domain for these values of m.

Problems 27–31 classify the primes in the ring $\mathbb{Z}[i]$ of Gaussian integers. This ring is a Euclidean domain and therefore is a unique factorization domain. Members of this ring will be written as $a + bi$, and it is understood that a and b are in $\mathbb{Z}$. Put $N(a+bi) = (a+bi)(a-bi) = a^2 + b^2$.

27. Let $a + bi$ be prime in $\mathbb{Z}[i]$. Prove that
    (a) $a - bi$ is prime.
    (b) $N(a+bi)$ is a power of some positive prime p in $\mathbb{Z}$.
    (c) $N(a+bi)$ equals p or p^2 when p is as in (b).
    (d) $N(a+bi) = p^2$ in (c) forces $a + bi = p$, apart from a unit factor.

28. Prove that no prime $a+bi$ in $\mathbb{Z}[i]$ has $N(a+bi) = p$ with p of the form $4n+3$. Conclude that every positive prime in $\mathbb{Z}$ of the form $4n+3$ is a prime in $\mathbb{Z}[i]$.

29. Prove that the only primes $a+bi$ of $\mathbb{Z}[i]$ for which $N(a+bi)$ equals 2 or 2^2 are $1+i$ and its associates, for which $N(a+bi) = 2$.

30. Prove that if p is a positive prime in $\mathbb{Z}$ of the form $4n+1$, then -1 is a square in the finite field $\mathbb{F}_p$.

31. Let p be a positive prime in $\mathbb{Z}$ of the form $4n+1$.
    (a) Prove that there exist ring homomorphisms φ_1 of $\mathbb{Z}[X]$ onto $\mathbb{F}_p[X]/(X^2+1)$ and φ_2 of $\mathbb{Z}[X]$ onto $\mathbb{Z}[i]/(p)$.
    (b) Prove that $\ker\varphi_1$ and $\ker\varphi_2$ are both equal to the ideal (p, X^2+1) in $\mathbb{Z}[X]$, and deduce a ring isomorphism $\mathbb{Z}[i]/(p) \cong \mathbb{F}_p[X]/(X^2+1)$.
    (c) Taking into account the results of Problems 27 and 30, show that p is not prime in $\mathbb{Z}[i]$ and is therefore of the form $p = N(a+bi) = a^2 + b^2$ for some prime $a+bi$ in $\mathbb{Z}[i]$.
    (d) Prove a uniqueness result for the decomposition $p = a^2 + b^2$, that if also $p = a'^2 + b'^2$, then $a' + b'i$ is an associate either of $a+bi$ or of $a-bi$.

Problems 32–35 establish a theory of **elementary divisors**. This theory provides a different uniqueness result, beyond the one in Corollary 8.28, to accompany the Fundamental Theorem of Finitely Generated Modules over a Principal Ideal Domain. When specialized to $\mathbb{K}[X]$ for a field $\mathbb{K}$, the theory yields the **rational canonical form** of a member of $M_n(\mathbb{K})$. Let R be a nonzero principal ideal domain. If C and D are members of $M_{mn}(R)$, let us say that C and D are **equivalent** if there exist A in $M_m(R)$ and B in $M_n(R)$ with $\det A$ in $R^\times$, $\det B$ in $R^\times$, and $D = ACB$. Fix m and n, and put $k = \min(m, n)$. If C is a member of $M_{mn}(R)$, its **diagonal entries** are the entries $C_{11}, C_{22}, \ldots, C_{kk}$. The matrix C will be called **diagonal** if its only nonzero entries are diagonal entries. Problems 26–31 of Chapter V are relevant for Problem 34.

32. (a) Suppose that C is a diagonal matrix in $M_{mn}(R)$ with $C_{11} \neq 0$. Show that C is equivalent to a matrix C' described as follows: all entries of C' are the same as those of C except possibly for the entries $C'_{21}, \ldots, C'_{k1}$ in the first column, and these satisfy $C'_{j1} = C_{jj}$.
    (b) By applying the algorithm of Lemma 8.26 to the matrix C' in (a), prove that any nonzero diagonal matrix C in $M_{mn}(R)$ is equivalent to a diagonal matrix C'' such that C''_{11} divides all the diagonal entries of C''.
    (c) By iterating the construction in (a) and (b), prove that any diagonal matrix C in $M_{mn}(R)$ is equivalent to a diagonal matrix D having the following properties: The nonzero diagonal entries of D are the entries D_{jj} with $1 \leq j \leq l$ for some integer l with $0 \leq l \leq k$. For each j with $1 \leq j < l$, D_{jj} divides $D_{j+1,j+1}$.

33. (a) Establish the following uniqueness theorem: Let D and E be diagonal matrices in $M_{mn}(R)$ whose diagonal entries satisfy the divisibility property in (c) of the previous problem. Prove that if D and E are equivalent, then they have the same number of nonzero entries, and their corresponding diagonal entries are associates.
    (b) Combine Corollary 8.29, Problem 32, and Problem 33a to establish the following elementary-divisors version of the **Fundamental Theorem of Finitely Generated Modules**: If R is a principal ideal domain, then any finitely generated unital R module M is the direct sum of a nonunique free R submodule $\bigoplus_{i=1}^{s} R$ of a well-defined finite rank $s \geq 0$ and the R submodule T of all members m of M such that $rm = 0$ for some $r \neq 0$ in R. In turn, the R submodule T is isomorphic to a direct sum $T \cong \bigoplus_{j=1}^{l} R/(d_j)$, where the d_j are nonzero nonunits in R such that d_j divides d_{j+1} for $1 \leq j < l$. The number of l of summands and the ideals (d_j) are uniquely determined by M.

34. (a) **(Rational decomposition)** Let $\mathbb{K}$ be a field, and let $L : V \to V$ be a $\mathbb{K}$ linear mapping from a finite-dimensional $\mathbb{K}$ vector space V to itself. By applying Theorem 8.25 and the results of the previous problems to V as a

$\mathbb{K}[X]$ module with $Xv = L(v)$, prove the following: V can be written as the direct sum of cyclic subspaces $V_1, \ldots, V_r$ under L in such a way that the minimal polynomial of L on V_j divides the minimal polynomial of L on V_{j+1} for $1 \le j < r$; moreover, the integer r and the minimal polynomials are uniquely determined by L, and any two linear mappings with the same r and matching minimal polynomials are similar over $\mathbb{K}$.

(b) **(Rational canonical form)** Interpret the result of (a) as saying something about similarity over $\mathbb{K}$ of any matrix in $M_{nn}(\mathbb{K})$ to a certain block diagonal matrix with blocks of the form in Problem 28 for Chapter V and with minimal polynomials having a suitable divisibility property.

35. Let $\mathbb{K}$ and $\mathbb{L}$ be fields with $\mathbb{K} \subseteq \mathbb{L}$, and suppose that two members of $M_n(\mathbb{K})$ are conjugate via $\mathrm{GL}(n, \mathbb{L})$. Prove that they are conjugate via $\mathrm{GL}(n, \mathbb{K})$.

Problems 36–39 concern symmetric polynomials in n indeterminates over a field. Let F be a field, and let $R = F[X_1, \ldots, X_n]$. If $\sigma \in \mathfrak{S}_n$ is a permutation, then there is a corresponding substitution homomorphism of rings $\sigma^* : R \to R$ fixing F and carrying each X_j into $X_{\sigma(j)}$. A **symmetric polynomial** A in R is a member of R for which $\sigma^* A = A$ for every permutation σ. The symmetric polynomials form a subring of R containing the constants. The main result about symmetric polynomials is that every symmetric polynomial is a polynomial in the "elementary symmetric polynomials"; these will be defined below.

36. Prove that the ring homomorphisms σ^* satisfy $(\sigma\tau)^* = \sigma^*\tau^*$. Deduce that each $\sigma^* : R \to R$ is an isomorphism.

37. Prove that the homogeneous-polynomial expansion of any symmetric polynomial is into symmetric polynomials.

38. For each permutation σ, let σ^{**} be the substitution homomorphism of $R[X] \cong F[X_1, \ldots, X_n, X]$ acting as σ^* on R and carrying X to itself.
    (a) Prove that $(\sigma\tau)^{**} = \sigma^{**}\tau^{**}$ and that each σ^{**} is a ring isomorphism of $R[X]$.
    (b) Prove that each coefficient in $R[X]$ of any polynomial fixed by all σ^{**} is a symmetric polynomial in R.
    (c) The polynomial $(X - X_1)(X - X_2)\cdots(X - X_n)$ is fixed by all σ^{**}, and its coefficients are called the **elementary symmetric polynomials**. Show that they are

$$E_1 = \sum_i X_i, \quad E_2 = \sum_{i<j} X_i X_j, \quad E_3 = \sum_{i<j<k} X_i X_j X_j, \ldots, \quad E_n = X_1 X_2 \cdots X_n.$$

39. Order the monomials of total degree m by saying that the monomial $a X_1^{k_1} \cdots X_n^{k_n}$ with $a \neq 0$ and $\sum k_j = m$ is greater than the monomial $a' X_1^{l_1} \cdots X_n^{l_n}$ with $a' \neq 0$ and $\sum l_j = m$ if the first j for which $k_j \neq l_j$ has $k_j > l_j$.

(a) If $A(X_1, \ldots, X_n)$ is a nonzero symmetric polynomial homogeneous of degree m and if $aX_1^{k_1} \cdots X_n^{k_n}$ is its nonzero monomial that is highest in the above order, why must it be true that $k_1 \geq k_2 \geq \cdots \geq k_n$?

(b) Verify that the largest monomial in $E_1^{c_1} \cdots E_n^{c_n}$ in the ordering is

$$X_1^{c_1+c_2+\cdots+c_n} X_2^{c_2+\cdots+c_n} \cdots X_n^{c_n}.$$

(c) Show that if $A(X_1, \ldots, X_n)$ is a nonzero symmetric polynomial homogeneous of degree m, then there exist a symmetric polynomial $M = E_1^{c_1} \cdots E_n^{c_n}$ homogeneous of degree m and a scalar r such that the largest monomials in A and rM are equal.

(d) With notation as in (c), show that $A - rM$ equals 0 or else the largest monomial of A is greater than the largest monomial of $A - rm$.

(e) Deduce that every symmetric polynomial is a polynomial in the elementary symmetric polynomials.

Problems 40–43 concern the Pfaffian of a $(2n)$-by-$(2n)$ alternating matrix $X = [x_{ij}]$ with entries in a field $\mathbb{K}$. Here "alternating" means that $x_{ij} = -x_{ji}$ for all i and j and $x_{ii} = 0$ for all i. The Pfaffian is the polynomial in the entries of X with integer coefficients given by

$$\text{Pfaff}(X) = \sum_{\substack{\text{certain } \tau\text{'s} \\ \text{in } \mathfrak{S}_{2n}}} (\operatorname{sgn} \tau) \prod_{k=1}^{n} x_{\tau(2k-1),\tau(2k)},$$

where the sum is taken over those permutations τ such that $\tau(2k-1) < \tau(2k)$ for $1 \leq k \leq n$ and such that $\tau(1) < \tau(3) < \cdots < \tau(2n-1)$. The Pfaffian was introduced in Problems 23–28 at the end of Chapter VI. It was shown in those problems that the Pfaffian satisfies $\det X = (\text{Pfaff}(X))^2$. The present problems will make use of that result but of no other results from Chapter VI. They will also make use of facts concerning continuous functions and connected open subsets of Euclidean space.

40. Prove by induction on m that the open subset of $\mathbb{C}^m$ on which a nonzero polynomial function $P(z_1, \ldots, z_m)$ is nonzero is pathwise connected and therefore connected.

41. For this problem let $\mathbb{K} = \mathbb{C}$.

(a) For any two matrices A and X in $M_{2n}(\mathbb{C})$ with X alternating, prove that $\text{Pfaff}(A^t X A) = \pm(\det A)\text{Pfaff}(X)$ with the sign depending on A and X.

(b) Fix X, and allow A to vary. Using Problem 40, prove that the sign is always positive in (a). That is, prove that $\text{Pfaff}(A^t X A) = (\det A)\text{Pfaff}(X)$.

42. For this problem let $\mathbb{K}$ be any field. By regarding the expressions $\text{Pfaff}(A^t X A)$ and $(\det A)\text{Pfaff}(X)$ as polynomials with coefficients in $\mathbb{Z}$ in the indeterminates A_{ij} for all i and j and the indeterminates X_{ij} for $i < j$, and using the principle of permanence of identities in Section V.2, prove that $\text{Pfaff}(A^t X A) = (\det A)\text{Pfaff}(X)$ whenever A and X are in $M_{2n}(\mathbb{K})$ and X is alternating.

43. Section VI.5 defines a particular alternating matrix J for which $\text{Pfaff}(J) = 1$. A **symplectic matrix** g over $\mathbb{K}$ is one for which $g^t J g = J$. Prove that every symplectic matrix has determinant 1.

Problems 44–47 concern Dedekind domains. Let R be such a domain. It is to be proved that each nonzero ideal I is doubly generated in the sense that $I = Ra + Rb$ for suitable members a and b of R.

44. Let $R_1, \dots, R_n$ be nonzero commutative rings with identity, not necessarily integral domains. Prove that if every ideal of each R_j is principal, then every ideal in $R_1 \times \cdots \times R_n$ is principal.

45. Let P be a nonzero prime ideal, and let k be a positive integer.
    (a) Prove that the only nonzero proper ideals in R/P^k are P/P^k, P^2/P^k, $\dots$, P^{k-1}/P^k.
    (b) Using the element π in the statement of Corollary 8.59, prove that each of the ideals in (a) is principal.

46. Combining Corollary 8.63 with Problems 44 and 45, conclude that the quotient of R by any nonzero proper ideal has only principal ideals.

47. Let I be a nonzero proper ideal in R. By letting a be any nonzero element of I and by applying (c) in the previous problem to the ideal $I/(a)$ of $R/(a)$, prove that $I = Ra + Rb$ for a suitable b in I.

Problems 48–53 introduce and classify "fractional ideals" in Dedekind domains. Let R be a Dedekind domain, regarded as a subring of its field of fractions F. A **fractional ideal** in F is a finitely generated R submodule of F.

48. Prove that the fractional ideals in F that lie in R are exactly the ordinary ideals of R.

49. Prove for any fractional ideal M that there exists a nonzero member a of F such that aM lies in R and hence is an ordinary ideal. Conclude that the product of two fractional ideals is a fractional ideal.

50. Prove that if I is a nonzero ideal of R and if I^{-1} is defined by

$$I^{-1} = \{x \in F \mid xR \subseteq I\},$$

    then I^{-1} is a fractional ideal in F. Conclude that if P is a prime ideal in R, then P^{-1} as defined in Lemma 8.58 is a fractional ideal in F.

51. Prove, by arguing with an ideal that is maximal among those for which the statement is false, that to any nonzero ideal I in R corresponds some fractional ideal M of F such that $IM = R$.

52. Prove in the notation of the previous two problems that $M = I^{-1}$.

53. Deduce that every nonzero fractional ideal is of the form IJ^{-1}, where I and J are nonzero ideals. Conclude that
    (a) the nonzero fractional ideals are exactly all products $\prod_{i=1}^{n} P_i^{k_i}$, where the P_i are distinct nonzero prime ideals and the k_i are arbitrary nonzero integers, positive or negative,
    (b) the nonzero fractional ideals form a group.