Chapter 2

Ring Fundamentals

2.1 Basic Definitions and Properties

2.1.1 Definitions and Comments

A *ring* R is an abelian group with a multiplication operation $(a, b) \to ab$ that is associative and satisfies the distributive laws: $a(b+c) = ab+ac$ and $(a+b)c = ab+ac$ for all $a, b, c \in R$. We will always assume that R has at least two elements, including a multiplicative identity 1_R satisfying $a1_R = 1_R a = a$ for all a in R. The multiplicative identity is often written simply as 1, and the additive identity as 0. If a, b, and c are arbitrary elements of R, the following properties are derived quickly from the definition of a ring; we sketch the technique in each case.

(1) $a0 = 0a = 0$ $[a0 + a0 = a(0 + 0) = a0;\ 0a + 0a = (0 + 0)a = 0a]$

(2) $(-a)b = a(-b) = -(ab)$ $\quad [0 = 0b = (a + (-a))b = ab + (-a)b,\ \text{so } (-a)b = -(ab)]$
$[0 = a0 = a(b + (-b)) = ab + a(-b),\ \text{so } a(-b) = -(ab)]$

(3) $(-1)(-1) = 1$ [take $a = 1, b = -1$ in (2)]

(4) $(-a)(-b) = ab$ [replace b by $-b$ in (2)]

(5) $a(b - c) = ab - ac$ $[a(b + (-c)) = ab + a(-c) = ab + (-(ac)) = ab - ac]$

(6) $(a - b)c = ac - bc$ $[(a + (-b))c = ac + (-b)c) = ac - (bc) = ac - bc]$

(7) $1 \neq 0$ [If $1 = 0$ then for all a we have $a = a1 = a0 = 0$, so $R = \{0\}$, contradicting the assumption that R has at least two elements]

(8) The multiplicative identity is unique [If $1'$ is another multiplicative identity then $1 = 11' = 1'$]

2.1.2 Definitions and Comments

If a and b are nonzero but $ab = 0$, we say that a and b are *zero divisors*; if $a \in R$ and for some $b \in R$ we have $ab = ba = 1$, we say that a is a *unit* or that a is *invertible*.

Note that ab need not equal ba; if this holds for all $a, b \in R$, we say that R is a *commutative ring*.

An *integral domain* is a commutative ring with no zero divisors.

A *division ring* or *skew field* is a ring in which every nonzero element a has a multiplicative inverse a^{-1} (i.e., $aa^{-1} = a^{-1}a = 1$). Thus the nonzero elements form a group under multiplication.

A *field* is a commutative division ring. Intuitively, in a ring we can do addition, subtraction and multiplication without leaving the set, while in a field (or skew field) we can do division as well.

Any finite integral domain is a field. To see this, observe that if $a \neq 0$, the map $x \to ax$, $x \in R$, is injective because R is an integral domain. If R is finite, the map is surjective as well, so that $ax = 1$ for some x.

The *characteristic* of a ring R (written Char R) is the smallest positive integer such that $n1 = 0$, where $n1$ is an abbreviation for $1 + 1 + \dots 1$ (n times). If $n1$ is never 0, we say that R has *characteristic 0*. Note that the characteristic can never be 1, since $1_R \neq 0$. If R is an integral domain and Char $R \neq 0$, then Char R must be a prime number. For if Char$R = n = rs$ where r and s are positive integers greater than 1, then $(r1)(s1) = n1 = 0$, so either $r1$ or $s1$ is 0, contradicting the minimality of n.

A *subring* of a ring R is a subset S of R that forms a ring under the operations of addition and multiplication defined on R. In other words, S is an additive subgroup of R that contains 1_R and is closed under multiplication. Note that 1_R is automatically the multiplicative identity of S, since the multiplicative identity is unique (see (8) of 2.1.1).

2.1.3 Examples

1. The integers $\mathbb{Z}$ form an integral domain that is not a field.

2. Let $\mathbb{Z}_n$ be the integers modulo n, that is, $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$ with addition and multiplication mod n. (If $a \in \mathbb{Z}_n$ then a is identified with all integers $a + kn$, $k = 0, \pm 1, \pm 2, \dots\}$. Thus, for example, in $\mathbb{Z}_9$ the multiplication of 3 by 4 results in 3 since $12 \equiv 3 \bmod 9$, and therefore 12 is identified with 3.

$\mathbb{Z}_n$ is a ring, which is an integral domain (and therefore a field, since $\mathbb{Z}_n$ is finite) if and only if n is prime. For if $n = rs$ then $rs = 0$ in $\mathbb{Z}_n$; if n is prime then every nonzero element in $\mathbb{Z}_n$ has a multiplicative inverse, by Fermat's little theorem 1.3.4.

Note that by definition of characteristic, any field of prime characteristic p contains an isomorphic copy of $\mathbb{Z}_p$. Any field of characteristic 0 contains a copy of $\mathbb{Z}$, hence a copy of the rationals $\mathbb{Q}$.

3. If $n \geq 2$, then the set $M_n(R)$ of all n by n matrices with coefficients in a ring R forms a noncommutative ring, with the identity matrix I_n as multiplicative identity. If we identify the element $c \in R$ with the diagonal matrix cI_n, we may regard R as a subring of $M_n(R)$. It is possible for the product of two nonzero matrices to be zero, so that $M_n(R)$ is not an integral domain. (To generate a large class of examples, let E_{ij} be the matrix with 1 in row i, column j, and 0's elsewhere. Then $E_{ij}E_{kl} = \delta_{jk}E_{il}$, where δ_{jk} is 1 when $j = k$, and 0 otherwise.)

4. Let $1, i, j$ and k be basis vectors in 4-dimensional Euclidean space, and define multiplication of these vectors by

$$i^2 = j^2 = k^2 = -1, \quad ij = k, \quad jk = i, \quad ki = j, \quad ji = -ij, \quad kj = -jk, \quad ik = -ki \quad (1)$$

Let H be the set of all linear combinations $a + bi + cj + dk$ where a, b, c and d are real numbers. Elements of H are added componentwise and multiplied according to the above rules, i.e.,

$$(a + bi + cj + dk)(x + yi + zj + wk) = (ax - by - cz - dw) + (ay + bx + cw - dz)i \\ + (az + cx + dy - bw)j + (aw + dx + bz - cy)k.$$

H (after Hamilton) is called the ring of *quaternions*. In fact H is a division ring; the inverse of $a + bi + cj + dk$ is $(a^2 + b^2 + c^2 + d^2)^{-1}(a - bi - cj - dk)$.

H can also be represented by 2 by 2 matrices with complex entries, with multiplication of quaternions corresponding to ordinary matrix multiplication. To see this, let

$$\mathbf{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{i} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \ \mathbf{j} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \ \mathbf{k} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix};$$

a direct computation shows that $\mathbf{1}, \mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ obey the multiplication rules (1) given above. Thus we may identify the quaternion $a + bi + cj + dk$ with the matrix

$$a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = \begin{bmatrix} a + bi & c + di \\ -c + di & a - bi \end{bmatrix}$$

(where in the matrix, i is $\sqrt{-1}$, not the quaternion i).

The set of 8 elements ± 1, $\pm i$, $\pm j$, $\pm k$ forms a group under multiplication; it is called the *quaternion group*.

5. If R is a ring, then $R[X]$, the set of all polynomials in X with coefficients in R, is also a ring under ordinary polynomial addition and multiplication, as is $R[X_1, \dots, X_n]$, the set of polynomials in n variables $X_i, 1 \le i \le n$, with coefficients in R. Formally, the polynomial $A(X) = a_0 + a_1X + \cdots + a_nX^n$ is simply the sequence $(a_0, \dots, a_n)$; the symbol X is a placeholder. The product of two polynomials $A(X)$ and $B(X)$ is a polynomial whose X^k-coefficient is $a_0b_k + a_1b_{k-1} + \cdots + a_kb_0$. If we wish to evaluate a polynomial on R, we use the *evaluation map*

$$a_0 + a_1X + \cdots + a_nX^n \to a_0 + a_1x + \cdots + a_nx^n$$

where x is a particular element of R. A nonzero polynomial can evaluate to 0 at all points of R. For example, $X^2 + X$ evaluates to 0 on $\mathbb{Z}_2$, the field of integers modulo 2, since $1 + 1 = 0 \bmod 2$. We will say more about evaluation maps in Section 2.5, when we study polynomial rings.

6. If R is a ring, then $R[[X]]$, the set of *formal power series*

$$a_0 + a_1X + a_2X^2 + \dots$$

with coefficients in R, is also a ring under ordinary addition and multiplication of power series. The definition of multiplication is purely formal and convergence is never mentioned; we simply define the coefficient of X^n in the product of $a_0 + a_1X + a_2X^2 + \dots$ and $b_0 + b_1X + b_2X^2 + \dots$ to be $a_0b_n + a_1b_{n-1} + \cdots + a_{n-1}b_1 + a_nb_0$.

In Examples 5 and 6, if R is an integral domain, so are $R[X]$ and $R[[X]]$. In Example 5, look at leading coefficients to show that if $f(X) \neq 0$ and $g(X) \neq 0$, then $f(X)g(X) \neq 0$. In Example 6, if $f(X)g(X) = 0$ with $f(X) \neq 0$, let a_i be the first nonzero coefficient of $f(X)$. Then $a_ib_j = 0$ for all j, and therefore $g(X) = 0$.

2.1.4 Lemma

The *generalized associative law* holds for multiplication in a ring. There is also a *generalized distributive law*:

$$(a_1 + \cdots + a_m)(b_1 + \cdots + b_n) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j.$$

Proof. The argument for the generalized associative law is exactly the same as for groups; see the beginning of Section 1.1. The generalized distributive law is proved in two stages. First set $m = 1$ and work by induction on n, using the left distributive law $a(b+c) = ab + ac$. Then use induction on m and the right distributive law $(a+b)c = ac + bc$ on $(a_1 + \cdots + a_m + a_{m+1})(b_1 + \cdots + b_n)$. ♣

2.1.5 Proposition

$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$, the *binomial theorem*, is valid in any ring, if $ab = ba$.

Proof. The standard proof via elementary combinatorial analysis works. Specifically, $(a+b)^n = (a+b)\dots(a+b)$, and we can expand this product by multiplying an element (a or b) from object 1 (the first $(a+b)$) times an element from object 2 times ... times an element from object n, in all possible ways. Since $ab = ba$, these terms are of the form $a^k b^{n-k}, 0 \le k \le n$. The number of terms corresponding to a given k is the number of ways of selecting k objects from a collection of n, namely $\binom{n}{k}$. ♣

Problems For Section 2.1

1. If R is a field, is $R[X]$ a field always? sometimes? never?
2. If R is a field, what are the units of $R[X]$?
3. Consider the ring of formal power series with rational coefficients.
   (a) Give an example of a nonzero element that does not have a multiplicative inverse, and thus is not a unit.
   (b) Give an example of a nonconstant element (one that is not simply a rational number) that does have a multiplicative inverse, and therefore is a unit.
4. Let $\mathbb{Z}[i]$ be the ring of *Gaussian integers* $a + bi$, where $i = \sqrt{-1}$ and a and b are integers. Show that $\mathbb{Z}[i]$ is an integral domain that is not a field.
5. Continuing Problem 4, what are the units of $\mathbb{Z}[i]$?
6. Establish the following quaternion identities:
   (a) $(x_1 + y_1 i + z_1 j + w_1 k)(x_2 - y_2 i - z_2 j - w_2 k)$
   $= (x_1x_2 + y_1y_2 + z_1z_2 + w_1w_2) + (-x_1y_2 + y_1x_2 - z_1w_2 + w_1z_2)i$
   $+ (-x_1z_2 + z_1x_2 - w_1y_2 + y_1w_2)j + (-x_1w_2 + w_1x_2 - y_1z_2 + z_1y_2)k$
   (b) $(x_2 + y_2 i + z_2 j + w_2 k)(x_1 - y_1 i - z_1 j - w_1 k)$
   $= (x_1x_2 + y_1y_2 + z_1z_2 + w_1w_2) + (x_1y_2 - y_1x_2 + z_1w_2 - w_1z_2)i$
   $+ (x_1z_2 - z_1x_2 + w_1y_2 - y_1w_2)j + (x_1w_2 - w_1x_2 + y_1z_2 - z_1y_2)k$

(c) The product of a quaternion $h = a + bi + cj + dk$ and its *conjugate* $h^* = a - bi - cj - dk$ is $a^2 + b^2 + c^2 + d^2$. If q and t are quaternions, then $(qt)^* = t^*q^*$.

7. Use Problem 6 to establish *Euler's Identity* for real numbers x_r, y_r, z_r, w_r, $r = 1, 2$:

$$\begin{aligned}(x_1^2 + y_1^2 + z_1^2 + w_1^2)(x_2^2 + y_2^2 + z_2^2 + w_2^2) = {} & (x_1x_2 + y_1y_2 + z_1z_2 + w_1w_2)^2 \\ & + (x_1y_2 - y_1x_2 + z_1w_2 - w_1z_2)^2 \\ & + (x_1z_2 - z_1x_2 + w_1y_2 - y_1w_2)^2 \\ & + (x_1w_2 - w_1x_2 + y_1z_2 - z_1y_2)^2\end{aligned}$$

8. Recall that an endomorphism of a group G is a homomorphism of G to itself. Thus if G is abelian, an endomorphism is a function $f\colon G \to G$ such that $f(a + b) = f(a) + f(b)$ for all $a, b \in G$. Define addition of endomorphisms in the natural way, $(f+g)(a) = f(a)+g(a)$, and define multiplication as functional composition, $(fg)(a) = f(g(a))$. Show that the set End G of endomorphisms of G becomes a ring under these operations.

9. Continuing Problem 8, what are the units of End G?

10. It can be shown that every positive integer is the sum of 4 squares. A key step is to prove that if n and m can be expressed as sums of 4 squares, so can nm. Do this using Euler's identity, and illustrate for the case $n = 34, m = 54$.

11. Which of the following collections of n by n matrices form a ring under matrix addition and multiplication?

    (a) symmetric matrices

    (b) matrices whose entries are 0 except possibly in column 1

    (c) lower triangular matrices ($a_{ij} = 0$ for $i < j$)

    (d) upper triangular matrices ($a_{ij} = 0$ for $i > j$)

2.2 Ideals, Homomorphisms, and Quotient Rings

Let $f\colon R \to S$, where R and S are rings. Rings are, in particular, abelian groups under addition, so we know what it means for f to be a group homomorphism: $f(a + b) = f(a) + f(b)$ for all a, b in R. It is then automatic that $f(0_R) = 0_S$ (see (1.3.11)). It is natural to consider mappings f that preserve multiplication as well as addition, i.e.,

$$f(a + b) = f(a) + f(b) \text{ and } f(ab) = f(a)f(b) \text{ for all } a, b \in R.$$

But here it does not follow that f maps the multiplicative identity 1_R to the multiplicative identity 1_S. We have $f(a) = f(a1_R) = f(a)f(1_R)$, but we cannot multiply on the left by $f(a)^{-1}$, which might not exist. We avoid this difficulty by only considering functions f that have the desired behavior.

2.2.1 Definition

If $f\colon R \to S$, where R and S are rings, we say that f is a *ring homomorphism* if $f(a+b) = f(a) + f(b)$ and $f(ab) = f(a)f(b)$ for all $a, b \in R$, and $f(1_R) = 1_S$.

2.2.2 Example

Let $f\colon \mathbb{Z} \to M_n(R)$, $n \geq 2$, be defined by $f(n) = nE_{11}$ (see 2.1.3, Example 3). Then we have $f(a+b) = f(a) + f(b), f(ab) = f(a)f(b)$, but $f(1) \neq I_n$. Thus f is not a ring homomorphism.

In Chapter 1, we proved the basic isomorphism theorems for groups, and a key observation was the connection between group homomorphisms and normal subgroups. We can prove similar theorems for rings, but first we must replace the normal subgroup by an object that depends on multiplication as well as addition.

2.2.3 Definitions and Comments

Let I be a subset of the ring R, and consider the following three properties:

(1) I is an additive subgroup of R.

(2) If $a \in I$ and $r \in R$ then $ra \in I$; in other words, $rI \subseteq I$ for every $r \in R$.

(3) If $a \in I$ and $r \in R$ then $ar \in I$; in other words, $Ir \subseteq I$ for every $r \in R$.

If (1) and (2) hold, I is said to be a *left ideal* of R. If (1) and (3) hold, I is said to be a *right ideal* of R. If all three properties are satisfied, I is said to be an *ideal* (or *two-sided ideal*) of R, a *proper ideal* if $I \neq R$, a *nontrivial ideal* if I is neither R nor $\{0\}$.

If $f\colon R \to S$ is a ring homomorphism, its *kernel* is

$$\ker f = \{r \in R\colon f(r) = 0\};$$

exactly as in (1.3.13), f is injective if and only if $\ker f = \{0\}$.

Now it follows from the definition of ring homomorphism that $\ker f$ is an ideal of R. The kernel must be a proper ideal because if $\ker f = R$ then f is identically 0, in particular, $f(1_R) = 1_S = 0_S$, a contradiction (see (7) of 2.1.1). Conversely, every proper ideal is the kernel of a ring homomorphism, as we will see in the discussion to follow.

2.2.4 Construction of Quotient Rings

Let I be a proper ideal of the ring R. Since I is a subgroup of the additive group of R, we can form the quotient group R/I, consisting of cosets $r + I$, $r \in R$. We define multiplication of cosets in the natural way:

$$(r + I)(s + I) = rs + I.$$

To show that multiplication is well-defined, suppose that $r + I = r' + I$ and $s + I = s' + I$, so that $r' - r$ is an element of I, call it a, and $s' - s$ is an element of I, call it b. Thus

$$r's' = (r + a)(s + b) = rs + as + rb + ab,$$

and since I is an ideal, we have $as \in I$, $rb \in I$, and $ab \in I$. Consequently, $r's' + I = rs + I$, so the multiplication of two cosets is independent of the particular representatives r and s that we choose. From our previous discussion of quotient groups, we know that the cosets of the ideal I form a group under addition, and the group is abelian because R

itself is an abelian group under addition. Since multiplication of cosets $r+I$ and $s+I$ is accomplished simply by multiplying the coset representatives r and s in R and then forming the coset $rs+I$, we can use the ring properties of R to show that the cosets of I form a ring, called the *quotient ring* of R by I. The identity element of the quotient ring is 1_R+I, and the zero element is 0_R+I. Furthermore, if R is a commutative ring, so is R/I. The fact that I is proper is used in verifying that R/I has at least two elements. For if $1_R+I=0_R+I$, then $1_R=1_R-0_R\in I$; thus for any $r\in R$ we have $r=r1_R\in I$, so that $R=I$, a contradiction.

2.2.5 Proposition

Every proper ideal I is the kernel of a ring homomorphism.

Proof. Define the *natural* or *canonical* map $\pi\colon R\to R/I$ by $\pi(r)=r+I$. We already know that π is a homomorphism of abelian groups and its kernel is I (see (1.3.12)). To verify that π preserves multiplication, note that

$$\pi(rs)=rs+I=(r+I)(s+I)=\pi(r)\pi(s);$$

since

$$\pi(1_R)=1_R+I=1_{R/I},$$

π is a ring homomorphism. ♣

2.2.6 Proposition

Suppose $f\colon R\to S$ is a ring homomorphism and the only ideals of R are $\{0\}$ and R. (In particular, if R is a division ring, then R satisfies this hypothesis.) Then f is injective.

Proof. Let $I=\ker f$, an ideal of R (see 2.2.3). If $I=R$ then f is identically zero, and is therefore not a legal ring homomorphism since $f(1_R)=1_S\neq 0_S$. Thus $I=\{0\}$, so that f is injective.

If R is a division ring, then in fact R has no nontrivial left or right ideals. For suppose that I is a left ideal of R and $a\in I$, $a\neq 0$. Since R is a division ring, there is an element $b\in R$ such that $ba=1$, and since I is a left ideal, we have $1\in I$, which implies that $I=R$. If I is a right ideal, we choose the element b such that $ab=1$. ♣

2.2.7 Definitions and Comments

If X is a nonempty subset of the ring R, then $\langle X\rangle$ will denote the *ideal generated by* X, that is, the smallest ideal of R that contains X. Explicitly,

$$\langle X\rangle = RXR = \text{the collection of all finite sums of the form } \sum_i r_i x_i s_i$$

with $r_i, s_i\in R$ and $x_i\in X$. To show that this is correct, verify that the finite sums of the given type form an ideal containing X. On the other hand, if J is any ideal containing X, then all finite sums $\sum_i r_i x_i s_i$ must belong to J.

If R is commutative, then $rxs = rsx$, and we may as well drop the s. In other words:

$$\text{In a commutative ring, } \langle X\rangle = RX = \text{all finite sums } \sum_i r_i x_i,\ r_i \in R, x_i \in X.$$

An ideal generated by a single element a is called a *principal ideal* and is denoted by $\langle a\rangle$ or (a). In this case, $X = \{a\}$, and therefore:

$$\text{In a commutative ring, the principal ideal generated by } a \text{ is } \langle a\rangle = \{ra \colon r \in R\},$$

the set of all multiples of a, sometimes denoted by Ra.

2.2.8 Definitions and Comments

In an arbitrary ring, we will sometimes need to consider the *sum* of two ideals I and J, defined as $\{x + y \colon x \in I, y \in J\}$. It follows from the distributive laws that $I + J$ is also an ideal. Similarly, the sum of two left [resp. right] ideals is a left [resp. right] ideal.

Problems For Section 2.2

1. What are the ideals in the ring of integers?
2. Let $M_n(R)$ be the ring of n by n matrices with coefficients in the ring R. If C_k is the subset of $M_n(R)$ consisting of matrices that are 0 except perhaps in column k, show that C_k is a left ideal of $M_n(R)$. Similarly, if R_k consists of matrices that are 0 except perhaps in row k, then R_k is a right ideal of $M_n(R)$.
3. In Problem 2, assume that R is a division ring, and let E_{ij} be the matrix with 1 in row i, column j, and 0's elsewhere.
   (a) If $A \in M_n(R)$, show that $E_{ij}A$ has row j of A as its i^{th}row, with 0's elsewhere.
   (b) Now suppose that $A \in C_k$. Show that $E_{ij}A$ has a_{jk} in the ik position, with 0's elsewhere, so that if a_{jk} is not zero, then $a_{jk}^{-1}E_{ij}A = E_{ik}$.
   (c) If A is a nonzero matrix in C_k with $a_{jk} \neq 0$, and C is any matrix in C_k, show that
   $$\sum_{i=1}^{n} c_{ik}a_{jk}^{-1}E_{ij}A = C.$$
4. Continuing Problem 3, if a nonzero matrix A in C_k belongs to the left ideal I of $M_n(R)$, show that every matrix in C_k belongs to I. Similarly, if a nonzero matrix A in R_k belongs to the right ideal I of $M_n(R)$, every matrix in R_k belongs to I.
5. Show that if R is a division ring, then $M_n(R)$ has no nontrivial two-sided ideals.
6. In $R[X]$, express the set I of polynomials with no constant term as $\langle f\rangle$ for an appropriate f, and thus show that I is a principal ideal.
7. Let R be a commutative ring whose only proper ideals are $\{0\}$ and R. Show that R is a field.
8. Let R be the ring $\mathbb{Z}_n$ of integers modulo n, where n may be prime or composite. Show that every ideal of R is principal.

2.3 The Isomorphism Theorems For Rings

The basic ring isomorphism theorems may be proved by adapting the arguments used in Section 1.4 to prove the analogous theorems for groups. Suppose that I is an ideal of the ring R, f is a ring homomorphism from R to S with kernel K, and π is the natural map, as indicated in Figure 2.3.1. To avoid awkward analysis of special cases, let us make a blanket assumption that any time a quotient ring R_0/I_0 appears in the statement of a theorem, the ideal I_0 is proper.

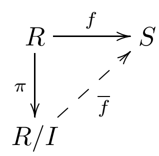


Figure 2.3.1

2.3.1 Factor Theorem For Rings

Any ring homomorphism whose kernel contains I can be factored through R/I. In other words, in Figure 2.3.1 there is a unique ring homomorphism $\overline{f}\colon R \to S$ that makes the diagram commutative. Furthermore,

(i) $\overline{f}$ is an epimorphism if and only if f is an epimorphism;

(ii) $\overline{f}$ is a monomorphism if and only if $\ker f = I$;

(iii) $\overline{f}$ is an isomorphism if and only if f is an epimorphism and $\ker f = I$.

Proof. The only possible way to define $\overline{f}$ is $\overline{f}(a+I) = f(a)$. To verify that $\overline{f}$ is well-defined, note that if $a + I = b + I$, then $a - b \in I \subseteq K$, so $f(a-b) = 0$, i.e.,$f(a) = f(b)$. Since f is a ring homomorphism, so is $\overline{f}$. To prove (i), (ii) and (iii), the discussion in (1.4.1) may be translated into additive notation and copied. ♣

2.3.2 First Isomorphism Theorem For Rings

If $f\colon R \to S$ is a ring homomorphism with kernel K, then the image of f is isomorphic to R/K.

Proof. Apply the factor theorem with $I = K$, and note that f is an epimorphism onto its image. ♣

2.3.3 Second Isomorphism Theorem For Rings

Let I be an ideal of the ring R, and let S be a subring of R. Then:

(a) $S + I (= \{x + y\colon x \in S, y \in I\})$ is a subring of R;

(b) I is an ideal of $S + I$;

(c) $S \cap I$ is an ideal of S;

(d) $(S+I)/I$ is isomorphic to $S/(S \cap I)$, as suggested by the "parallelogram" or "diamond" diagram in Figure 2.3.2.

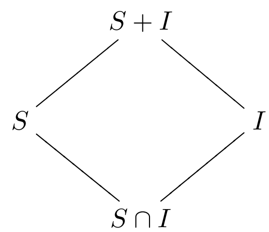


Figure 2.3.2

Proof. (a) Verify directly that $S+I$ is an additive subgroup of R that contains 1_R (since $1_R \in S$ and $0_R \in I$) and is closed under multiplication. For example, if $a \in S$, $x \in I$, $b \in S$, $y \in I$, then $(a+x)(b+y) = ab + (ay + xb + xy) \in S+I$.

(b) Since I is an ideal of R, it must be an ideal of the subring $S+I$.

(c) This follows from the definitions of subring and ideal.

(d) Let $\pi \colon R \to R/I$ be the natural map, and let π_0 be the restriction of π to S. Then π_0 is a ring homomorphism whose kernel is $S \cap I$ and whose image is $\{a+I \colon a \in S\} = (S+I)/I$. (To justify the last equality, note that if $s \in S$ and $x \in I$ we have $(s+x)+I = s+I$.) By the first isomorphism theorem for rings, $S/\ker\ \pi_0$ is isomorphic to the image of π_0, and (d) follows. ♣

2.3.4 Third Isomorphism Theorem For Rings

Let I and J be ideals of the ring R, with $I \subseteq J$. Then J/I is an ideal of R/I, and $R/J \cong (R/I)/(J/I)$.

Proof. Define $f \colon R/I \to R/J$ by $f(a+I) = a+J$. To check that f is well-defined, suppose that $a+I = b+I$. Then $a-b \in I \subseteq J$, so $a+J = b+J$. By definition of addition and multiplication of cosets in a quotient ring, f is a ring homomorphism. Now

$$\ker f = \{a+I \colon a+J = J\} = \{a+I \colon a \in J\} = J/I$$

and

$$\operatorname{im} f = \{a+J \colon a \in R\} = R/J$$

(where im denotes image). The result now follows from the first isomorphism theorem for rings. ♣

2.3.5 Correspondence Theorem For Rings

If I is an ideal of the ring R, then the map $S \to S/I$ sets up a one-to-one correspondence between the set of all subrings of R containing I and the set of all subrings of R/I, as well as a one-to-one correspondence between the set of all ideals of R containing I and the set of all ideals of R/I. The inverse of the map is $Q \to \pi^{-1}(Q)$, where π is the canonical map: $R \to R/I$.

Proof. The correspondence theorem for groups yields a one-to-one correspondence between additive subgroups of R containing I and additive subgroups of R/I. We must check that subrings correspond to subrings and ideals to ideals. If S is a subring of R then S/I is closed under addition, subtraction and multiplication. For example, if s and s' belong to S, we have $(s+I)(s'+I) = ss'+I \in S/I$. Since $1_R \in S$ we have $1_R+I \in S/I$, proving that S/I is a subring of R/I. Conversely, if S/I is a subring of R/I, then S is closed under addition, subtraction and multiplication, and contains the identity, hence is a subring or R. For example, if $s, s' \in S$ then $(s+I)(s'+I) \in S/I$, so that $ss'+I = t+I$ for some $t \in S$, and therefore $ss' - t \in I$. But $I \subseteq S$, so $ss' \in S$.

Now if J is an ideal of R containing I, then J/I is an ideal of R/I by the third isomorphism theorem for rings. Conversely, let J/I be an ideal of R/I. If $r \in R$ and $x \in J$ then $(r+I)(x+I) \in J/I$, that is, $rx + I \in J/I$. Thus for some $j \in J$ we have $rx - j \in I \subseteq J$, so $rx \in J$. A similar argument shows that $xr \in J$, and that J is an additive subgroup of R. It follows that J is an ideal of R. ♣

We now consider the Chinese remainder theorem, which is an abstract version of a result in elementary number theory. Along the way, we will see a typical application of the first isomorphism theorem for rings; in fact the development of any major theorem of algebra is likely to include an appeal to one or more of the isomorphism theorems. The following observations may make the ideas easier to visualize.

2.3.6 Definitions and Comments

(i) If a and b are integers that are congruent modulo n, then $a - b$ is a multiple of n. Thus $a - b$ belongs to the ideal I_n consisting of all multiples of n in the ring $\mathbb{Z}$ of integers. Thus we may say that a is congruent to b modulo I_n. In general, if $a, b \in R$ and I is an ideal of R, we say that $a \equiv b \bmod I$ if $a - b \in I$.

(ii) The integers a and b are relatively prime if and only if the integer 1 can be expressed as a linear combination of a and b. Equivalently, the sum of the ideals I_a and I_b is the entire ring $\mathbb{Z}$. In general, we say that the ideals I and J in the ring R are *relatively prime* if $I + J = R$.

(iii) If I_{n_i} consists of all multiples of n_i in the ring of integers $(i = 1, \dots k)$, then the intersection $\cap_{i=1}^{k} I_{n_i}$ is I_r where r is the least common multiple of the n_i. If the n_i are relatively prime in pairs, then r is the product of the n_i.

(iv) If $R_1, \dots, R_n$ are rings, the *direct product* of the R_i is defined as the ring of n-tuples $(a_1, \dots, a_n)$, $a_i \in R_i$, with componentwise addition and multiplication, that is, with

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n),$$
$$(a_1, \dots, a_n)(b_1, \dots, b_n) = (a_1 b_1, \dots, a_n b_n).$$

The zero element is $(0, \dots, 0)$ and the multiplicative identity is $(1, \dots, 1)$.

2.3.7 Chinese Remainder Theorem

Let R be an arbitrary ring, and let $I_1, \dots, I_n$ be ideals in R that are relatively prime in pairs, that is, $I_i + I_j = R$ for all $i \neq j$.

(1) If $a_1 = 1$ (the multiplicative identity of R) and $a_j = 0$ (the zero element of R) for $j = 2, \dots, n$, then there is an element $a \in R$ such that $a \equiv a_i \bmod I_i$ for all $i = 1, \dots, n$.

(2) More generally, if $a_1, \dots, a_n$ are arbitrary elements of R, there is an element $a \in R$ such that $a \equiv a_i \bmod I_i$ for all $i = 1, \dots, n$.

(3) If b is another element of R such that $b \equiv a_i \bmod I_i$ for all $i = 1, \dots, n$, then $b \equiv a \bmod I_1 \cap I_2 \cap \dots \cap I_n$. Conversely, if $b \equiv a \bmod \cap_{i=1}^n I_i$, then $b \equiv a_i \bmod I_i$ for all i.

(4) $R/\bigcap_{i=1}^n I_i$ is isomorphic to the direct product $\prod_{i=1}^n R/I_i$.

Proof. (1) If $j > 1$ we have $I_1 + I_j = R$, so there exist elements $b_j \in I_1$ and $c_j \in I_j$ such that $b_j + c_j = 1$; thus

$$\prod_{j=2}^n (b_j + c_j) = 1.$$

Expand the left side and observe that any product containing at least one b_j belongs to I_1, while $c_2 \dots c_n$ belongs to $\prod_{j=2}^n I_j$, the collection of all finite sums of products $x_2 \dots x_n$ with $x_j \in I_j$. Thus we have elements $b \in I_1$ and $a \in \prod_{j=2}^n I_j$ (a subset of each I_j) with $b + a = 1$. Consequently, $a \equiv 1 \bmod I_1$ and $a \equiv 0 \bmod I_j$ for $j > 1$, as desired.

(2) By the argument of part (1), for each i we can find c_i with $c_i \equiv 1 \bmod I_i$ and $c_i \equiv 0 \bmod I_j, j \neq i$. If $a = a_1 c_1 + \dots + a_n c_n$, then a has the desired properties. To see this, write $a - a_i = a - a_i c_i + a_i(c_i - 1)$, and note that $a - a_i c_i$ is the sum of the $a_j c_j, j \neq i$, and is therefore congruent to 0 mod I_i.

(3) We have $b \equiv a_i \bmod I_i$ for all i iff $b - a \equiv 0 \bmod I_i$ for all i, that is, iff $b - a \in \cap_{i=1}^n I_i$, and the result follows.

(4) Define $f\colon R \to \prod_{i=1}^n R/I_i$ by $f(a) = (a + I_1, \dots, a + I_n)$. If $a_1, \dots, a_n \in R$, then by part (2) there is an element $a \in R$ such that $a \equiv a_i \bmod I_i$ for all i. But then $f(a) = (a_1 + I_1, \dots, a_n + I_n)$, proving that f is surjective. Since the kernel of f is the intersection of the ideals I_j, the result follows from the first isomorphism theorem for rings. ♣

The concrete version of the Chinese remainder theorem can be recovered from the abstract result; see Problems 3 and 4.

Problems For Section 2.3

1. Show that the group isomorphisms of Section 1.4, Problems 1 and 2, are ring isomorphisms as well.
2. Give an example of an ideal that is not a subring, and a subring that is not an ideal.
3. If the integers $m_i, i = 1, \dots, n$, are relatively prime in pairs, and $a_1, \dots, a_n$ are arbitrary integers, show that there is an integer a such that $a \equiv a_i \bmod m_i$ for all i, and that any two such integers are congruent modulo $m_1 \dots m_n$.
4. If the integers $m_i, i = 1, \dots, n$, are relatively prime in pairs and $m = m_1 \dots m_n$, show that there is a ring isomorphism between $\mathbb{Z}_m$ and the direct product $\prod_{i=1}^n \mathbb{Z}_{m_i}$. Specifically, $a \bmod m$ corresponds to $(a \bmod m_1, \dots, a \bmod m_n)$.
5. Suppose that $R = R_1 \times R_2$ is a direct product of rings. Let R_1' be the ideal $R_1 \times \{0\} = \{(r_1, 0)\colon r_1 \in R_1\}$, and let R_2' be the ideal $\{(0, r_2)\colon r_2 \in R_2\}$. Show that $R/R_1' \cong R_2$ and $R/R_2' \cong R_1$.
6. If $I_1, \dots, I_n$ are ideals, the *product* $I_1 \dots I_n$ is defined as the set of all finite sums $\sum_i a_{1i}a_{2i} \dots a_{ni}$, where $a_{ki} \in I_k$, $k = 1, \dots, n$. [See the proof of part (1) of 2.3.7; a brief check shows that the product of ideals is an ideal.]

   Assume that R is a commutative ring. Under the hypothesis of the Chinese remainder theorem, show that the intersection of the ideals I_i coincides with their product.
7. Let $I_1, \dots, I_n$ be ideals in the ring R. Suppose that $R/\cap_i I_i$ is isomorphic to $\prod_i R/I_i$ via $a + \cap_i I_i \to (a + I_1, \dots, a + I_n)$. Show that the ideals I_i are relatively prime in pairs.

2.4 Maximal and Prime Ideals

If I is an ideal of the ring R, we might ask"What is the smallest ideal containing I" and "What is the largest ideal containing I". Neither of these questions is challenging; the smallest ideal is I itself, and the largest ideal is R. But if I is a proper ideal and we ask for a maximal proper ideal containing I, the question is much more interesting.

2.4.1 Definition

A *maximal ideal* in the ring R is a proper ideal that is not contained in any strictly larger proper ideal.

2.4.2 Theorem

Every proper ideal I of the ring R is contained in a maximal ideal. Consequently, every ring has at least one maximal ideal.

Proof. The argument is a prototypical application of Zorn's lemma. Consider the collection of all proper ideals containing I, partially ordered by inclusion. Every chain $\{J_t, t \in T\}$ of proper ideals containing I has an upper bound, namely the union of the chain. (Note that the union is still a proper ideal, because the identity 1_R belongs to none of the ideals J_t.) By Zorn, there is a maximal element in the collection, that is, a maximal ideal containing I. Now take $I = \{0\}$ to conclude that every ring has at least one maximal ideal. ♣

We have the following characterization of maximal ideals.

2.4.3 Theorem

Let M be an ideal in the commutative ring R. Then M is a maximal ideal if and only if R/M is a field.

Proof. Suppose M is maximal. We know that R/M is a ring (see 2.2.4); we need to find the multiplicative inverse of the element $a + M$ of R/M, where $a + M$ is not the zero element, i.e., $a \notin M$. Since M is maximal, the ideal $Ra + M$, which contains a and is therefore strictly larger than M, must be the ring R itself. Consequently, the identity element 1 belongs to $Ra + M$. If $1 = ra + m$ where $r \in R$ and $m \in M$, then

$$(r + M)(a + M) = ra + M = (1 - m) + M = 1 + M \text{ (since } m \in M),$$

proving that $r + M$ is the multiplicative inverse of $a + M$.

Conversely, if R/M is a field, then M must be a proper ideal. If not, then $M = R$, so that R/M contains only one element, contradicting the requirement that $1 \neq 0$ in R/M (see (7) of 2.1.1). By (2.2.6), the only ideals of R/M are $\{0\}$ and R/M, so by the correspondence theorem 2.3.5, there are no ideals properly between M and R. Therefore M is a maximal ideal. ♣

If in (2.4.3) we relax the requirement that R/M be a field, we can identify another class of ideals.

2.4.4 Definition

A *prime ideal* in a commutative ring R is a proper ideal P such that for any two elements a, b in R,

$$ab \in P \text{ implies that } a \in P \text{ or } b \in P.$$

We can motivate the definition by looking at the ideal (p) in the ring of integers. In this case, $a \in (p)$ means that p divides a, so that (p) will be a prime ideal if and only if

$$p \text{ divides } ab \text{ implies that } p \text{ divides } a \text{ or } p \text{ divides } b,$$

which is equivalent to the requirement that p be a prime number.

2.4.5 Theorem

If P is an ideal in the commutative ring R, then P is a prime ideal if and only if R/P is an integral domain.

Proof. Suppose P is prime. Since P is a proper ideal, R/P is a ring. We must show that if $(a+P)(b+P)$ is the zero element P in R/P, then $a+P=P$ or $b+P=P$, i.e., $a \in P$ or $b \in P$. This is precisely the definition of a prime ideal.

Conversely, if R/P is an integral domain, then, as in (2.4.3), P is a proper ideal. If $ab \in P$, then $(a+P)(b+P)$ is zero in R/P, so that $a+P=P$ or $b+P=P$, i.e., $a \in P$ or $b \in P$. ♣

2.4.6 Corollary

In a commutative ring, a maximal ideal is prime.

Proof. This is immediate from (2.4.3) and (2.4.5). ♣

2.4.7 Corollary

Let $f\colon R \to S$ be an epimorphism of commutative rings. Then:

(i) If S is a field then $\ker f$ is a maximal ideal of R;

(ii) If S is an integral domain then $\ker f$ is a prime ideal of R.

Proof. By the first isomorphism theorem (2.3.2), S is isomorphic to $R/\ker f$, and the result now follows from (2.4.3) and (2.4.5). ♣

2.4.8 Example

Let $\mathbb{Z}[X]$ be the set of all polynomials $f(X) = a_0 + a_1X + \cdots + a_nX^n, n = 0, 1, \ldots$ in the indeterminate X, with integer coefficients. The ideal generated by X, that is, the collection of all multiples of X, is

$$\langle X \rangle = \{f(X) \in \mathbb{Z}[X]\colon a_0 = 0\}.$$

The ideal generated by 2 is

$$\langle 2 \rangle = \{f(X) \in \mathbb{Z}[X]\colon \text{all } a_i \text{ are even integers.}\}$$

Both $\langle X \rangle$ and $\langle 2 \rangle$ are proper ideals, since $2 \notin \langle X \rangle$ and $X \notin \langle 2 \rangle$. In fact we can say much more. Consider the ring homomorphisms $\varphi\colon \mathbb{Z}[X] \to \mathbb{Z}$ and $\psi\colon \mathbb{Z}[X] \to \mathbb{Z}_2$ given by $\varphi(f(X)) = a_0$ and $\psi(f(X)) = \overline{a}_0$, where $\overline{a}_0$ is a_0 reduced modulo 2. We will show that both $\langle X \rangle$ and $\langle 2 \rangle$ are prime ideals that are not maximal.

First note that by (2.4.7), $\langle X \rangle$ is prime because it is the kernel of φ. Then observe that $\langle X \rangle$ is not maximal because it is properly contained in $\langle 2, X \rangle$, the ideal generated by 2 and X.

To verify that $\langle 2 \rangle$ is prime, note that it is the kernel of the homomorphism from $\mathbb{Z}[X]$ onto $\mathbb{Z}_2[X]$ that takes $f(X)$ to $\overline{f}(X)$, where the overbar indicates that the coefficients of $f(X)$ are to be reduced modulo 2. Since $\mathbb{Z}_2[X]$ is an integral domain (see the comment at the end of 2.1.3), $\langle 2 \rangle$ is a prime ideal. Since $\langle 2 \rangle$ is properly contained in $\langle 2, X \rangle$, $\langle 2 \rangle$ is not maximal.

Finally, $\langle 2, X\rangle$ is a maximal ideal, since

$$\ker\psi = \{a_0 + Xg(X)\colon a_0 \text{ is even and } g(X) \in \mathbb{Z}[X]\} = \langle 2, X\rangle.$$

Thus $\langle 2, X\rangle$ is the kernel of a homomorphism onto a field, and the result follows from (2.4.7).

2.4.9 Problems For Section 2.4

1. We know from Problem 1 of Section 2.2 that in the ring of integers, all ideals I are of the form $\langle n\rangle$ for some $n \in \mathbb{Z}$, and since $n \in I$ implies $-n \in I$, we may take n to be nonnegative. Let $\langle n\rangle$ be a nontrivial ideal, so that n is a positive integer greater than 1. Show that $\langle n\rangle$ is a prime ideal if and only if n is a prime number.
2. Let I be a nontrivial prime ideal in the ring of integers. Show that in fact I must be maximal.
3. Let $F[[X]]$ be the ring of formal power series with coefficients in the field F (see (2.1.3), Example 6). Show that $\langle X\rangle$ is a maximal ideal.
4. Perhaps the result of Problem 3 is a bit puzzling. Why can't we argue that just as in (2.4.8), $\langle X\rangle$ is properly contained in $\langle 2, X\rangle$, and therefore $\langle X\rangle$ is not maximal?
5. Let I be a proper ideal of $F[[X]]$. Show that $I \subseteq \langle X\rangle$, so that $\langle X\rangle$ is the unique maximal ideal of $F[[X]]$. (A commutative ring with a unique maximal ideal is called a *local ring.*)
6. Show that every ideal of $F[[X]]$ is principal, and specifically every nonzero ideal is of the form (X^n) for some $n = 0, 1, \ldots$.
7. Let $f\colon R \to S$ be a ring homomorphism, with R and S commutative. If P is a prime ideal of S, show that the preimage $f^{-1}(P)$ is a prime ideal of R.
8. Show that the result of Problem 7 does not hold in general when P is a maximal ideal.
9. Show that a prime ideal P cannot be the intersection of two strictly larger ideals I and J.

2.5 Polynomial Rings

In this section, *all rings are assumed commutative.* To see a good reason for this restriction, consider the *evaluation map* (also called the *substitution map*) E_x, where x is a fixed element of the ring R. This map assigns to the polynomial $a_0 + a_1X + \cdots + a_nX^n$ in $R[X]$ the value $a_0 + a_1x + \cdots + a_nx^n$ in R. It is tempting to say that "obviously", E_x is a ring homomorphism, but we must be careful. For example,

$$E_x[(a + bX)(c + dX)] = E_x(ac + (ad + bc)X + bdX^2) = ac + (ad + bc)x + bdx^2,$$
$$E_x(a + bX)E_x(c + dX) = (a + bx)(c + dx) = ac + adx + bxc + bxdx,$$

and these need not be equal if R is not commutative.

The *degree*, abbreviated deg, of a polynomial $a_0 + a_1X + \cdots + a_nX^n$ (with *leading coefficient* $a_n \neq 0$) is n; it is convenient to define the degree of the zero polynomial as $-\infty$. If f and g are polynomials in $R[X]$, where R is a *field*, ordinary long division allows us to express f as $qg + r$, where the degree of r is less than the degree of g. We have a similar result over an arbitrary commutative ring, if g is *monic*, i.e., the leading coefficient of g is 1. For example (with $R = \mathbb{Z}$), we can divide $2X^3 + 10X^2 + 16X + 10$ by $X^2 + 3X + 5$:

$$2X^3 + 10X^2 + 16X + 10 = 2X(X^2 + 3X + 5) + 4X^2 + 6X + 10.$$

The remainder $4X^2 + 6X + 10$ does not have degree less than 2, so we divide it by $X^2 + 3X + 5$:

$$4X^2 + 6X + 10 = 4(X^2 + 3X + 5) - 6X - 10.$$

Combining the two calculations, we have

$$2X^3 + 10X^2 + 16X + 10 = (2X + 4)(X^2 + 3X + 5) + (-6X - 10)$$

which is the desired decomposition.

2.5.1 Division Algorithm

If f and g are polynomials in $R[X]$, with g monic, there are unique polynomials q and r in $R[X]$ such that $f = qg + r$ and deg $r <$ deg g. If R is a field, g can be any nonzero polynomial.

Proof. The above procedure, which works in any ring R, shows that q and r exist.If $f = qg+r = q_1g+r_1$ where r and r_1 are of degree less than deg g, then $g(q-q_1) = r_1 - r$. But if $q - q_1 \neq 0$, then, since g is monic, the degree of the left side is at least deg g, while the degree of the right side is less than deg g, a contradiction. Therefore $q = q_1$, and consequently $r = r_1$. ♣

2.5.2 Remainder Theorem

If $f \in R[X]$ and $a \in R$, then for some unique polynomial $q(X)$ in $R[X]$ we have

$$f(X) = q(X)(X - a) + f(a);$$

hence $f(a) = 0$ if and only if $X - a$ divides $f(X)$.

Proof. By the division algorithm, we may write $f(X) = q(X)(X - a) + r(X)$ where the degree of r is less than 1, i.e., r is a constant. Apply the evaluation homomorphism $X \to a$ to show that $r = f(a)$. ♣

2.5.3 Theorem

If R is an integral domain, then a nonzero polynomial f in $R[X]$ of degree n has at most n roots in R, counting multiplicity.

Proof. If $f(a_1) = 0$, then by (2.5.2), possibly applied several times, we have $f(X) = q_1(X)(X - a_1)^{n_1}$, where $q_1(a_1) \neq 0$ and the degree of q_1 is $n - n_1$. If a_2 is another root of f, then $0 = f(a_2) = q_1(a_2)(a_2 - a_1)^{n_1}$. But $a_1 \neq a_2$ and R is an integral domain, so $q_1(a_2)$ must be 0, i.e. a_2 is a root of $q_1(X)$. Repeating the argument, we have $q_1(X) = q_2(X)(X-a_2)^{n_2}$, where $q_2(a_2) \neq 0$ and $\deg q_2 = n-n_1-n_2$. After n applications of (2.5.2), the quotient becomes constant, and we have $f(X) = c(X-a_1)^{n_1} \dots (X-a_k)^{n_k}$ where $c \in R$ and $n_1 + \cdots + n_k = n$. Since R is an integral domain, the only possible roots of f are $a_1, \dots, a_k$. ♣

2.5.4 Example

Let $R = \mathbb{Z}_8$, which is not an integral domain. The polynomial $f(X) = X^3$ has four roots in R, namely 0, 2, 4 and 6.

Problems For Section 2.5

In Problems 1-4, we review the Euclidean algorithm. Let a and b be positive integers, with $a > b$. Divide a by b to obtain

$$a = bq_1 + r_1 \text{ with } 0 \leq r_1 < b,$$

then divide b by r_1 to get

$$b = r_1 q_2 + r_2 \text{ with } 0 \leq r_2 < r_1,$$

and continue in this fashion until the process terminates:

$$\begin{aligned} r_1 &= r_2 q_3 + r_3,\ 0 \leq r_3 < r_2, \\ &\vdots \\ r_{j-2} &= r_{j-1} q_j + r_j,\ 0 \leq r_j < r_{j-1}, \\ r_{j-1} &= r_j q_{j+1} \end{aligned}$$

1. Show that the greatest common divisor of a and b is the last remainder r_j.
2. If d is the greatest common divisor of a and b, show that there are integers x and y such that $ax + by = d$.
3. Define three sequences by

$$\begin{aligned} r_i &= r_{i-2} - q_i r_{i-1} \\ x_i &= x_{i-2} - q_i x_{i-1} \\ y_i &= y_{i-2} - q_i y_{i-1} \end{aligned}$$

for $i = -1, 0, 1, \ldots$ with initial conditions $r_{-1} = a$, $r_0 = b$, $x_{-1} = 1$, $x_0 = 0$, $y_{-1} = 0$, $y_0 = 1$. (The q_i are determined by dividing r_{i-2} by r_{i-1}.) Show that we can generate all steps of the algorithm, and at each stage, $r_i = ax_i + by_i$.

4. Use the procedure of Problem 3 (or any other method) to find the greatest common divisor d of $a = 123$ and $b = 54$, and find integers x and y such that $ax + by = d$.
5. Use Problem 2 to show that $\mathbb{Z}_p$ is a field if and only if p is prime.
6. If $a(X)$ and $b(X)$ are polynomials with coefficients in a field F, the Euclidean algorithm can be used to find their greatest common divisor. The previous discussion can be taken over verbatim, except that instead of writing

$$a = q_1 b + r_1 \text{ with } 0 \leq r_1 < b,$$

we write

$$a(X) = q_1(X)b(X) + r_1(X) \text{ with } \deg r_1(X) < \deg b(X).$$

The greatest common divisor can be defined as the monic polynomial of highest degree that divides both $a(X)$ and $b(X)$.

Let $f(X)$ and $g(X)$ be polynomials in $F[X]$, where F is a field. Show that the ideal I generated by $f(X)$ and $g(X)$, i.e., the set of all linear combinations $a(X)f(X) + b(X)g(X)$, with $a(X), b(X) \in F[X]$, is the principal ideal $J = \langle d(X) \rangle$ generated by the greatest common divisor $d(X)$ of $f(X)$ and $g(X)$.

7. (*Lagrange Interpolation Formula*) Let $a_0, a_1, \ldots, a_n$ be distinct points in the field F, and define

$$P_i(X) = \prod_{j \neq i} \frac{X - a_j}{a_i - a_j}, \quad i = 0, 1, \ldots, n;$$

then $P_i(a_i) = 1$ and $P_i(a_j) = 0$ for $j \neq i$. If $b_0, b_1, \ldots, b_n$ are arbitrary elements of F (not necessarily distinct), use the P_i to find a polynomial $f(X)$ of degree n or less such that $f(a_i) = b_i$ for all i.

8. In Problem 7, show that $f(X)$ is the unique polynomial of degree n or less such that $f(a_i) = b_i$ for all i.
9. Suppose that f is a polynomial in $F[X]$, where F is a field. If $f(a) = 0$ for every $a \in F$, it does not in general follow that f is the zero polynomial. Give an example.
10. Give an example of a field F for which it does follow that $f = 0$.

2.6 Unique Factorization

If we are asked to find the greatest common divisor of two integers, say 72 and 60, one method is to express each integer as a product of primes; thus $72 = 2^3 \times 3^2$, $60 = 2^2 \times 3 \times 5$. The greatest common divisor is the product of terms of the form p^e, where for each prime appearing in the factorization, we use the minimum exponent. Thus $\gcd(72, 60) = 2^2 \times 3^1 \times 5^0 = 12$. To find the least common multiple, we use the maximum exponent: $\text{lcm}(72, 60) = 2^3 \times 3^2 \times 5^1 = 360$. The key idea is that every integer (except 0, 1 and -1) can be uniquely represented as a product of primes. It is natural to ask whether there are integral domains other than the integers in which unique factorization is possible. We now begin to study this question; throughout this section, *all rings are assumed to be integral domains.*

2.6.1 Definitions

Recall from (2.1.2) that a *unit* in a ring R is an element with a multiplicative inverse. The elements a and b are *associates* if $a = ub$ for some unit u.

Let a be a nonzero nonunit; a is said to be *irreducible* if it cannot be represented as a product of nonunits. In other words, if $a = bc$, then either b or c must be a unit.

Again let a be a nonzero nonunit; a is said to be *prime* if whenever a divides a product of terms, it must divide one of the factors. In other words, if a divides bc, then a divides b or a divides c (a divides b means that $b = ar$ for some $r \in R$). It follows from the definition that if p is any nonzero element of R, then p is prime if and only if $\langle p \rangle$ is a prime ideal.

The units of $\mathbb{Z}$ are 1 and -1, and the irreducible and the prime elements coincide. But these properties are not the same in an arbitrary integral domain.

2.6.2 Proposition

If a is prime, then a is irreducible, but not conversely.

Proof. We use the standard notation $r|s$ to indicate that r divides s. Suppose that a is prime, and that $a = bc$. Then certainly $a|bc$, so by definition of prime, $a|b$ or $a|c$, say $a \mid b$. If $b = ad$ then $b = bcd$, so $cd = 1$ and therefore c is a unit. (Note that b cannot be 0, for if so, $a = bc = 0$, which is not possible since a is prime.) Similarly, if $a|c$ with $c = ad$ then $c = bcd$, so $bd = 1$ and b is a unit. Therefore a is irreducible.

To give an example of an irreducible element that is not prime, consider $R = \mathbb{Z}[\sqrt{-3}\,] = \{a + ib\sqrt{3} \colon a, b \in \mathbb{Z}\}$; in R, 2 is irreducible but not prime. To see this, first suppose that we have a factorization of the form

$$2 = (a + ib\sqrt{3}\,)(c + id\sqrt{3}\,);$$

take complex conjugates to get

$$2 = (a - ib\sqrt{3}\,)(c - id\sqrt{3}\,).$$

Now multiply these two equations to obtain

$$4 = (a^2 + 3b^2)(c^2 + 3d^2).$$

Each factor on the right must be a divisor of 4, and there is no way that $a^2 + 3b^2$ can be 2. Thus one of the factors must be 4 and the other must be 1. If, say, $a^2 + 3b^2 = 1$, then $a = \pm 1$ and $b = 0$. Thus in the original factorization of 2, one of the factors must be a unit, so 2 is irreducible. Finally, 2 divides the product $(1 + i\sqrt{3}\,)(1 - i\sqrt{3}\,)$ $(= 4)$, so if 2 were prime, it would divide one of the factors, which means that 2 divides 1, a contradiction since $1/2$ is not an integer. ♣

The distinction between irreducible and prime elements disappears in the presence of unique factorization.

2.6.3 Definition

A *unique factorization domain* (UFD) is an integral domain R satisfying the following properties:

(UF1) Every nonzero element a in R can be expressed as $a = up_1 \dots p_n$, where u is a unit and the p_i are irreducible.

(UF2): If a has another factorization, say $a = vq_1 \dots q_m$, where v is a unit and the q_i are irreducible, then $n = m$ and, after reordering if necessary, p_i and q_i are associates for each i.

Property UF1 asserts the existence of a factorization into irreducibles, and UF2 asserts uniqueness.

2.6.4 Proposition

In a unique factorization domain, a is irreducible if and only if a is prime.

Proof. By (2.6.2), prime implies irreducible, so assume a irreducible, and let a divide bc. Then we have $ad = bc$ for some $d \in R$. We factor d, b and c into irreducibles to obtain

$$aud_1 \dots d_r = vb_1 \dots b_s wc_1 \dots c_t$$

where u, v and w are units and the d_i, b_i and c_i are irreducible. By uniqueness of factorization, a, which is irreducible, must be an associate of some b_i or c_i. Thus a divides b or a divides c. ♣

2.6.5 Definitions and Comments

Let A be a nonempty subset of R, with $0 \notin A$. The element d is a *greatest common divisor* (gcd) of A if d divides each a in A, and whenever e divides each a in A, we have $e|d$.

If d' is another gcd of A, we have $d|d'$ and $d'|d$, so that d and d' are associates. We will allow ourselves to speak of "the" greatest common divisor, suppressing but not forgetting that the gcd is determined up to multiplication by a unit.

The elements of A are said to be *relatively prime* (or the set A is said to be relatively prime) if 1 is a greatest common divisor of A.

The nonzero element m is a *least common multiple* (lcm) of A if each a in A divides m, and whenever $a|e$ for each a in A, we have $m|e$.

Greatest common divisors and least common multiples always exist for finite subsets of a UFD; they may be found by the technique discussed at the beginning of this section.

We will often use the fact that for any $a, b \in R$, we have $a|b$ if and only if $\langle b \rangle \subseteq \langle a \rangle$. This follows because $a|b$ means that $b = ac$ for some $c \in R$. For short, *divides means contains.*

It would be useful to be able to recognize when an integral domain is a UFD. The following criterion is quite abstract, but it will help us to generate some explicit examples.

2.6.6 Theorem

Let R be an integral domain.

(1) If R is a UFD then R satisfies the *ascending chain condition* (acc) *on principal ideals*: If $a_1, a_2, \dots$ belong to R and $\langle a_1 \rangle \subseteq \langle a_2 \rangle \subseteq \dots$, then the sequence eventually stabilizes, that is, for some n we have $\langle a_n \rangle = \langle a_{n+1} \rangle = \langle a_{n+2} \rangle = \dots$.

(2) If R satisfies the ascending chain condition on principal ideals, then R satisfies UF1, that is, every nonzero element of R can be factored into irreducibles.

(3) If R satisfies UF1 and in addition, every irreducible element of R is prime, then R is a UFD.

Thus R is a UFD if and only if R satisfies the ascending chain condition on principal ideals and every irreducible element of R is prime.

Proof. (1) If $\langle a_1 \rangle \subseteq \langle a_2 \rangle \subseteq \dots$ then $a_{i+1} | a_i$ for all i. Therefore the prime factors of a_{i+1} consist of some (or all) of the prime factors of a_i. Multiplicity is taken into account here; for example, if p^3 is a factor of a_i, then p^k will be a factor of a_{i+1} for some $k \in \{0, 1, 2, 3\}$. Since a_1 has only finitely many prime factors, there will come a time when the prime factors are the same from that point on, that is, $\langle a_n \rangle = \langle a_{n+1} \rangle = \dots$ for some n.

(2) Let a_1 be any nonzero element. If a_1 is irreducible, we are finished, so let $a_1 = a_2 b_2$ where neither a_2 nor b_2 is a unit. If both a_2 and b_2 are irreducible, we are finished, so we can assume that one of them, say a_2, is not irreducible. Since a_2 divides a_1 we have $\langle a_1 \rangle \subseteq \langle a_2 \rangle$, and in fact the inclusion is proper because $a_2 \notin \langle a_1 \rangle$. (If $a_2 = c a_1$ then $a_1 = a_2 b_2 = c a_1 b_2$, so b_2 is a unit, a contradiction.) Continuing, we have $a_2 = a_3 b_3$ where neither a_3 nor b_3 is a unit, and if, say, a_3 is not irreducible, we find that $\langle a_2 \rangle \subset \langle a_3 \rangle$. If a_1 cannot be factored into irreducibles, we obtain, by an inductive argument, a strictly increasing chain $\langle a_1 \rangle \subset \langle a_2 \rangle \subset \dots$ of principal ideals.

(3) Suppose that $a = u p_1 p_2 \dots p_n = v q_1 q_2 \dots q_m$ where the p_i and q_i are irreducible and u and v are units. Then p_1 is a prime divisor of $v q_1 \dots q_m$, so p_1 divides one of the q_i, say q_1. But q_1 is irreducible, and therefore p_1 and q_1 are associates. Thus we have, up to multiplication by units, $p_2 \dots p_n = q_2 \dots q_m$. By an inductive argument, we must have $m = n$, and after reordering, p_i and q_i are associates for each i. ♣

We now give a basic sufficient condition for an integral domain to be a UFD.

2.6.7 Definition

A *principal ideal domain* (PID) is an integral domain in which every ideal is principal, that is, generated by a single element.

2.6.8 Theorem

Every principal ideal domain is a unique factorization domain. For short, PID implies UFD.

Proof. If$\langle a_1\rangle \subseteq \langle a_2\rangle \subseteq \dots$, let $I = \cup_i \langle a_i\rangle$. Then I is an ideal, necessarily principal by hypothesis. If $I = \langle b\rangle$ then b belongs to some $\langle a_n\rangle$, so $I \subseteq \langle a_n\rangle$. Thus if $i \geq n$ we have $\langle a_i\rangle \subseteq I \subseteq \langle a_n\rangle \subseteq \langle a_i\rangle$. Therefore $\langle a_i\rangle = \langle a_n\rangle$ for all $i \geq n$, so that R satisfies the acc on principal ideals.

Now suppose that a is irreducible. Then $\langle a\rangle$ is a proper ideal, for if $\langle a\rangle = R$ then $1 \in \langle a\rangle$, so that a is a unit. By the acc on principal ideals, $\langle a\rangle$ is contained in a maximal ideal I. (Note that we need not appeal to the general result (2.4.2), which uses Zorn's lemma.) If $I = \langle b\rangle$ then b divides the irreducible element a, and b is not a unit since I is proper. Thus a and b are associates, so $\langle a\rangle = \langle b\rangle = I$. But I, a maximal ideal, is prime by (2.4.6), hence a is prime. The result follows from (2.6.6). ♣

The following result gives a criterion for a UFD to be a PID. (Terminology: the *zero ideal* is $\{0\}$; a *nonzero ideal* is one that is not $\{0\}$.)

2.6.9 Theorem

R is a PID if and only if R is a UFD and every nonzero prime ideal of R is maximal.

Proof. Assume R is a PID; then R is a UFD by (2.6.8). If $\langle p\rangle$ is a nonzero prime ideal of R, then $\langle p\rangle$ is contained in the maximal ideal $\langle q\rangle$, so that q divides the prime p. Since a maximal ideal must be proper, q cannot be a unit, so that p and q are associates. But then $\langle p\rangle = \langle q\rangle$ and $\langle p\rangle$ is maximal.

The proof of the converse is given in the exercises. ♣

Problems For Section 2.6

Problems 1-6 form a project designed to prove that if R is a UFD and every nonzero prime ideal of R is maximal, then R is a PID.

Let I be an ideal of R; since $\{0\}$ is principal, we can assume that $I \neq \{0\}$. Since R is a UFD, every nonzero element of I can be written as $up_1 \dots p_t$ where u is a unit and the p_i are irreducible, hence prime. Let $r = r(I)$ be the minimum such t. We are going to prove by induction on r that I is principal.

1. If $r = 0$, show that $I = \langle 1\rangle = R$.
2. If the result holds for all $r < n$, let $r = n$, with $up_1 \dots p_n \in I$, hence $p_1 \dots p_n \in I$. Since p_1 is prime, $\langle p_1\rangle$ is a prime ideal, necessarily maximal by hypothesis. By (2.4.3), $R/\langle p_1\rangle$ is a field. If b belongs to I but not to $\langle p_1\rangle$, show that for some $c \in R$ we have $bc - 1 \in \langle p_1\rangle$.
3. By Problem 2, $bc - dp_1 = 1$ for some $d \in R$. Show that this implies that $p_2 \dots p_n \in I$, which contradicts the minimality of n. Thus if b belongs to I, it must also belong to $\langle p_1\rangle$, that is, $I \subseteq \langle p_1\rangle$.
4. Let $J = \{x \in R \colon xp_1 \in I\}$. Show that J is an ideal.
5. Show that $Jp_1 = I$.
6. Since $p_1 \dots p_n = (p_2 \dots p_n)p_1 \in I$, we have $p_2 \dots p_n \in J$. Use the induction hypothesis to conclude that I is principal.

7. Let p and q be prime elements in the integral domain R, and let $P = \langle p \rangle$ and $Q = \langle q \rangle$ be the corresponding prime ideals. Show that it is not possible for P to be a proper subset of Q.
8. If R is a UFD and P is a nonzero prime ideal of R, show that P contains a nonzero principal prime ideal.

2.7 Principal Ideal Domains and Euclidean Domains

In Section 2.6, we found that a principal ideal domain is a unique factorization domain, and this exhibits a class of rings in which unique factorization occurs. We now study some properties of PID's, and show that any integral domain in which the Euclidean algorithm works is a PID. If I is an ideal in $\mathbb{Z}$, in fact if I is simply an additive subgroup of $\mathbb{Z}$, then I consists of all multiples of some positive integer n; see Section 1.1, Problem 6. Thus $\mathbb{Z}$ is a PID.

Now suppose that A is a nonempty subset of the PID R. The ideal $\langle A \rangle$ generated by A consists of all finite sums $\sum r_i a_i$ with $r_i \in R$ and $a_i \in A$; see (2.2.7). We show that if d is a greatest common divisor of A, then d generates A, and conversely.

2.7.1 Proposition

Let R be a PID, with A a nonempty subset of R. Then d is a greatest common divisor of A if and only if d is a generator of $\langle A \rangle$.

Proof. Let d be a gcd of A, and assume that $\langle A \rangle = \langle b \rangle$. Then d divides every $a \in A$, so d divides all finite sums $\sum r_i a_i$. In particular d divides b, hence $\langle b \rangle \subseteq \langle d \rangle$; that is, $\langle A \rangle \subseteq \langle d \rangle$. But if $a \in A$ then $a \in \langle b \rangle$, so b divides a. Since d is a gcd of A, it follows that b divides d, so $\langle d \rangle$ is contained in $\langle b \rangle = \langle A \rangle$. We conclude that $\langle A \rangle = \langle d \rangle$, proving that d is a generator of $\langle A \rangle$.

Conversely, assume that d generates $\langle A \rangle$. If $a \in A$ then a is a multiple of d, so $d \mid a$. By (2.2.7), d can be expressed as $\sum r_i a_i$, so any element that divides everything in A divides d. Therefore d is a gcd of A. ♣

2.7.2 Corollary

If d is a gcd of A, where A is a nonempty subset of the PID R, then d can be expressed as a finite linear combination $\sum r_i a_i$ of elements of A with coefficients in R.

Proof. By (2.7.1), $d \in \langle A \rangle$, and the result follows from (2.2.7). ♣

As a special case, we have the familiar result that the greatest common divisor of two integers a and b can be expressed as $ax + by$ for some integers x and y.

The Euclidean algorithm in $\mathbb{Z}$ is based on the division algorithm: if a and b are integers and $b \neq 0$, then a can be divided by b to produce a quotient and remainder. Specifically, we have $a = bq + r$ for some $q, r \in \mathbb{Z}$ with $|r| < |b|$. The Euclidean algorithm performs equally well for polynomials with coefficients in a field; the absolute value of an integer is replaced by the degree of a polynomial. It is possible to isolate the key property that makes the Euclidean algorithm work.

2.7.3 Definition

Let R be an integral domain. R is said to be a *Euclidean domain* (ED) if there is a function Ψ from $R \setminus \{0\}$ to the nonnegative integers satisfying the following property:

If a and b are elements of R, with $b \neq 0$, then a can be expressed as $bq + r$ for some $q, r \in R$, where either $r = 0$ or $\Psi(r) < \Psi(b)$.

We can replace"$r = 0$ or $\Psi(r) < \Psi(b)$" by simply "$\Psi(r) < \Psi(b)$" if we define $\Psi(0)$ to be $-\infty$.

In any Euclidean domain, we may use the Euclidean algorithm to find the greatest common divisor of two elements; see the problems in Section 2.5 for a discussion of the procedure in $\mathbb{Z}$ and in $F[X]$, where F is a field.

A Euclidean domain is automatically a principal ideal domain, as we now prove.

2.7.4 Theorem

If R is a Euclidean domain, then R is a principal ideal domain. For short, ED implies PID.

Proof. Let I be an ideal of R. If $I = \{0\}$ then I is principal, so assume $I \neq \{0\}$. Then $\{\Psi(b)\colon b \in I, b \neq 0\}$ is a nonempty set of nonnegative integers, and therefore has a smallest element n. Let b be any nonzero element of I such that $\Psi(b) = n$; we claim that $I = \langle b \rangle$. For if a belongs to I then we have $a = bq + r$ where $r = 0$ or $\Psi(r) < \Psi(b)$. Now $r = a - bq \in I$ (because a and b belong to I), so if $r \neq 0$ then $\Psi(r) < \Psi(b)$ is impossible by minimality of $\Psi(b)$. Thus b is a generator of I. ♣

The most familiar Euclidean domains are $\mathbb{Z}$ and $F[X]$, with F a field. We now examine some less familiar cases.

2.7.5 Example

Let $\mathbb{Z}[\sqrt{d}\,]$ be the ring of all elements $a + b\sqrt{d}$, where $a, b \in \mathbb{Z}$. If $d = -2, -1, 2$ or 3, we claim that $\mathbb{Z}[\sqrt{d}\,]$ is a Euclidean domain with

$$\Psi(a + b\sqrt{d}\,) = |a^2 - db^2|.$$

Since the $a + b\sqrt{d}$ are real or complex numbers, there are no zero divisors, and $\mathbb{Z}[\sqrt{d}\,]$ is an integral domain. Let $\alpha, \beta \in \mathbb{Z}[\sqrt{d}\,], \beta \neq 0$, and divide α by β to get $x + y\sqrt{d}$. Unfortunately, x and y need not be integers, but at least they are rational numbers. We can find integers reasonably close to x and y; let x_0 and y_0 be integers such that $|x - x_0|$ and $|y - y_0|$ are at most $1/2$. Let

$$q = x_0 + y_0\sqrt{d}, \quad r = \beta((x - x_0) + (y - y_0)\sqrt{d}\,);$$

then

$$\beta q + r = \beta(x + y\sqrt{d}\,) = \alpha.$$

We must show that $\Psi(r) < \Psi(\beta)$. Now

$$\Psi(a + b\sqrt{d}\,) = |(a + b\sqrt{d}\,)(a - b\sqrt{d}\,)|,$$

and it follows from (Problem 4 that for all $\gamma, \delta \in \mathbb{Z}[\sqrt{d}\,]$ we have

$$\Psi(\gamma\delta) = \Psi(\gamma)\Psi(\delta).$$

(When $d = -1$, this says that the magnitude of the product of two complex numbers is the product of the magnitudes.) Thus $\Psi(r) = \Psi(\beta)[(x - x_0)^2 - d(y - y_0)^2]$, and the factor in brackets is at most $\frac{1}{4} + |d|(\frac{1}{4}) \le \frac{1}{4} + \frac{3}{4} = 1$. The only possibility for equality occurs when $d = 3$ ($d = -3$ is excluded by hypothesis) and $|x - x_0| = |y - y_0| = \frac{1}{2}$. But in this case, the factor in brackets is $|\frac{1}{4} - 3(\frac{1}{4})| = \frac{1}{2} < 1$. We have shown that $\Psi(r) < \Psi(\beta)$, so $\mathbb{Z}[\sqrt{d}\,]$ is a Euclidean domain.

When $d = -1$, we obtain the *Gaussian integers* $a + bi$, $a, b \in \mathbb{Z}$, $i = \sqrt{-1}$.

Problems For Section 2.7

1. Let $A = \{a_1, \dots, a_n\}$ be a finite subset of the PID R. Show that m is a least common multiple of A iff m is a generator of the ideal $\cap_{i=1}^n \langle a_i \rangle$.
2. Find the gcd of $11 + 3i$ and $8 - i$ in the ring of Gaussian integers.
3. Suppose that R is a Euclidean domain in which $\Psi(a) \le \Psi(ab)$ for all nonzero elements $a, b \in R$. Show that $\Psi(a) \ge \Psi(1)$, with equality if and only if a is a unit in R.
4. Let $R = \mathbb{Z}[\sqrt{d}\,]$, where d is any integer, and define $\Psi(a + b\sqrt{d}\,) = |a^2 - db^2|$. Show that for all nonzero α and β, $\Psi(\alpha\beta) = \Psi(\alpha)\Psi(\beta)$, and if d is not a perfect square, then $\Psi(\alpha) \le \Psi(\alpha\beta)$.
5. Let $R = \mathbb{Z}[\sqrt{d}\,]$ where d is not a perfect square. Show that 2 is not prime in R. (Show that 2 divides $d^2 - d$.)
6. If d is a negative integer with $d \le -3$, show that 2 is irreducible in $\mathbb{Z}[\sqrt{d}\,]$.
7. Let $R = \mathbb{Z}[\sqrt{d}\,]$ where d is a negative integer. We know (see (2.7.5)) that R is an ED, hence a PID and a UFD, for $d = -1$ and $d = -2$. Show that for $d \le -3$, R is not a UFD.
8. Find the least common multiple of $11 + 3i$ and $8 - i$ in the ring of Gaussian integers.
9. If $\alpha = a + bi$ is a Gaussian integer, let $\Psi(\alpha) = a^2 + b^2$ as in (2.7.5). If $\Psi(\alpha)$ is prime in $\mathbb{Z}$, show that α is prime in $\mathbb{Z}[i]$.

2.8 Rings of Fractions

It was recognized quite early in mathematical history that the integers have a significant defect: the quotient of two integers need not be an integer. In such a situation a mathematician is likely to say "I want to be able to divide one integer by another, and I will". This will be legal if the computation takes place in a field F containing the integers $\mathbb{Z}$. Any such field will do, since if a and b belong to F and $b \ne 0$, then $a/b \in F$. How do we know that a suitable F exists? With hindsight we can take F to be the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, or the complex numbers $\mathbb{C}$. In fact, $\mathbb{Q}$ is the smallest field containing $\mathbb{Z}$, since any field $F \supseteq \mathbb{Z}$ contains a/b for all $a, b \in \mathbb{Z}$, $b \neq 0$, and consequently $F \supseteq \mathbb{Q}$.

For an arbitrary integral domain D, the same process that leads from the integers to the rationals allows us to construc a field F whose elements are (essentially) fractions a/b, $a, b \in D$, $b \neq 0$. F is called the *field of fractions* or *quotient field* of D. The mathematician's instinct to generalize then leads to the following question: If R is an arbitrary commutative ring (not necessarily an integral domain), can we still form fractions with numerator and denominator in R? Difficulties quickly arise; for example, how do we make sense of $\frac{a}{b}\frac{c}{d}$ when $bd = 0$? Some restriction must be placed on the allowable denominators, and we will describe a successful approach shortly. Our present interest is in the field of fractions of an integral domain, but later we will need the more general development. Since the ideas are very similar, we will give the general construction now.

2.8.1 Definitions and Comments

Let S be a subset of the ring R; we say that S is *multiplicative* if (a) $0 \notin S$, (b) $1 \in S$, and (c) whenever a and b belong to S, we have $ab \in S$. We can merge (b) and (c) by stating that S is closed under multiplication, if we regard 1 as the empty product. Here are some standard examples of multiplicative sets.

(1) The set of all nonzero elements of an integral domain

(2) The set of all elements of a commutative ring R that are not zero divisors,

(3) $R \setminus P$, where P is a prime ideal of the commutative ring R

If S is a multiplicative subset of the commutative ring R, we define the following equivalence relation on $R \times S$:

$$(a, b) \sim (c, d) \text{ if and only if for some } s \in S \text{ we have } s(ad - bc) = 0.$$

If we are constructing the field of fractions of an integral domain, then (a, b) is our first approximation to a/b. Also, since the elements $s \in S$ are never 0 and R has no zero divisors, we have $(a, b) \sim (c, d)$ iff $ad = bc$, and this should certainly be equivalent to $a/b = c/d$.

Let us check that we have a legal equivalence relation. (Commutativity of multiplication will be used many times to slide an element to a more desirable location in a formula. There is also a theory of rings of fractions in the *non*commutative case, but we will not need the results, and in view of the serious technical difficulties that arise, we will not discuss this area.)

Reflexivity and symmetry follow directly from the definition. For transitivity, suppose that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. Then for some elements s and t in S we have

$$s(ad - bc) = 0 \text{ and } t(cf - de) = 0.$$

Multiply the first equation by tf and the second by sb, and add the results to get

$$std(af - be) = 0,$$

which implies that $(a, b) \sim (e, f)$, proving transitivity.

If $a \in R$ and $b \in S$, we define the fraction $\frac{a}{b}$ to be the equivalence class of the pair (a, b). The set of all equivalence classes is denoted by $S^{-1}R$, and in view of what we are about to prove, is called the *ring of fractions of R by S*. The term *localization of R by S* is also used, because it will turn out that in Examples (1) and (3) above, $S^{-1}R$ is a local ring (see Section 2.4, Problem 5).

We now make the set of fractions into a ring in a natural way.

addition $\dfrac{a}{b} + \dfrac{c}{d} = \dfrac{ad + bc}{bd}$

multiplication $\dfrac{a}{b}\dfrac{c}{d} = \dfrac{ac}{bd}$

additive identity $\dfrac{0}{1}$ $(= \dfrac{0}{s}$ for any $s \in S)$

additive inverse $-(\dfrac{a}{b}) = \dfrac{-a}{b}$

multiplicative identity $\dfrac{1}{1}$ $(= \dfrac{s}{s}$ for any $s \in S)$

2.8.2 Theorem

Let S be a multiplicative subset of the commutative ring R. With the above definitions, $S^{-1}R$ is a commutative ring. If R is an integral domain, so is $S^{-1}R$. If R is an integral domain and $S = R \setminus \{0\}$, then $S^{-1}R$ is a field (the *field of fractions* or *quotient field* of R).

Proof. First we show that addition is well-defined. If $a_1/b_1 = c_1/d_1$ and $a_2/b_2 = c_2/d_2$, then for some $s, t \in S$ we have

$$s(a_1d_1 - b_1c_1) = 0 \quad \text{and} \quad t(a_2d_2 - b_2c_2) = 0 \tag{1}$$

Multiply the first equation of (1) by tb_2d_2 and the second equation by sb_1d_1, and add the results to get

$$st[(a_1b_2 + a_2b_1)d_1d_2 - (c_1d_2 + c_2d_1)b_1b_2] = 0.$$

Thus

$$\frac{a_1b_2 + a_2b_1}{b_1b_2} = \frac{c_1d_2 + c_2d_1}{d_1d_2},$$

in other words,

$$\frac{a_1}{b_1} + \frac{a_2}{b_2} = \frac{c_1}{d_1} + \frac{c_2}{d_2}$$

so that addition of fractions does not depend on the particular representative of an equivalence class.

Now we show that multiplication is well-defined. We follow the above computation as far as (1), but now we multiply the first equation by ta_2d_2, the second by sc_1b_1, and add. The result is

$$st[a_1a_2d_1d_2 - b_1b_2c_1c_2] = 0$$

which implies that

$$\frac{a_1}{b_1}\frac{a_2}{b_2} = \frac{c_1}{d_1}\frac{c_2}{d_2},$$

as desired. We now know that the fractions in $S^{-1}R$ can be added and multiplied in exactly the same way as ratios of integers, so checking the defining properties of a commutative ring essentially amounts to checking that the rational numbers form a commutative ring; see Problems 3 and 4 for some examples.

Now assume that R is an integral domain. It follows that if a/b is zero in $S^{-1}R$, i.e., $a/b = 0/1$, then $a = 0$ in R. (For some $s \in S$ we have $sa = 0$, and since R is an integral domain and $s \neq 0$, we must have $a = 0$.) Thus if $\frac{a}{b}\frac{c}{d} = 0$, then $ac = 0$, so either a or c is 0, and consequently either a/b or c/d is zero. Therefore $S^{-1}R$ is an integral domain.

If R is an integral domain and $S = R \setminus \{0\}$, let a/b be a nonzero element of $S^{-1}R$. Then both a and b are nonzero, so $a, b \in S$. By definition of multiplication we have $\frac{a}{b}\frac{b}{a} = \frac{1}{1}$. Thus a/b has a multiplicative inverse, so $S^{-1}R$ is a field. ♣

When we go from the integers to the rational numbers, we don't lose the integers in the process, in other words, the rationals contain a copy of the integers, namely, the rationals of the form $a/1$, $a \in \mathbb{Z}$. So a natural question is whether $S^{-1}R$ contains a copy of R.

2.8.3 Proposition

Define $f\colon R \to S^{-1}R$ by $f(a) = a/1$. Then f is a ring homomorphism. If S has no zero divisors then f is a monomorphism, and we say that R can be *embedded* in $S^{-1}R$. In particular:

(i) A commutative ring R can be embedded in its *complete* (or *full*) ring of fractions ($S^{-1}R$, where S consists of all non-divisors of zero in R).

(ii) An integral domain can be embedded in its quotient field.

Proof. We have $f(a+b) = \frac{a+b}{1} = \frac{a}{1} + \frac{b}{1} = f(a) + f(b)$, $f(ab) = \frac{ab}{1} = \frac{a}{1}\frac{b}{1} = f(a)f(b)$, and $f(1) = \frac{1}{1}$, proving that f is a ring homomorphism. If S has no zero divisors and $f(a) = a/1 = 0/1$, then for some $s \in S$ we have $sa = 0$, and since s cannot be a zero divisor, we have $a = 0$. Thus f is a monomorphism. ♣

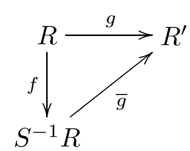
2.8.4 Corollary

The quotient field F of an integral domain R is the smallest field containing R.

Proof. By (2.8.3), we may regard R as a subset of F, so that F is a field containing R. But if L is any field containing R, then all fractions a/b, $a, b \in R$, must belong to L. Thus $F \subseteq L$. ♣

Problems For Section 2.8

1. If the integral domain D is in fact a field, what is the quotient field of D?
2. If D is the set $F[X]$ of all polynomials over a field F, what is the quotient field of D?
3. Give a detailed proof that addition in a ring of fractions is associative.
4. Give a detailed proof that the distributive laws hold in a ring of fractions.
5. Let R be an integral domain with quotient field F, and let h be a ring monomorphism from R to a field L. Show that h has a unique extension to a monomorphism from F to L.
6. Let h be the ring homomorphism from $\mathbb{Z}$ to $\mathbb{Z}_p$, p prime, given by $h(x) = x \bmod p$. Why can't the analysis of Problem 5 be used to show that h extends to a monomorphism of the rationals to $\mathbb{Z}_p$? (This can't possibly work since $\mathbb{Z}_p$ is finite, but what goes wrong?)
7. Let S be a multiplicative subset of the commutative ring R, with $f\colon R \to S^{-1}R$ defined by $f(a) = a/1$. If g is a ring homomorphism from R to a commutative ring R' and $g(s)$ is a unit in R' for each $s \in S$, we wish to find a ring homomorphism $\overline{g}\colon S^{-1}R \to R'$ such that $\overline{g}(f(a)) = g(a)$ for every $a \in R$, i.e., such that the diagram below is commutative. Show that there is only one conceivable way to define $\overline{g}$.



8. Show that the mapping you have defined in Problem 7 is a well-defined ring homomorphism.

2.9 Irreducible Polynomials

2.9.1 Definitions and Comments

In (2.6.1) we defined an irreducible element of a ring; it is a nonzero nonunit which cannot be represented as a product of nonunits. If R is an integral domain, we will refer to an irreducible element of $R[X]$ as an *irreducible polynomial*. Now in $F[X]$, where F is a field, the units are simply the nonzero elements of F (Section 2.1, Problem 2). Thus in this case, an irreducible element is a polynomial of degree at least 1 that cannot be factored into two polynomials of lower degree. A polynomial that is not irreducible is said to be *reducible* or *factorable*. For example, $X^2 + 1$, regarded as an element of $\mathbb{R}[X]$, where $\mathbb{R}$ is the field of real numbers, is irreducible, but if we replace $\mathbb{R}$ by the larger field $\mathbb{C}$ of complex numbers, $X^2 + 1$ is factorable as $(X - i)(X + i)$, $i = \sqrt{-1}$. We say that $X^2 + 1$ is *irreducible over* $\mathbb{R}$ but not *irreducible over* $\mathbb{C}$.

Now consider $D[X]$, where D is a unique factorization domain but not necessarily a field, for example, $D = \mathbb{Z}$. The polynomial $12X + 18$ is not an irreducible element of $\mathbb{Z}[X]$ because it can be factored as the product of the two nonunits 6 and $2X + 3$.

It is convenient to factor out the greatest common divisor of the coefficients (6 in this case). The result is a *primitive polynomial*, one whose *content* (gcd of coefficients) is 1. A primitive polynomial will be irreducible if and only if it cannot be factored into two polynomials of lower degree.

In this section, we will compare irreducibility over a unique factorization domain D and irreducibility over the quotient field F of D. Here is the key result.

2.9.2 Proposition

Let D be a unique factorization domain with quotient field F. Suppose that f is a nonzero polynomial in $D[X]$ and that f can be factored as gh, where g and h belong to $F[X]$. Then there is a nonzero element $\lambda \in F$ such that $\lambda g \in D[X]$ and $\lambda^{-1}h \in D[X]$. Thus if f is factorable over F, then it is factorable over D. Equivalently, if f is irreducible over D, then f is irreducible over F.

Proof. The coefficients of g and h are quotients of elements of D. If a is the least common denominator for g (technically, the least common multiple of the denominators of the coefficients of g), let $g^* = ag \in D[X]$. Similarly, let $h^* = bh \in D[X]$. Thus $abf = g^*h^*$ with $g^*, h^* \in D[X]$ and $c = ab \in D$.

Now if p is a prime factor of c, we will show that either p divides all coefficients of g^* or p divides all coefficients of h^*. We do this for all prime factors of c to get $f = g_0h_0$ with $g_0, h_0 \in D[X]$. Since going from g to g_0 involves only multiplication or division by nonzero constants in D, we have $g_0 = \lambda g$ for some nonzero $\lambda \in F$. But then $h_0 = \lambda^{-1}h$, as desired.

Now let

$$g^*(X) = g_0 + g_1X + \cdots + g_sX^s, \ h^*(X) = h_0 + h_1X + \cdots + h_tX^t.$$

Since p is a prime factor of $c = ab$ and $abf = g^*h^*, p$ must divide all coefficients of g^*h^*. If p does not divide every g_i and p does not divide every h_i, let g_u and h_v be the coefficients of minimum index not divisible by p. Then the coefficient of X^{u+v} in g^*h^* is

$$g_0h_{u+v} + g_1h_{u+v-1} + \cdots + g_uh_v + \cdots + g_{u+v-1}h_1 + g_{u+v}h_0.$$

But by choice of u and v, p divides every term of this expression except g_uh_v, so that p cannot divide the entire expression. So there is a coefficient of g^*h^* not divisible by p, a contradiction. ♣

The technique of the above proof yields the following result.

2.9.3 Gauss' Lemma

Let f and g be nonconstant polynomials in $D[X]$, where D is a unique factorization domain. If c denotes content, then $c(fg) = c(f)c(g)$, up to associates. In particular, the product of two primitive polynomials is primitive.

Proof. By definition of content we may write $f = c(f)f^*$ and $g = c(g)g^*$ where f^* and g^* are primitive. Thus $fg = c(f)c(g)f^*g^*$. It follows that $c(f)c(g)$ divides every coefficient of fg, so $c(f)c(g)$ divides $c(fg)$. Now let p be any prime factor of $c(fg)$; then p divides $c(f)c(g)f^*g^*$, and the proof of (2.9.2) shows that either p divides every coefficient of $c(f)f^*$ or p divides every coefficient of $c(g)g^*$. If, say, p divides every coefficient of $c(f)f^*$, then (since p is prime) either p divides $c(f)$ or p divides every coefficient of f^*. But f^* is primitive, so that p divides $c(f)$, hence p divides $c(f)c(g)$. We conclude that $c(fg)$ divides $c(f)c(g)$, and the result follows. ♣

2.9.4 Corollary of the Proof of (2.9.3)

If h is a nonconstant polynomial in $D[X]$ and $h = ah^*$ where h^* is primitive and $a \in D$, then a must be the content of h.

Proof. Since a divides every coefficient of h, a must divide $c(h)$. If p is any prime factor of $c(h)$, then p divides every coefficient of ah^*, and as in (2.9.3), either p divides a or p divides every coefficient of h^*, which is impossible by primitivity of h^*. Thus $c(h)$ divides a, and the result follows. ♣

Proposition 2.9.2 yields a precise statement comparing irreducibility over D with irreducibility over F.

2.9.5 Proposition

Let D be a unique factorization domain with quotient field F. If f is a nonconstant polynomial in $D[X]$, then f is irreducible over D if and only if f is primitive and irreducible over F.

Proof. If f is irreducible over D, then f is irreducible over F by (2.9.2). If f is not primitive, then $f = c(f)f^*$ where f^* is primitive and $c(f)$ is not a unit. This contradicts the irreducibility of f over D. Conversely, if $f = gh$ is a factorization of the primitive polynomial f over D, then g and h must be of degree at least 1. Thus neither g nor h is a unit in $F[X]$, so $f = gh$ is a factorization of f over F. ♣

Here is another basic application of (2.9.2).

2.9.6 Theorem

If R is a unique factorization domain, so is $R[X]$.

Proof. If $f \in R[X]$, $f \neq 0$, then f can be factored over the quotient field F as $f = f_1 f_2 \dots f_k$, where the f_i are irreducible polynomials in $F[X]$. (Recall that $F[X]$ is a Euclidean domain, hence a unique factorization domain.) By (2.9.2), for some nonzero $\lambda_1 \in F$ we may write $f = (\lambda_1 f_1)(\lambda_1^{-1} f_2 \dots f_k)$ with $\lambda_1 f_1$ and $\lambda_1^{-1} f_2 \dots f_k$ in $R[X]$. Again by (2.9.2), we have

$$\lambda_1^{-1} f_2 \dots f_k = f_2 \lambda_1^{-1} f_3 \dots f_k = (\lambda_2 f_2)(\lambda_2^{-1} \lambda_1^{-1} f_3 \dots f_k)$$

with $\lambda_2 f_2$ and $\lambda_2^{-1}\lambda_1^{-1} f_3 \dots f_k \in R[X]$. Continuing inductively, we express f as $\prod_{i=1}^k \lambda_i f_i$ where the $\lambda_i f_i$ are in $R[X]$ and are irreducible over F. But $\lambda_i f_i$ is the product of its content and a primitive polynomial (which is irreducible over F, hence over R by (2.9.5)). Furthermore, the content is either a unit or a product of irreducible elements of the UFD R, and these elements are irreducible in $R[X]$ as well. This establishes the existence of a factorization into irreducibles.

Now suppose that $f = g_1 \cdots g_r = h_1 \cdots h_s$, where the g_i and h_i are nonconstant irreducible polynomials in $R[X]$. (Constant polynomials cause no difficulty because R is a UFD.) By (2.9.5), the g_i and h_i are irreducible over F, and since $F[X]$ is a UFD, we have $r = s$ and, after reordering if necessary, g_i and h_i are associates (in $F[X]$) for each i. Now $g_i = c_i h_i$ for some constant $c_i \in F$, and we have $c_i = a_i/b_i$ with $a_i, b_i \in R$. Thus $b_i g_i = a_i h_i$, with g_i and h_i primitive by (2.9.5). By (2.9.4), $b_i g_i$ has content b_i and $a_i h_i$ has content a_i. Therefore a_i and b_i are associates, which makes c_i a unit in R, which in turn makes g_i and h_i associates in $R[X]$, proving uniqueness of factorization. ♣

The following result is often used to establish irreducibility of a polynomial.

2.9.7 Eisenstein's Irreducibility Criterion

Let R be a UFD with quotient field F, and let $f(X) = a_n X^n + \cdots + a_1 X + a_0$ be a polynomial in $R[X]$, with $n \geq 1$ and $a_n \neq 0$. If p is prime in R and p divides a_i for $0 \leq i < n$, but p does not divide a_n and p^2 does not divide a_0, then f is irreducible over F. Thus by (2.9.5), if f is primitive then f is irreducible over R.

Proof. If we divide f by its content to produce a primitive polynomial f^*, the hypothesis still holds for f^*. (Since p does not divide a_n, it is not a prime factor of $c(f)$, so it must divide the i^{th} coefficient of f^* for $0 \leq i < n$.) If we can prove that f^* is irreducible over R, then by (2.9.5), f^* is irreducible over F, and therefore so is f. Thus we may assume without loss of generality that f is primitive, and prove that f is irreducible over R.

Assume that $f = gh$, with $g(X) = g_0 + \cdots + g_r X^r$ and $h(X) = h_0 + \cdots + h_s X^s$. If $r = 0$ then g_0 divides all coefficients a_i of f, so g_0 divides $c(f)$, hence $g(= g_0)$ is a unit. Thus we may assume that $r \geq 1$, and similarly $s \geq 1$. By hypothesis, p divides $a_0 = g_0 h_0$ but p^2 does not divide a_0, so p cannot divide both g_0 and h_0. Assume that p fails to divide h_0, so that p divides g_0; the argument is symmetrical in the other case. Now $g_r h_s = a_n$, and by hypothesis, p does not divide a_n, so that p does not divide g_r. Let i be the smallest integer such that p does not divide g_i; then $1 \leq i \leq r < n$ (since $r + s = n$ and $s \geq 1$). Now

$$a_i = g_0 h_i + g_1 h_{i-1} + \cdots + g_i h_0$$

and by choice of i, p divides $g_0, \dots, g_{i-1}$. But p divides the entire sum a_i, so p must divide the last term $g_i h_0$. Consequently, either p divides g_i, which contradicts the choice of i, or p divides h_0, which contradicts our earlier assumption. Thus there can be no factorization of f as a product of polynomials of lower degree; in other words, f is irreducible over R. ♣

Problems For Section 2.9

1. (The *rational root test*, which can be useful in factoring a polynomial over $\mathbb{Q}$.)
   Let $f(X) = a_nX^n + \cdots + a_1X + a_0 \in \mathbb{Z}[X]$. If f has a rational root u/v where u and v are relatively prime integers and $v \neq 0$, show that v divides a_n and u divides a_0.
2. Show that for every positive integer n, there is at least one irreducible polynomial of degree n over the integers.
3. If $f(X) \in \mathbb{Z}[X]$ and p is prime, we can reduce all coefficients of f modulo p to obtain a new polynomial $f_p(X) \in \mathbb{Z}_p[X]$. If f is factorable over $\mathbb{Z}$, then f_p is factorable over $\mathbb{Z}_p$. Therefore if f_p is irreducible over $\mathbb{Z}_p$, then f is irreducible over $\mathbb{Z}$. Use this idea to show that the polynomial $X^3 + 27X^2 + 5X + 97$ is irreducible over $\mathbb{Z}$. (Note that Eisenstein does not apply.)
4. If we make a change of variable $X = Y + c$ in the polynomial $f(X)$, the result is a new polynomial $g(Y) = f(Y + c)$. If g is factorable over $\mathbb{Z}$, so is f since $f(X) = g(X - c)$. Thus if f is irreducible over $\mathbb{Z}$, so is g. Use this idea to show that $X^4+4X^3+6X^2+4X+4$ is irreducible over $\mathbb{Z}$.
5. Show that in $\mathbb{Z}[X]$, the ideal $\langle n, X\rangle, n \geq 2$, is not principal, and therefore $\mathbb{Z}[X]$ is a UFD that is not a PID.
6. Show that if F is a field, then $F[X, Y]$, the set of all polynomials $\sum a_{ij}X^iY^j$, $a_{ij} \in F$, is not a PID since the ideal $\langle X, Y\rangle$ is not principal.
7. Let $f(X, Y) = X^2 + Y^2 + 1 \in \mathbb{C}[X, Y]$, where $\mathbb{C}$ is the field of complex numbers. Write f as $Y^2 + (X^2 + 1)$ and use Eisenstein's criterion to show that f is irreducible over $\mathbb{C}$.
8. Show that $f(X, Y) = X^3 + Y^3 + 1$ is irreducible over $\mathbb{C}$.