II YEAR - M.Sc. MATHEMATICS

M.S. University

Directorate of Distance & Continuing Education

Tirunelveli

RING THEORY AND LATTICES

Prepared by

Dr. T. Anitha Baby

மனோன்மணியம் சுந்தரனார் பல்கலைக்கழகம்

MANONMANIAM SUNDARANAR UNIVERSITY

TIRUNELVELI-627 012

தொலைநிலை தொடர் கல்வி இயக்ககம்

DIRECTORATE OF DISTANCE AND CONTINUING EDUCATION

M.Sc. MATHEMATICS

II YEAR

RING THEORY AND LATTICES

Sub. Code: SMAE41

Prepared by

Dr. T. ANITHA BABY

Assistant Professor

Department of Mathematics

Women's Christian College, Nagercoil - 629001

RING THEORY AND LATTICES

UNIT	DETAILS
I	Ring homomorphism- ideals and Quotient ring -More ideals and Quotient rings – the field of quotients of integral domain
II	Euclidean ring- A particular ring
III	Polynomial rings- polynomials over rational field- Polynomial ring over commutative rings
IV	Certain radical of a ring-Jacobson radical of a ring- Semi simple ring-Nil radical
V	Partially ordered sets and lattices – distributivity and modularity- the theorem of Jordan Holder – Boolean algebra

Text Books:

1. I. N. Herstein, Topics in Algebra, 2rd Edition, Wiley Student Edition, New Delhi, 2006
2. David M. Burton, A first course in Rings and Ideals, Addison Wesley Publishing Company, the University of Michigan, 1970.
3. I. Nathan Jacobson, Basic Algebra, Yale University, W. H. Freeman and Company, New York,1985

Contents

Chapter 1

Unit 1

1.1 Basics of Ring Theory

In this section, we collect basic definitions and results on rings.

Definition 1.1.1. A *ring* $(R, +, \cdot)$ is a nonempty set R together with binary operations '+'and '$\cdot$'defined on R, which satisfy the following conditions:

(i) $(R, +)$ is an abelian group

(ii) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$, for all $a, b, c \in R$

(iii) $a \cdot (b + c) = a \cdot b + a \cdot c$, for all $a, b, c \in R$

(iv) $(a + b) \cdot c = a \cdot c + b \cdot c$, for all $a, b, c \in R$.

Definition 1.1.2. A ring R is called *commutative* if for every $a, b \in R$, $a{\cdot}b = b{\cdot}a$. A ring R which is not commutative is called a *noncommutative* ring.

Example 1.1.3. Let $M_n(\mathbb{Z})$ denote the set of all $n \times n$ matrices over the ring of integers $\mathbb{Z}$. Let '+'and '$\cdot$'denote the usual matrix addition and multiplication, respectively. Then $(M_n(\mathbb{Z}), +, \cdot)$ is a noncommutative ring.

Definition 1.1.4. Let R be a ring. An element $e \in R$ is called an *identity* element if $ea = ae = a$ for all $a \in R$. The identity element of a ring R is denoted by '1'.

Definition 1.1.5. Let R be a ring with identity. An element $u \in R$ is called a *unit* element if there exists $v \in R$ such that $uv = 1 = vu$. The set of all units in R is denoted by $R^{\times}$.

Definition 1.1.6. A ring R with identity is called a *division ring* if every nonzero element of R is a unit. A commutative division ring R is called a *field.*

Definition 1.1.7. A nonzero element x of a ring R is a *left zero-divisor*(*right zero-divisor*) if there exists a nonzero element $y \in R$ such that $xy =$

0 ($yx = 0$). An element that is both a left and a right zero-divisor is simply called a *zero-divisor.* The set of all nonzero zero-divisors of R is denoted by $Z(R)^*$.

Definition 1.1.8. A commutative ring R is called an *integral domain* if R has no zero-divisors.

Definition 1.1.9. An ideal I of a ring R is called a *proper ideal* if $I \neq \{0\}$ and $I \neq R$.

Definition 1.1.10. An ideal P of a ring R is called a *prime ideal* if $P \neq R$ and for all $a, b \in R$, $ab \in P$ implies $a \in P$ or $b \in P$.

Definition 1.1.11. Let R be a commutative ring with identity. An ideal I of R is called a *principal ideal* if $I = \langle a \rangle = \{ra : r \in R\}$ for some $a \in R$.

Definition 1.1.12. An element x of a ring R is called *nilpotent* if there exists some positive integer n such that $x^n = 0$.

Definition 1.1.13. Let R be a ring. The *characteristic* of R is the least positive integer n such that $na = 0$ for all $a \in R$. If no such positive integer exists, then R is said to be of *characteristic zero.*

Remark 1.1.14. The set $\mathfrak{N}(R)$ of all nilpotent elements in R is an ideal. The ideal $\mathfrak{N}(R)$ is called the *nilradical* of R.

Definition 1.1.15. An ideal I of R is called a *nil-ideal* if each element of I is a nilpotent. An ideal I of R is called *nilpotent* if $I^k = (0)$ for some positive integer $k \geq 1$.

Definition 1.1.16. Let R be a ring with identity. The *Jacobson radical* of R, denoted by $\mathfrak{J}(R)$, is the intersection of all maximal ideals of R.

Definition 1.1.17. A ring R is called *reduced* if it contains no nonzero nilpotent elements.

Definition 1.1.18. An ideal I of a ring R is called an *annihilating-ideal* if there exists a nonzero ideal I' of R such that $II' = (0)$.

Definition 1.1.19. A ring R is said to be a *local ring* if it contains unique maximal ideal.

Definition 1.1.20. A ring R is said to satisfy the *descending chain condition* of ideals if, for every chain of ideals $I_1 \supseteq I_2 \supseteq I_3 \supseteq \dots$, there exists a positive integer m such that $I_k = I_m$ for all $k \geq m$. A ring R is said to be an *Artinian* if it satisfies the descending chain condition of ideals.

Definition 1.1.21. A ring R is said to satisfy the *ascending chain condition* of ideals if, for every chain of ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \ldots$, there exists a positive integer m such that $I_k = I_m$ for all $k \geq m$. A ring R is said to be *Noetherian* if it satisfies the ascending chain condition of ideals.

Remark 1.1.22. If R is an Artinian ring, then $\mathfrak{N}(R) = \mathfrak{J}(R)$.

1.2 Homomorphisms

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

Definition 1.2.1. A mapping ϕ from the ring R into the ring R' is said to be a **homomorphism** if

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

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

for all $a, b \in R$.

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

A useful observation to make is that a homomorphism of one ring, R, into another, R', is, if we totally ignore the multiplications in both these rings, at least a homomorphism of R into R' when we consider them as abelian groups under their respective additions. Therefore, as far as addition is concerned, all the properties about homomorphisms of groups proved carry over. In particular, merely restating Lemma 1.2.2 for the case of the additive group of a ring yields for us

Lemma 1.2.2. *lf ϕ is a homomorphism of R into R', then*

1. $\phi(0) = 0$.

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

If both R and R' have the respective unit elements 1 and $1'$ for their multiplications it need not follow that $\phi(1) = 1'$. However, if R' is an integral domain, or if R' is arbitrary but ϕ is onto, then $\phi(1) = 1'$ is indeed true. In the case of groups, given a homomorphism we associated with this homomorphism a certain subset of the group which we called the kernel of the homomorphism.

Definition 1.2.3. If ϕ is a homomorphism of R into R' then the kernel of ϕ, $I(\phi)$, is the set of all elements $a \in R$ such that $\phi(a) = 0$, the zero-element of R'.

Lemma 1.2.4. *If ϕ is a homomorphism of R into R' with kernel $I(\phi)$, then*

1. I(ϕ) is a subgroup of R under addition.

2. If $a \in I(\phi)$ and $r \in R$ then both ar and ra are in I(ϕ).

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

To see (2), suppose that $a \in I(\phi)$, $r \in R$. Then $\phi(a) = 0$ so that $\phi(ar) = \phi(a)\phi(r) = 0\phi(r) = 0$ by Lemma 1.2.2. Similarly $\phi(ra) = 0$. Thus by defining property of $I(\phi)$ both ar and ra are in $I(\phi)$.

Before proceeding we examine these concepts for certain examples.

□

Example 1.2.5. Let R and R' be two arbitrary rings and define $\phi(a) = 0$ for all $a \in R$. Trivially ϕ is a homomorphism and $I(\phi) = R$ and hence ϕ is called the zero-homomorphism.

Example 1.2.6. Let R be a ring, $R' = R$ and define $\phi(x) = x$ for every $x \in R$. Clearly ϕ is a homomorphism and $I(\phi)$ consists only of 0.

Example 1.2.7. Let $J(\sqrt{2})$ be all real numbers of the form $m + n\sqrt{2}$ where m, n are integers; $J(\sqrt{2})$ forms a ring under the usual addition and multiplication of real numbers.

Define $\phi : J(\sqrt{2}) \longrightarrow J(\sqrt{2})$ by $\phi(m + n\sqrt{2}) = m - n\sqrt{2}$. Then ϕ is a homomorphism of $J(\sqrt{2})$ onto $J(\sqrt{2})$ and its kernel $I(\phi) = \{0\}$.

Example 1.2.8. Let J be the ring of integers, J_n, the ring of integers modulo n. Define $\phi : J \longrightarrow J_n$ by $\phi(a) =$ remainder of a on division by n. Then ϕ is a homomorphism of J onto J_n and that the kernel, $I(\phi)$, of ϕ consists of all multiples of n.

Example 1.2.9. Let $R = C[0, 1]$ be the set of all continuous, real-valued functions on $[0, 1]$. Then $(R, +, \cdot)$ is a commutative ring with identity. Define $\phi : R \longrightarrow \mathbb{R}$ by $\phi(f(x)) = f(\frac{1}{2})$. Then ϕ is a homomorphism of R onto $\mathbb{R}$ and its kernel $I(\phi) = \{f \in R : f(\frac{1}{2}) = 0\}$.

Definition 1.2.10. A homomorphism of R into R' is said to be an isomorphism if it is a one-to-one mapping.

Definition 1.2.11. Two rings are said to be isomorphic if there is an isomorphism of one onto the other.

Lemma 1.2.12. *The homomorphism ϕ of R into R' is one to one if and only if $I(\phi) = (0)$.*

Proof. Suppose ϕ is one to one. Let $x \in I(\phi)$. Then $\phi(x) = 0' = \phi(0)$ and by hypothesis, $x = 0$ and hence $I(\phi) = (0)$.

Conversely, assume that $I(\phi) = (0)$. Let $x, y \in R$. Suppose $\phi(x) = \phi(y)$. Then $\phi(x - y) = 0'$ and so $x - y \in I(\phi)$. By hypothesis, $x = y$ and ϕ is one to one. □

1.3 Ideals and Quotient Rings

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

Definition 1.3.1. A nonempty subset J of R is said to be a (two-sided) ideal of R if

1. J is a subgroup of R under addition.

2. For every $u \in J$ and $r \in R$, both ur and ru are in J.

Definition 1.3.2. An ideal M of a ring R is called a *maximal ideal* if $M \neq R$ and the only ideals containing M are M and R.

Lemma 1.3.3. *If U is an ideal of the ring R, then R/U is a ring and is a homomorphic image of R.*

Proof. Given an ideal U of a ring R, let $R/U = \{a+U : a \in R\}$ be the set of all the distinct cosets of U in R and $a+U+b+U = (a+b)+U$. Since $(U,+)$ is a subgroup of $(R,+)$, $(R/U,+)$ is a group. Since R is an abelian group under addition, $(R/U,+)$ is an abelian group. From this $0+U$ is an additive identity in R/U and $-a+U$ is an additive inverse of $a+U$ in R/U

Define $(a+U)(b+U) = ab+U$. If $a+U = a'+U$ and $b+U = b'+U$, then under our definition of the multiplication, $(a+U)(b+U) = (a'+U)(b'+U)$. Equivalently, it must be established that $ab+U = a'b'+U$. To this end we first note that since $a+U = a'+U$, $a = a'+u_1$ where $u_1 \in U$; similarly $b = b'+u_2$ where $u_2 \in U$. Thus $ab = (a'+u_1)(b+u_2) = a'b'+u_1b'+a'u_2+u_1u_2$; since U is an ideal of R, $u_1b' \in U$, $a'u_2 \in U$, and $u_1u_2 \in U$.

Consequently $u_1b'+a'u_2+u_1u_2 = u_3 \in U$. But then $ab = a'b'+u_3$, from which we deduce that $ab+U = a'b'+u_3+U$, and since $u_3 \in U$, $u_3+U = U$. Hence $ab+U = a'b'+U$.

If $X = a+U$, $Y = b+U$, $Z = c+U$ are three elements of R/U, where $a,b,c \in R$, then $(X+Y)Z = ((a+U)+(b+U))(c+U) = ((a+b)+U)(c+U) = (a+b)c+U = ac+bc+U = (ac+U)+(bc+U) =$

$(a+U)(c+U)+(b+U)(c+U) = XZ+YZ$. Similar way we get $Z(X+Y) = ZX+ZY$ and hence $(R/U,+,\cdot)$ is a ring.

Clearly, if R is commutative then so is R/U, for $(a+U)(b+U) = ab+U = ba+U = (b+U)(a+U)$. If R has a unit element 1, then R/U has a unit element $1+U$.

Define $\phi : R \rightarrow R/U$ by $\phi(a) = a+U$ for all $a \in R$. Then $\phi(a+b) = a+b+U = a+U+b+U = \phi(a)+\phi(b)$ and $\phi(ab)ab+U = (a+U)(b+U) = \phi(a)\phi(b)$ and so ϕ is ring homomorphism. For any $x = a+U \in R/U$, $x = a+U = \phi(a)$ and so ϕ is onto. Let $a \in I(\phi)$. Then $\phi(a) = 0+U$ implies $a+U = 0+U$ and so $a \in U$. Hence $I(\phi) = U$. □

Theorem 1.3.4. *Let R, R' be rings and ϕ a homomorphism of R onto R' with kernel U. Then R' is isomorphic to R/U. Moreover there is a one-to-one correspondence between the set of ideals of R' and the set of ideals of R which contain U. This correspondence can be achieved by associating with an ideal W' in R' the ideal W in R defined by $W = \{x \in R \mid \phi(x) \in W'\}$. With W so defined, R/W is isomorphic to R'/W'.*

1.4 More Ideals and Quotient Rings

We continue the discussion of ideals and quotient rings. We now ask the explicit question: Under what conditions is the homomorphic image of a ring a field? For commutative rings we give a complete answer in this section.

Lemma 1.4.1. *Let R be a commutative ring with unit element whose only ideals are (0) and R itself. Then R is a field.*

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

So, suppose that $a = 1 = 0$ is in R. Consider the set $Ra = \{xa \mid x \in R\}$. We claim that Ra is an ideal of R. In order to establish this as fact we must show that it is a subgroup of R under addition and that if $u \in Ra$ and $r \in R$ then ru is also in Ra. (We only need to check that ru is in Ra for then ur also is since $ru = ur$.)

Now, if $u, v \in Ra$, then $u = r_1a$, $v = r_2a$ for some $r_1, r_2 \in R$. Thus $u + v = r_1a + r_2a = (r_1 + r_2)a \in Ra$; similarly $-u = -r_1a = (-r_1)a \in Ra$. Hence Ra is an additive subgroup of R. Moreover, if $r \in R$, $ru = r(r_1a) = (rr_1)a \in Ra$. Ra therefore satisfies all the defining conditions for an ideal of R, hence is an ideal of R.

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

can be realized as a multiple of a; that is, there exists an element $b \in R$ such that $ba = 1$. □

Definition 1.4.2. An ideal $M \neq R$ in a ring R is said to be a **maximal ideal** of R if whenever U is an ideal of R such that $M \subset U \subset R$, then either $R = U$ or $M = U$.

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

Example 1.4.3. Let R be the ring of integers, and let U be an ideal of R. Since U is a subgroup of R under addition, from our results in group theory, we know that U consists of all the multiples of a fixed integer n_0 ; we write this as $U = (n_0)$. What values of n_0 lead to maximal ideals?

We first assert that if p is a prime number then $P = (p)$ is a maximal ideal of R. For if U is an ideal of R and $U \subset P$, then $U = (n_0)$ for some integer n_0 . Since $p \in P \subset U$, $p = mn_0$ for some integer m; because p is a prime this implies that $n_0 = 1$ or $n_0 = p$. If $n_0 = p$, then $P \subset U = (n_0) \subset P$, so that $U = P$ follows; if $n_0 = 1$, then $1 \in U$, hence

$r = lr \in U$ for all $r \in R$ whence $U = R$ follows. Thus no ideal, other than R or P itself, can be put between P and R, from which we deduce that P is maximal.

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

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

Example 1.4.4. Let R be the ring of all the real-valued, continuous functions on the closed unit interval. Let

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

M is certainly an ideal of R. Moreover, it is a maximal ideal of R, for if the ideal U contains M and $U \neq M$, then there is a function $g(x) \in U$, $g(x) \notin M$. Since $g(x) \notin M$, $g(\frac{1}{2}) = \alpha \neq 0$. Now $h(x) = g(x) - \alpha$ is such that $h(\frac{1}{2}) = g(\frac{1}{2}) - \alpha = 0$, so that $h(x) \in M \subset U$. But $g(x)$ is also in U; therefore $\alpha = g(x) - h(x) \in U$ and so $1 = \alpha\alpha^{-1} \in U$. Thus for any function $t(x) \in R$, $t(x) = 1t(x) \in U$, in consequence of which $U = R$. M is therefore a maximal ideal of R. Similarly if γ is a real number $0 \leqq \gamma \leqq 1$, then $M_\gamma = \{f(x) \in R \mid f(\gamma) = 0\}$ is a maximal ideal of R.

Example 1.4.5. If R is a field, then (0) is a maximal ideal in R

Example 1.4.6. If $R = \mathbb{Z}_n$, where $n = p_1^{k_1} \cdots p_t^{k_t}$, where p_i's are distinct primes, then (p_i) is only maximal ideal in R

Theorem 1.4.7. *If R is a commutative ring with unit element and M is an ideal of R, then M is a maximal ideal of R if and only if R/M is a field.*

Proof. Suppose, first, that M is an ideal of R such that R/M is a field. Since R/M is a field its only ideals are (0) and R/M itself. But by Theorem 1.3.4 there is a one-to-one correspondence between the set of ideals of R/M and the set of ideals of R which contain M. The ideal M of R corresponds to the ideal (0) of R/M whereas the ideal $R$of R corresponds to the ideal R/M of R/M in this one-to-one mapping. Thus there is no ideal between M and R other than these two, whence M is a maximal ideal.

On the other hand, if M is a maximal ideal of R, by the correspondence mentioned above R/M has only (0) and itself as ideals. Furthermore R/M is commutative and has a unit element since R enjoys both these properties. All the conditions of Lemma 1.4.1 are fulfilled for R/Mso we can conclude that, R/M is a field. □

1.5 The Field of Quotients of an Integral Domain

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

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

Definition 1.5.1. A ring R can be **imbedded** in a ring R' if there is an isomorphism of R into R'. (If R and R' have unit elements 1 and $1'$ we insist, in addition, that this isomorphism takes 1 onto $1'$.)

R' will be called an **over-ring** or **extension** of R if R can be imbedded in R'. With this understanding of imbedding we prove

Theorem 1.5.2. *Every integral domain can be imbedded in a field.*

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

1. When is $a/b = c/d$?

2. What is $(a/b) + (c/d)$?

3. What is $(a/b)(c/d)$?

In answer to 1, what could be more natural than to insist that $a/b = c/d$ if and only if $ad = bc$? As for 2 and 3, why not try the obvious, that is, define

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

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

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

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

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

1. If $(a, b) \in \mathscr{M}$, then $(a, b) \sim (a, b)$ since $ab = ba$.

2. If $(a, b), (c, d) \in \mathscr{M}$ and $(a, b) \sim (c, d)$, then $ad = bc$, hence $cb = da$, and so $(c, d) \sim (a, b)$.

3. If $(a, b), (c, d), (e, f)$ are all in $\mathscr{M}$ and $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$, then $ad = bc$ and $cf = de$. Thus $bcf = bde$, and since $bc = ad$, it follows that $adf = bde$. Since D is commutative, this relation becomes $afd = bed$; since, moreover, D is an integral domain and $d \neq 0$, this relation further implies that $af = be$. But then $(a, b) \sim (e, f)$ and our relation is transitive. Let $[a, b]$ be the equivalence class in $\mathscr{M}$ of (a, b), and let F be the set of all such equivalence classes $[a, b]$ where $a, b \in D$ and $b \neq 0$. F is the candidate for the field we are seeking. In order to create out of F a field we must introduce an addition and a multiplication for its elements and then show that under these operations F forms a field. We first dispose of the addition. Motivated by our heuristic discussion at the beginning of the proof we define $[a, b] + [c, d] = [ad + bc, bd]$

Since D is an integral domain and both $b \neq 0$ and $d \neq 0$ we have that $bd \neq 0$; this, at least, tells us that $[ad + bc, bd] \in F$. We now assert that this addition is well defined, that is, if $[a, b] = [a', b']$ and $[c, d] = [c', d']$, then $[a, b] + [c, d] = [a', b'] + [c', d']$. To see that this is so, from $[a, b] = [a', b']$ we have that $ab' = ba'$; from $[c, d] = [c', d']$ we have that $cd' = dc'$.

What we need is that these relations force the equality of $[a, b] + [c, d]$ and $[a', b'] + [c', d']$. From the definition of addition this boils down to showing that $[ad + bc, bd] = [a'd' + b'c', b'd']$, or, in equivalent terms, that $(ad + bc)b'd' = bd(a'd' + b'c')$. Using $ab' = ba'$, $cd' = dc'$ this becomes: $(ad+bc)b'd' = adb'd'+bcb'd' = ab'dd'+bb'cd' = ba'dd'+bb'dc' = bd(a'd' + b'c')$, which is the desired equality.

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

We now turn to the multiplication in F. Again motivated by our preliminary heuristic discussion we define $[a, b][c, d] = [ac, bd]$. As in the case of addition, since $b \neq 0$, $d \neq 0$, $bd \neq 0$ and so $[ac, bd] \in F$. A computation, very much in the spirit of the one just carried out, proves that if $[a, b] = [a', b']$ and $[c, d] = [c', d']$ then $[a, b][c, d] = [a', b'][c', d']$. One can now show that the nonzero elements of F (that is, all the elements $[a, b]$ where $a \neq 0$) form an abelian group under multiplication in which $[d, d]$ acts as the unit element and where $[c, d]^{-1} = [d, c]$.

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

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

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

Remark 1.5.3. If F is a field, then the field of quotient of F is F itself.

Example 1.5.4.

1. $\mathbb{Q}$ is the field of quotient of $\mathbb{Z}$
2. If F is a field, then $F[x]$ is an integral domain and so $F(x)$ is the field of quotient of $F[x]$.

Chapter 2

Unit 2

2.1 Euclidean Rings

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

Definition 2.1.1. An integral domain R is said to be a **Euclidean ring** if for every $a \neq 0$ in R there is defined a nonnegative integer $d(a)$ such that

1. For all $a, b \in R$, both nonzero, $d(a) \leqq d(ab)$.

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

We do not assign a value to $d(O)$. The integers serve as an example of a Euclidean ring, where $d(a) =$ absolute value of a acts as the required function. In the next section we shall see that the Gaussian integers also form a Euclidean ring. Out of that observation, and the results developed in this part, we shall prove a classic theorem in number theory due to Fermat, namely, that every prime number of the form $4n + 1$ can be written as the sum of two squares.

Theorem 2.1.2. *Let R be a Euclidean ring and let A be an ideal of R. Then there exists an element $a_0 \in A$ such that A consists exactly of all a_0x as x ranges over R.* □

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

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

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

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

Definition 2.1.3. An integral domain R with unit element is a **principal ideal ring** if every ideal A in R is of the form $A = (a)$ for some $a \in R$.

.

Corollary 2.1.4. *A Euclidean ring possesses a unit element.*

Proof. Let R be a Euclidean ring; then R is certainly an ideal of R, so that by Theorem 2.1.2 we may conclude that $R = (u_0)$ for some $u_0 \in R$. Thus every element in R is a multiple of u_0. Therefore, in particular, $u_0 = u_0 c$ for some $c \in R$. If $a \in R$ then $a = xu_0$ for some $x \in R$, hence $ac = (xu_0)c = x(u_0 c) = xu_0 = a$. Thus c is seen to be the required unit element. □

Definition 2.1.5. If $a \neq 0$ and b are in a commutative ring R then a is said to divide b if there exists a $c \in R$ such that $b = ac$. We shall use the symbol $a \mid b$ to represent the fact that a divides b and $a \nmid b$ to mean that a does not divide b.

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

Remark 2.1.6.

1. If $a \mid b$ and $b \mid c$ then $a \mid c$.

2. If $a \mid b$ and $a \mid c$ then $a \mid (b \pm c)$.

3. If $a \mid b$ then $a \mid bx$ for all $x \in R$.

Definition 2.1.7. If $a, b \in R$ then $d \in R$ is said to be a greatest common divisor of a and b if

1. $d \mid a$ and $d \mid b$.

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

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

Lemma 2.1.8. *Let R be a Euclidean ring. Then any two elements a and b in R have a greatest common divisor d. Moreover $d = \lambda a + \mu b$ for some $\lambda, \mu \in R$.*

□

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

Since A is an ideal of R, there exists an element $d \in A$ such that every element in A is a multiple of d. By hint of the fact that $d \in A$ and that every element of A is of the form $ra + sb$, $d = \lambda a + \mu b$ for some $\lambda, \mu \in R$. Now by the above corollary, R has a unit element 1; thus

$a = la + Ob \in A$, $b = Oa + 1b \in A$. Being in A, they are both multiples of d, whence $d \mid a$ and $d \mid b$.

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

Definition 2.1.9. Let R be a commutative ring with unit element. An element $a \in R$ is a **unit** in R if there exists an element $b \in R$ such that $ab = 1$.

A unit in a ring is an element whose inverse is also in the ring.

Lemma 2.1.10. *Let R be an integral domain with unit element and suppose that for $a, b \in R$ both $a \mid b$ and $b \mid a$ are true. Then $a = ub$, where u is a unit in R.*

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

Definition 2.1.11. Let R be a commutative ring with unit element. Two elements a and b in R are said to be **associates** if $b = ua$ for some unit u in R.

The relation of being associates is an equivalence relation. Note that in a Euclidean ring any two greatest common divisors of two given elements are associates.

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

Lemma 2.1.12. *Let R be a Euclidean ring and $a, b \in R$. If $b \neq 0$ is not a unit in R, then $d(a) < d(ab)$.*

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

Definition 2.1.13. In the Euclidean ring R a nonunit π is said to be a **prime element** of R if whenever $\pi = ab$, where a, b are in R, then one of a or b is a unit in R.

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

Lemma 2.1.14. *Let R be a Euclidean ring. Then every element in R is either a unit in R or can be written as the product of a finite number of prime elements of R.*

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

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

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

If a is a prime element of R there is nothing to prove. So suppose that $a = bc$ where neither b nor c is a unit in R. By above Lemma, $d(b) < d(bc) = d(a)$ and $d(c) < d(bc) = d(a)$. Thus by our induction hypothesis b and c can be written as a product of a finite number of prime elements of R; $b = \pi_1\pi_2\cdots\pi_n$, $c = \pi'_1\pi'_2\cdots\pi'_m$ where the π's and π''s are prime elements of R. Consequently $a = bc = \pi_1\pi_2\cdots\pi_n\pi'_1\pi'_2\cdots\pi'_m$ and in this way a has been factored as a product of a finite number of prime elements. This completes the proof. □

Definition 2.1.15. In the Euclidean ring R, a and b in R are said to be **relatively prime** if their greatest common divisor is a unit of R.

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

Lemma 2.1.16. *Let R be a Euclidean ring. Suppose that for $a, b, c \in R$, $a \mid bc$ but $(a, b) = 1$. Then $a \mid c$.*

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

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

Lemma 2.1.17. *If π is a prime element in the Euclidean ring R and $\pi \mid ab$ where $a, b \in R$ then π divides at least one of a or b.*

Proof. Suppose that π does not divide a; then $(\pi, a)=1$. Applying Lemma 2.1.16 we are led to $\pi \mid b$. □

Corollary 2.1.18. *If π is a prime element in the Euclidean ring R and $\pi \mid a_1a_2\cdots a_n$ then π divides at least one $a_1, a_2, \ldots, a_n$.*

Theorem 2.1.19. ***(Unique Factorization Theorem)*** *Let R be a Euclidean ring and $a \neq 0$ a nonunit in R. Suppose that $a=\pi_1\pi_2\cdots\pi_n=\pi_1'\pi_2'\cdots\pi_m'$ where the π_i and π_j' are prime elements of R. Then $n=m$ and each π_i, $1 \leq i \leq \pi$ is an associate of some π_j', $1 \leq j \leq m$ and conversely each π_k' is an associate of some π_q.*

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

From above arguments, we have that every nonzero element in a Euclidean ring R can be uniquely written (up to associates) as a product of prime elements or is a unit in R.

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

Now we proved that any ideal A in the Euclidean ring R is of the form $A = (a_0)$ where $(a_0) = \{xa_0 \mid x \in R\}$. □

Lemma 2.1.20. *The ideal $A = (a_0)$ is a maximal ideal of the Euclidean ring R if and only if a_0 is a prime element of R.*

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

$A \subset B$. We claim that $A \neq B$ and that $B \neq R$.

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

Conversely, suppose that a_0 is a prime element of R and that U is an ideal of R such that $A = (a_0) \subset U \subset R$. By above Theorem, $U = (u_0)$ Since $a_0 \in A \subset U = (u_0)$, $a_0 = xu_0$ for some $x \in R$. But a_0 is a prime element of R, from which it follows that either x or u_0 is a unit in R. If u_0 is a unit in R then $U = R$. If, on the other hand, x is a unit in R, then $x^{-1} \in R$ and the relation $a_0 = xu_0$ becomes $u_0 = x^{-1}a_0 \in A$ since A is an ideal of R. This implies that $U \subset A$; together with $A \subset U$ we conclude that $U = A$. Therefore there is no ideal of R which fits strictly between A and R. This means that A is a maximal ideal of R. □

2.2 A Particular Euclidean Ring

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

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

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

1. $d(x)$ is a nonnegative integer for every $x \neq 0 \in J[i]$.

2. $d(x) \leq d(xy)$ for every $y \neq 0$ in $J[i]$.

3. Given $u, v \in J[i]$ there exist $t, r \in J[i]$ such that $v = tu + r$ where $r = 0$ or $d(r) < d(u)$.

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

Theorem 2.2.1. $J[i]$ *is a Euclidean ring.*

Proof. As was remarked in the discussion above, to prove Theorem 2.2.1 we merely must show that, given $x, y \in J[i]$ there exists $t, r \in J[i]$

such that $y = tx + r$ where $r = 0$ or $d(r) < d(x)$.

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

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

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

Lemma 2.2.2. *Let p be a prime integer and suppose that for some integer c relatively prime to p we can find integers x and y such that $x^2 + y^2 = cp$. Then p can be written as the sum of squares of two integers, that is, there exist integers a and b such that $p = a^2 + b^2$.*

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

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

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

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

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

$a^2 + b^2 = p$ and the lemma is thereby established.

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

Lemma 2.2.3. *If p is a prime number of the form $4n + 1$, then we can solve the congruence $x^2 \equiv -1 \mod p$.*

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

$$x \equiv (-1)(-2)(-3)\cdots\Big(-(\frac{p-1}{2})\Big)$$

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

$$\begin{aligned} x^2 &\equiv \Big(1 \cdot 2 \cdots \frac{p-1}{2}\Big)(-1)(-2)\cdots\Big(-(\frac{p-1}{2})\Big) \\ &\equiv 1 \cdot 2 \cdots \frac{p-1}{2}\frac{p+1}{2}\cdots(p-1) \\ &\equiv (p-1)! = 1 \mod p \end{aligned}$$

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

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

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

□

Theorem 2.2.4. ***(FERMAT)*** *If p is a prime number of the form $4n+1$, then $p = a^2 + b^2$ for some integers a, b.*

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

Chapter 3

Unit 3

3.1 Polynomial Rings

Let F be a field. By the **ring of polynomials** in the indeterminate, x, written as $F[x]$, we mean the set of all symbols $a_0+a_1x+\cdots+a_nx^n$, where n can be any nonnegative integer and where the coefficients $a_1, a_2, \ldots, a_n$ are all in F.

In order to make a ring out of $F[x]$ we must be able to recognize when two elements in it are equal, we must be able to add and multiply elements of $F[x]$ so that the axioms defining a ring hold true for $F[x]$. This will be our initial goal.

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

Definition 3.1.1. If $p(x) = a_0 + a_1x + \cdots + a_mx^m$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$ are in $F[x]$, then $p(x) = q(x)$ if and only if for every integer $i \geq 0$, $a_i = b_i$.

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

Definition 3.1.2. If $p(x) = a_0 + a_1x + \cdots + a_mx^m$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$ are both in $F[x]$, then $p(x) + q(x) = c_0 + c_1x + \cdots + c_tx^t$ where for each i, $c_i = a_i + b_i$.

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

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

Definition 3.1.3. If $p(x) = a_0 + a_1x + \cdots + a_mx^m$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$, then $p(x)q(x) = c_0 + c_1x + \cdots + c_kx^k$ where

$$c_t = a_tb_0 + a_{t-1}b_1 + a_{t-2}b_2 + \cdots + a_0b_t.$$

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

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

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

$$\begin{aligned}
c_0 &= a_0b_0 = 1.2 = 2\ ,\\
c_1 &= a_1b_0 + a_0b_1 = 1.2 + 1.0 = 2,\\
c_2 &= a_2b_0 + a_1b_1 + a_0b_2 = (-1)(2) + 1.0 + 1.1 = -1,\\
c_3 &= a_3b_0 + a_2b_1 + a_1b_2 + a_0b_3 = (0)(2) + (-1)(0) + 1.1 + 1.1 = 2,\\
c_4 &= a_4b_0 + a_3b_1 + a_2b_2 + a_1b_3 + a_0b_4\\
&= (0)(2) + (0)(0) + (-1)(1) + (1)(1) + 1(0) = 0,\\
c_5 &= a_5b_0 + a_4b_1 + a_3b_2 + a_2b_3 + a_1b_4 + a_0b_5\\
&= (0)(2) + (0)(0) + (0)(1) + (-1)(1) + (1)(0) + (0)(0) = -1,\\
c_6 &= a_6b_0 + a_5b_1 + a_4b_2 + a_3b_3 + a_2b_4 + a_1b_5 + a_0b_6\\
&= (0)(2) + (0)(0) + (0)(1) + (0)(1) + (-1)(0) + (1)(0) + (1)(0) = 0,\\
c_7 &= c_8 = \cdots = 0.
\end{aligned}$$

Therefore according to our definition,

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

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

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

Definition 3.1.4. If $f(x) = a_0 + a_1x + \cdots + a_nx^n \neq 0$ and $a_n \neq 0$ then the **degree** of $f(x)$, written as deg $f(x)$, is n.

That is, the degree of $f(x)$ is the largest integer i for which the ith coefficient of $f(x)$ is not 0. We do not define the degree of the zero polynomial. We say a polynomial is a constant if its degree is 0. The degree function defined on the nonzero elements of $F[x]$ will provide us with the function $d(x)$ needed in order that $F[x]$ be a Euclidean ring.

Lemma 3.1.5. *If $f(x), g(x)$ are two nonzero elements of $F[x]$, then deg $(f(x)g(x))$ = deg $f(x)$ + deg $g(x)$.*

Proof. Suppose that $f(x) = a_0 + a_1x + \cdots + a_mx^m$ and $g(x) = b_0 + b_1x + \cdots + b_nx^n$ and that $a_m \neq 0$ and $b_n \neq 0$. Therefore deg $f(x) = m$ and deg $g(x) = n$. By definition, $f(x)g(x) = c_0 + c_1x + \cdots + c_kx^k$ where $c_1 = a_tb_0 + a_{t-1}b_1 + \cdots + a_1b_{t-1} + a_0b_t$. We claim that $cm + n = a_mb_n \neq 0$ and $c_i = 0$ for $i > m + n$. That $c_{m+n} = a_mb_n$ can be seen at a glance by its definition. What about c_i for $i > m + n$? c_i is the sum of terms of the form a_jb_{i-j}; since $i = j + (i - j) > m + n$ then either $j > m$ or $(i - j) > n$. But then one of a_i or b_{i-j} is 0, so that $a_jb_{i-j} = 0$; since c_i is the sum of a bunch of zeros it itself is 0, and our claim has been established. Thus the highest nonzero coefficient of $f(x)g(x)$ is c_{m+n} whence deg $f(x)g(x) = m + n =$ deg $f(x)$ + deg $g(x)$. □

Corollary 3.1.6. *If $f(x), g(x)$ are nonzero elements in $F[x]$ then deg $f(x) \leq$deg $f(x)g(x)$.*

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

Corollary 3.1.7. *If F is a field, then $F[x]$ is an integral domain.*

We leave the proof of this corollary to the reader.

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

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

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

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

Lemma 3.1.8. ***(The Division Algorithm)*** *Given two polynomials $f(x)$ and $g(x) \neq 0$ in $F[x]$, then there exist two polynomials $t(x)$ and $r(x)$ in $F[x]$ such that $f(x) = t(x)g(x) + r(x)$ where $r(x) = 0$ or deg $r(x) <$ deg $g(x)$.*

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

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

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

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

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

In view of division algorithm, we have

Corollary 3.1.9. *$F[x]$ is a Euclidean ring.*

Lemma 3.1.10. *$F[x]$ is a principal ideal ring.*

Lemma 3.1.11. *Given two polynomials $f(x), g(x)$ in $F[x]$ they have a greatest common divisor $d(x)$ which can be realized as $d(x) = \lambda(x)f(x) + \mu(x)g(x)$.*

Definition 3.1.12. A polynomial $p(x)$ in $F[x]$ is said to be **irreducible** over F if whenever $p(x) = a(x)b(x)$ with $a(x), b(x) \in F[x]$, then one of $a(x)$ or $b(x)$ has degree 0 (i.e., is a constant).

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

Lemma 3.1.13. *Any polynomial in $F[x]$ can be written in a unique manner as a product of irreducible polynomials in $F[x]$.*

Lemma 3.1.14. *The ideal $A = (p(x))$ in $F[x]$ is a maximal ideal if and only if $p(x)$ is irreducible over F.*

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

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

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

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

$(a_0\alpha + 2a_2\beta + 2a_1\gamma) + (a_1\alpha + a_0\beta + 2a_2\gamma)t + (a_2\alpha + a_1\beta + a_0\gamma)t^2 = 1$;thus

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

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

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

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

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

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

$$a_0^3 + 2a_1^3 + 4a_2^3 = 6a_0a_1a_2$$

is the solution $a_0 = a_1 = a_2 = 0$. We now proceed to show this. If a solution exists in rationals, by clearing of denominators we can show that a solution exists where a_0, a_1, a_2 are integers. Thus we may assume that a_0, a_1, a_2 are integers satisfying the above equation. We now assert that we may assume that a_0, a_1, a_2 have no common divisor other than 1, for if $a_0 = b_0d$, $a_1 = b_1d$, and $a_2 = b_2d$, where d is their greatest common divisor, then substituting in the above equation we obtain $d^3(b_0^3 + 2b_1^3 + 4b_2^{)} = d^3(6b_0b_1b_2)$, and so $b_0^3 + 2b_1^3 + 4b_2^3 = 6b_0b_1b_2$. The problem has thus been reduced to proving that the above equation has no solutions in integers which are relatively prime. But then the above equation implies that a_0^3 is even, so that a_0 is even; substituting $a_0 = 2\alpha_0$ in the above equation gives us $4\alpha_0^3 + a_1^3 + 2a_2^3 = 6a_0a_1a_2$.

Thus a_1^3, and so, a_1 is even; $a_1 = 2\alpha_1$. Substituting in the above equation we obtain $2\alpha_0^3 + 4\alpha_1^3 + a_2^3 = 6a_0a_1a_2$.Thus a_2^3, and so a_2 , is even! But then a_0, a_1, a_2 have 2 as a common factor! This contradicts that they are relatively prime, and we have proved that the equation $a_0^3 + 2a_1^3 + 4a_2^3 = 6a_0a_1a_2$ has no rational solution other than $a_0 = a_1 = a_2 = 0$. Therefore we can solve for α, β, γ and $F[x]/(x^3 - 2)$ is seen, directly, to be a field.

3.2 Polynomials over the Rational Field

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

Definition 3.2.1. The polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n$, where the $a_0, a_1, a_2, \ldots, a_n$ are integers is said to be primitive if the greatest common divisor of $a_0, a_1, ..., a_n$ is 1.

Lemma 3.2.2. *If $f(x)$ and $g(x)$ are primitive polynomials, then $f(x)g(x)$ is a primitive polynomial.*

Proof. Let $f(x) = a_0+a_1x+\cdots+a_nx^n$ and $g(x) = b_0+b_1x+\cdots+b_mx^m$. Suppose that the lemma was false; then all the coefficients of $f(x)g(x)$

would be divisible by some integer larger than 1, hence by some prime number p. Since $f(x)$ is primitive, p does not divide some coefficient a_i. Let a_i be the first coefficient of $f(x)$ which p does not divide. Similarl let b_k be the first coefficient of $g(x)$ which p does not divide. In $f(x)g(x)$ the coefficient of x^{j+k}, c_{j+k} is

$$\begin{aligned} c_{j+k} &= a_j b_k + (a_{j+1}b_{k-1} + a_{j+2}b_{k-2} + \cdots + a_{j+k}b_0) \\ &\quad + (a_{j-1}b_{k+1} + a_{j-2}b_{k+2} + \cdots + a_0 b_{j+k}). \end{aligned}$$

Now by our choice of $b_k, p \mid b_{k-1}, b_{k-2}, \ldots$ so that $p \mid (a_{j+l}b_{k-1} + a_{j+2}b_{k-2} + \cdots + a_{j+k}b_0)$. Similarly, by our choice of a_j, $p \mid a_{j-1}, a_{j-2}, \ldots$ so that $p \mid (a_{j-1}b_{k+1} + a_{j-2}b_{k+2} \cdots + a_0 b_{k+i})$. By assumption, $p \mid c_{j+k}$. Thus by the above equation, $p \mid a_j b_k$, which is nonsense since $p \nmid a_j$ and $p \nmid b_k$. This proves the lemma. □

Definition 3.2.3. The **content** of the polynomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n$, where the a's are integers, is the greatest common divisor of the integers $a_0, a_1, \ldots, a_n$.

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

Theorem 3.2.4. ***(Gauss' Lemma)*** *If the primitive polynomial $f(x)$ can be factored as the product of two polynomials having rational coefficients, it can be factored as the product of two polynomials having integer* *coefficients.*

Proof. Suppose that $f(x) = u(x)v(x)$ where $u(x)$ and $v(x)$ have rational coefficients. By clearing of denominators and taking out common factors we can then write $f(x) = (a/b)\lambda(x)\mu(x)$ where a and b are integers and where both $\lambda(x)$ and $\mu(x)$ have integer coefficients and are primitive. Thus $bf(x) = a\lambda(x)\mu(x)$.

The content of the left-hand side is b, since $f(x)$ is primitive; since both $\lambda(x)$ and $\mu(x)$ are primitive, $\lambda(x)\mu(x)$ is primitive, so that the content of the right-hand side is a. Therefore $a = b$, $(a/b) = 1$, and $f(x) = \lambda(x)\mu(x)$ where $\lambda(x)$ and $\mu(x)$ have integer coefficients. This is the assertion of the theorem. □

Definition 3.2.5. A polynomial is said to be **integer monic** if all its coefficients are integers and its highest coefficient is 1.

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

Corollary 3.2.6. *If an integer monic polynomial factors as the product of two nonconstant polynomials having rational coefficients then it factors as the product of two integer monic polynomials*

Theorem 3.2.7. ***(The Eisentein Criterion)*** *Let* $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ *be a polynomial with integer coefficients. Suppose that* *for some prime number* p*,* $p \nmid a_n$*,* $p \mid a_1, p \mid a_2, \ldots, p \mid a_0$ *,* $p^2 \nmid a_0$*. Then* $f(x)$ *is irreducible over the rationals.*

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

$$f(x) = (b_0 + b_1 x + \cdots + b_r x^r)(c_0 + c_1 x + \cdots + c_s x^s),$$

where the b's and c's are integers and where $r > 0$ and $s > 0$. Reading off the coefficients we first get $a_0 = b_0 c_0$. Since $p \mid a_0$, p must divide one of b_0 or c_0 . Since $p^2 \nmid a_0$, p cannot divide both b_0 and c_0. Suppose that $p \mid b_0$, $p \nmid c_0$. Not all the coefficients $b_0, \ldots, b_r$ can be divisible by p; otherwise all the coefficients of $f(x)$ would be divisible by p, which is manifestly false since $p \nmid a_n$. Let b_k be the first b not divisible by p, $k \leq r < n$. Thus $p \mid b_{k-1}$ and the earlier b's. But $a_k = b_k c_0 + b_{k-1} c_1 + b_{k-2} c_2 + \cdots + b_0 c_k$, and $p \mid a_k$, $p \mid b_{k-1}$, b_{k-2} , $\cdots$, b_0 , so that $p \mid b_k c_0$. However, $p \nmid c_0$, $p \nmid b_k$, which conflicts with $p \mid b_k c_0$. This contradiction proves that we could not have factored $f(x)$ and so $f(x)$ is indeed irreducible. □

Example 3.2.8. *Let* $f(x) = x^3 - 3 \in \mathbb{Z}[x]$*. By the Eisentein Criterion,* $p = 3$ *and* $p|3$*,* $p|0$ *and* p *does not divide* 1*,* $f(x)$ *is irreducible over* $\mathbb{Z}$*.*

3.3 Polynomial Rings over Commutative Rings

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

Let R be a commutative ring with unit element. By the polynomial ring in x over R, $R[x]$, we shall mean the set of formal symbols $a_0+a_1x+\cdots+a_mx^m$, where $a_0, a_1, \ldots, a_m$ are in R, and where equality, addition, and multiplication are defined exactly as they were in Section 3.1. As in that section, $R[x]$ is a commutative ring with unit element.

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

Of particular importance is the case in which $R = F$ is a field; here we obtain the ring of polynomials in n-variables over a field. Of interest to us will be the influence of the structure of R on that of $R[x_1, \cdots, x_n]$.

Lemma 3.3.1. *If R is an integral domain, then so is $R[x]$.*

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

Corollary 3.3.2. *If R is an integral domain, then so is $R[x_1, \ldots, x_n]$.*

In particular, when F is a field, $F[x_1, \ldots, x_n]$ must be an integral domain.As such, we can construct its field of quotients; we call this the field of rational functions in $x_1, \ldots, x_n$ over F and denote it by $F(x_1, \ldots, x_n)$. This field plays a vital role in algebraic geometry. In arbitrary integral domains, R, with unit element. Two elements a, b in R are said to be **associates** if $a = ub$ where u is a unit in R.

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

An irreducible element is thus an element which cannot be factored in a "nontrivial" way.

Definition 3.3.3. An integral domain, R, with unit element is a **unique factorization domain** if

(i) Any nonzero element in R is either a unit or can be written as the

product of a finite number of irreducible elements of R.
(ii) The decomposition in part (a) is unique up to the order and associates of the irreducible elements.

From the above theorem, a Euclidean ring is a unique factorization domain. The converse, however, is false; for example, the ring $F[x_1, x_2]$, where F is a field, is not even a principal ideal ring, but as we shall soon see it is a unique factorization domain.

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

Lemma 3.3.4. *If R is a unique factorization domain and if a, b are in R, then a and b have a greatest common divisor (a, b) in R. Moreover, if a and b are relatively prime (i.e., $(a, b) = 1$), whenever $a \mid bc$ then $a \mid c$.*

Corollary 3.3.5. *If $a \in R$ is an irreducible element and $a \mid bc$, then $a \mid b$ or $a \mid c$.*

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

Given the polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$ in $R[x]$, then the content of $f(x)$ is defined to be the greatest common divisor of

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

Lemma 3.3.6. *If R is a unique factorization domain, then the product of two primitive polynomials in $R[x]$ is again a primitive polynomial in $R[x]$.*

Proof. Given $f(x), g(x)$ in $R[x]$ we can write $f(x) = af_1(x), g(x) = bg_1(x)$, where $a = c(f), b = c(g)$ and where $f_1(x)$ and $g_1(x)$ are primitive. Thus $f(x)g(x) = abf_1(x)g_1(x)$. By Lemma 3.3.6,$f_1(x)g_1(x)$ is primitive. Hence the content of $f(x)g(x)$ is ab, that is, it is $c(f)c(g)$. □

Corollary 3.3.7. *If R is a unique factorization domain and if $f(x), g(x)$ are in $R[x]$, then $c(fg) = c(f)c(g)$ (up to units).*

By a simple induction, the corollary extends to the product of a finite number of polynomials to read $c(f_1f_2\cdots f_k) = c(f_1)c(f_2)\cdots c(f_k)$. Let R be a unique factorization domain. Being an integral domain, it has a field of quotients F. We can consider $R[x]$ to be a subring of $F[x]$. Given any polynomial $f(x) \in F[x]$, then $f(x) = (f_0(x)/a)$, where $f_0(x) \in R[x]$ awhere $a \in R$.

It is natural to ask for the relation, in terms of reducibility and irreducibility, of a polynomial in $R[x]$ considered as a polynomial in the larger ring $F[x]$.

Lemma 3.3.8. *If $f(x)$ in $R[x]$ is both primitive and irreducible as an element of $R[x]$, then it is irreducible as an element of $F[x]$. Conversely, if the primitive element $f(x)$ in $R[x]$ is irreducible as an element of $F[x]$, it is also irreducible as an element of $R[x]$.*

Proof. Suppose that the primitive element $f(x)$ in $R[x]$ is irreducible in $R[x]$ but is reducible in $F[x]$. Thus $f(x) = g(x)h(x)$, where $g(x), h(x)$ are in $F[x]$ and are of positive degree. Now $g(x) = (g_0(x)/a)$, $h(x) = (h_0(x)/b)$, where $a, b \in R$ and where $g_0(x), h_0(x) \in R[x]$. Also $g_0(x) = \alpha g_1(x)$,$h_0(x) = \beta h_1(x)$, where $\alpha = c(g_0)$, $\beta = c(h_0)$, and $g_1(x), h_1(x)$ are primitive in $R[x]$. Thus $f(x) = (\alpha\beta/ab)g_1(x)h_1(x)$, whence $abf(x) = \alpha\beta g_1(x)h_1(x)$. By Lemma 3.3.6, $g_1(x)h_1(x)$ is primitive, whence the content of the righthand side is $\alpha\beta$. Since $f(x)$ is primitive, the content of the left-hand side is ab; but then $ab = \alpha\beta$; the implication of this is that $f(x) = g_1(x)h_1(x)$, and we have obtained a nontrivial factorization of $f(x)$ in $R[x]$, contrary to hypothesis. (Note: this factorization is nontrivial since each of $g_1(x), h_1(x)$ are of the same degree as $g(x)$, $h(x)$, so cannot be units in $R[x]$. We leave the converse half of the lemma as an exercise. □

Lemma 3.3.9. *If R is a unique factorization domain and if $p(x)$ is a* *primitive polynomial in $R[x]$, then it can be factored in a unique way as the product of irreducible elements in $R[x]$.*

Proof. When we consider $p(x)$ as an element in $F[x]$, we can factor it as $p(x) = p_1(x) \cdots p_k(x)$, where $p_1(x), p_2(x), \ldots, p_k(x)$ are irreducible polynomials in $F[x]$. Each $p_i(x) = (f_i(x)/a_i)$, where $f_i(x) \in R[x]$ and $a_i \in R$; moreover, $f_i(x) = c_i q_i(x)$, where $c_i = c(f_i)$ and where $q_i(x)$ is primitive in $R[x]$. Thus each $p_i(x) = (c_i q_i(x)/a_i)$, where $a_i, c_i \in R$ and where $q_i(x) \in R[x]$ is primitive. Since $p_i(x)$ is irreducible in $F[x]$, $q_i(x)$ must also be irreducible in $F[x]$, hence it is irreducible in $R[x]$.

Now

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

whence $a_1 a_2 \cdots a_k p(x) = c_1 c_2 \cdots c_k q_1(x) \cdots q_k(x)$. Using the primitivity of $p(x)$ and of $q_1(x) \cdots q_k(x)$, we can read off the content of the left-hand side as $a_1 a_2 \cdots a_k$ and that of the right-hand side as $c_1 c_2 \cdots c_k$. Thus $a_1 a_2 \cdots a_k = c_1 c_2 \cdots c_k$, hence $p(x) = q_1(x) \cdots q_k(x)$. We have factored $p(x)$, in $R[x]$, as a product of irreducible elements. Can we factor it in another way? If $p(x) = r_1(x) \cdots r_k(x)$, where the $r_i(x)$ are irreducible in $R[x]$, by the primitivity of $p(x)$, each $r_i(x)$ must be primitive, hence irreducible in $F[x]$ by Lemma 3.3.8. But by Lemma 3.1.5 we know unique factorization in $F[x]$; the net result of this is that the $r_i(x)$ and the $q_i(x)$ are equal (up to associates) in some order, hence $p(x)$ has a unique factorization as a product of irreducibles in $R[x]$.

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

Theorem 3.3.10. *If R is a unique factorization domain, then so is $R[x]$.*

Proof. Let $f(x)$ be an arbitrary element in $R[x]$. We can write $f(x)$ in a unique way as $f(x) = cf_1(x)$ where $c = c(f)$ is in R and where $f_1(x)$,in $R[x]$, is primitive. By Lemma 3.3.9 we can decompose $f_1(x)$ in a unique way as the product of irreducible elements of $R[x]$. What about c? Suppose that $c = a_1(x)a_2(x) \cdots a_m(x)$ in $R[x]$; then $O = \deg c = \deg (a_1(x)) + \deg (a_2(x)) + \cdots + \deg (a_m(x))$. Therefore, each $a_i(x)$ must be of degree 0, that is, it must be an element of R. In other words, the only factorizations of c as an element of $R[x]$ are those it had as an element of R. In particular, an irreducible element in R is still irreducible in $R[x]$. Since R is a unique factorization domain, c has a unique factorization as a product of irreducible elements of R, hence of $R[x]$.

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

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

Corollary 3.3.11. *If R is a unique factorization domain then so is $R[x_1, \ldots, x_n]$.*

Corollary 3.3.12. *If F is a field then $F[x_1, \ldots, x_n]$ is a unique factorization domain.*

Chapter 4

Unit 4

4.1 Semisimple ring

Definition 4.1.1. The **Jacobson radical** of a ring R, denoted by rad R, is the set

$$\text{rad } R = \cap\{M | M \text{ is a maximal ideal of } R\}.$$

If rad $R = \{0\}$, then R is said to be a ring without Jacobson radical or, more briefly, R is a **semisimple ring**.

The Jacobson radical always exists, since we know that any commutative ring with identity contains at least one maximal ideal. It is also immediately obvious from the definition that rad R forms an ideal of R which is contained in each maximal ideal.

To fix ideas, let us determine the Jacobson radical in several concrete rings.

Example 4.1.2. The ring Z of integers is a semisimple ring. For, according the maximal ideals of Z are precisely the principal ideals generated by the prime numbers; thus,

$$\text{rad } R = \cap\{(p)|p \text{ is a prime number}\}.$$

Since no nonzero integer is divisible by every prime, we see at once that rad $R = \{0\}$.

Example 4.1.3. A more penetrating illustration is furnished by the ring $R =$ map (X, F), where X is an arbitrary set and F a field. For any element $x \in X$, consider the function $1\tau_x f = f(x)$ which assigns to each function in R its value at x. It is easily checked that τ_x is a homomorphism of R into F; since R contains all the constant functions, this homomorphism actually maps onto the field F. Thus,its kernel is the maximal ideal

$$M_x = \{f \in R|f(x) = 0\}.$$

Because rad $R \subseteq \cap M_x = \{f \in R|f(x) = 0 \text{ for all } x \in X\} = \{0\}$, it follows that R must be a semisimple ring.

Example 4.1.4. For a final example, we turn to the ring $R[[x]]$ of formal power series. Here, there is a one-to-one correspondence between the maximal ideals M of R and maximal ideals M' of $R[[x]]$ in such a way that M' corresponds to M if and only if M' is generated by M and x. Thus,

$$\begin{aligned}\text{rad } R[[x]] &= \cap\{M' | M' \text{ is a maximal ideal of } R[[x]]\}\\ &= \cap\{(M,x)|M \text{ is a maximal ideal of } R\}\\ &= (\cap M, x) = (\text{rad } R, x).\end{aligned}$$

In particular, if R is taken to be a field F, we have rad $F[[x]] = (x)$, the principal ideal generated by x.

Our first theorem establishes a basic connection between the Jacobson radical and invertibility of ring elements.

Theorem 4.1.5. *Let I be an ideal of the ring R. Then $I \subseteq$ rad R if and only if each element of the coset $1 + I$ has an inverse in R.*

Proof. We begin by assuming that $I \subseteq$ rad R and that there is some element $a \in I$ for which $1 + a$ is not invertible. Our object, of course, is to derive a contradiction. By the corollary to Theorem 53, the element $1 + a$ must belong to some maximal ideal M of the ring R. Since $a \in$ rad R, a is also contained in M, and therefore $1 = (1 + a) - a$ lies in M. But this means that $M = R$, which is clearly impossible.

To prove the converse, suppose that each member of $1 + I$ has a multiplicative inverse in R, but $I \nsubseteq$ rad R. By definition of the Jacobson radical, there will exist a maximal ideal M of R with $I \nsubseteq M$. Now, if a is any element of I which is not in M, the maximality of M implies that the ideal $(M, a) = R$. Knowing this, the identity element 1 can be expressed in the form $1 = m + ra$ for suitable choice of $m \in M$ and $r \in R$. But then, $m = 1 - ra \in 1 + I$, so that m possesses an inverse.

The conclusion is untenable, since no proper ideal contains an invertible element.

The form which this result takes when I is the principal ideal generated by $a \in \text{rad } R$ furnishes a characterization of the Jacobson radical in terms of elements rather than ideals. Although actually a corollary to the theorem just proved, it is important enough to be singled out as a theorem. □

Theorem 4.1.6. *In any ring R, an element $a \in$ rad R if and only if $1 - ra$ is invertible for each $r \in R$.*

This theorem adapts itself to many uses. Three fairly short and instructive applications are presented below.

Corollary 4.1.7. *An element a is invertible in the ring R if and only if the coset $a+$ rad R is invertible in the quotient ring $R/$rad R.*

Proof. Assume that the coset $a+\text{rad } R$ has an inverse in $R/\text{rad } R$, so that

$$(a+\text{rad } R)(b+ \text{ rad } R) = 1+ \text{ rad } R$$

for some $b \in R$. Then $1 - ab$ lies in rad R. By above Theorem, $r = l$, to conclude that the product $ab = 1 - 1(1 - ab)$ is invertible; this, in turn, forces the element a to have an inverse in R. The other direction of the corollary is all but obvious. □

Corollary 4.1.8. *The only idempotent element in rad R is 0.*

Proof. Let the element $a \in$ rad R with $a^2 = a$. Taking $r = 1$ in the preceding theorem, we see that $1-a$ has an inverse in R; say $(1-a)b = 1$, where $b \in R$. This leads immediately to

$$a{=}a(1-a)b{=}(a-a^2)b{=}0,$$

which completes the proof. □

Corollary 4.1.9. *Every nil ideal of R is contained in rad R.*

Proof. Let N be a nil ideal of R and suppose that $a \in N$. For every $r \in R$, we then have $ra \in N$, so that the product ra is nilpotent. Therefore implies that $1 - ra$ is invertible in R. This shows that the element a lies in rad R, from which it follows that $N \subseteq$ rad R. □

Although the Jacobson radical of a ring R is not necessarily a nil ideal, very little restriction on R forces it to be nil. We shall see subsequently that, if every ideal of R is finitely generated, then rad R is not only nil but nilpotent.

This is a convenient place to also point out that a homomorphic image of a semisimple ring need not be semisimple. An explicit example of this situation can easily be obtained from the ring $mathbbZ$ of integers. While $\mathbb{Z}$ form a ring without a Jacobson radical, its homomorphic image $\mathbb{Z}_{p^n}$, (p a prime; $n > 1$) contains the nil ideal (p); appealing to above Corollary, we see that $\mathbb{Z}_{p^n}$ cannot be semisimple.

Example 4.1.10. Consider $F[[x]]$, the ring of formal power series over a field F. It is known that an element $f(x) = a_0 + a_1x + \cdots + a_nx^n + \cdots$ has an inverse in $F[[x]]$ if and only if the constant term $a_0 \neq 0$. This observation implies that if $g(x) = b_0 + b_1x + \cdots + b_nx^n + \cdots$, then

$$\begin{aligned}\text{rad } F[[x]] &= \{f(x) : 1 - f(x)g(x) \text{ is invertible for all } g(x) \in F[[x]]\}\\ &=\{f(x) : 1 - a_0b_0 \neq 0 \text{ for all } b_0 \in F\}\\ &=\{f(x) : a_0 = 0\} = (x).\end{aligned}$$

Thus, we have a second proof of the fact that the Jacobson radical of $F[[x]]$ is the principal ideal generated by x.

We next prove several results bearing on the Jacobson radical of quotient rings. The first of these provides a convenient method for manufacturing semisimple rings; its proof utilizes both implications of the last theorem.

Theorem 4.1.11. *For any ring R, the quotient ring $R/rad\ R$ is semisimple; that is, rad$(R/rad\ R)$ $= \{0\}$.*

Proof. Before becoming involved in details, let us remark that since rad R constitutes an ideal of R, we may certainly form the quotient ring $R/$rad R. To simplify notation somewhat, we will temporarily denote rad R by I.

Suppose that the coset $a + I \in$ rad (R/I). Our strategy is to show that the element $a \in I$, for then $a + I = I$, which would imply that rad (R/I) consists of only the zero element of R/I. Since $a + I$ is a member of rad (R/I),

$$(1+I)-(r+I)(a+I)=1-ra+I$$

is invertible in R/I for each choice of $r \in R$. Accordingly, there exists a coset $b+I$ (depending, of course, on both r and a) such that

$$(1-ra+I)(b+I)=1+I.$$

This is plainly equivalent to requiring

$$1-(b-rab) \in I = \text{rad } R.$$

From this, we conclude that the element

$$b-rab=1-1(1-b+rab)$$

has an inverse c in R. But then

$$(1-ra)(bc)=(b-rab)c=1,$$

so that $1ra$ possesses a multiplicative inverse in R. As this argument holds for every $r \in R$, it follows that $a \in \text{rad } R = I$, as desired.

Continuing this theme, let us express the Jacobson radical of the quotient ring R/I as a function of the radical of R. □

Theorem 4.1.12. *If I is an ideal of the ring R, then*
1) rad $(R/I) \supseteq \frac{rad\, R+\, I}{I}$ and,
2) whenever $I \subseteq$ rad R, rad (R/I) =(rad $R)/I$.

Proof. Perhaps the quickest way to establish the first assertion is by means of the Correspondence Theorem; using this, one has

$$\begin{aligned}\text{rad } (R/I) &= \cap\{M'|M' \text{ is a maximal ideal of } R/I\}\\ &= \cap\{\text{nat}_I M|M \text{ is a maximal ideal of } R \text{ with } I \subseteq M\}\\ &\supseteq \text{nat}_I(\text{rad } R + I) = \tfrac{rad\, R+ I}{I}\end{aligned}$$

which is the first part of our theorem (the crucial step requires the inclusion $\bigcap_{I\subseteq M} M \supseteq I+$ rad R).

With an eye to proving (2), notice that whenever $I \subseteq$ rad R, then

$$\text{rad } (R/I) \supseteq \tfrac{rad\, R+ I}{I} \supseteq (\text{rad } R)/I.$$

Thus, we need only to show the inclusion (rad $R)/I \supseteq$ rad (R/I). To this purpose, choose the coset $a + I \in$ rad(R/I) and let M be an arbitrary maximal ideal of R. Since $I \subseteq$ rad $R \subseteq M$, the image nat$_I M = M/I$ must be a maximal ideal of the quotient ring R/I. But then,

$$a + I \in \text{rad}(R/I) \subseteq M/I,$$

forcing the element a to lie in M. As this holds for every maximal ideal of R, it follows that $a \in$ rad R and so $a + I \in$ (rad $R)/I$. All in all, we have proved that rad $(R/I) \subseteq$ (rad $R)/I$, which, combined with our earlier inclusion, leads to (2). □

Theorem 4.1.13. *For any ring R, rad R is the smallest ideal I of R such that the quotient ring R/I is semisimple (in other words, if R/I is I semisimple ring, then rad $R \subseteq I$).*

Proof. It is already known that $R/$rad R is without Jacobson radical. Now, assume that I is any ideal of R for which the associated quotient ring R/I is semisimple. Using part (1) of the preceding theorem, we can then deduce the equality $(I+ \text{ rad } R)/I = I$. This in turn leads to the inclusion rad $R \subseteq I$, which is what we sought to prove.

This may be a good place to mention two theorems concerning the number of maximal ideals in a ring; these are of a rather special character, but typify the results that can be obtained. □

Theorem 4.1.14. *Let R be a principal ideal domain. Then, R is semisimple if and only if R is either a field or has an infinite number of maximal ideals.*

Proof. Let $\{p_i\}$ be the set of prime elements of R. Then the maximal ideals of R are simply the principal ideals (p_i). It follows that an element $a \in$ rad R if and only if a is divisible by each prime p_i if R has an infinite set of maximal ideals, then $a = 0$, since every nonzero noninvertible element of R is uniquely representable as a finite product of primes. On the other hand, if R contains only a finite number of primer $p_1, p_2, \ldots, p_n$, we have

$$\text{rad } R = \cap_{i=1}^{n}(p_i) = (p_1 p_2 \cdots p_n) \neq \{0\},$$

so that R cannot be semisimple. Finally, observe that if the set $\{p_i\}$ is empty, then each nonzero element of R is invertible and R is a field (in which case rad $R = \{0\}$). □

Corollary 4.1.15. *The ring $\mathbb{Z}$ of integers is semisimple.*

Theorem 4.1.16. *Let $\{M_i\}$, $i \in \mathscr{I}$, be the set of maximal ideals of the ring R. If, for each i, there exists an element $a_i \in M_i$ such that $1 - a_i \in$ rad $R - M_i$, then $\{M_i\}$ is a finite set.*

Proof. Suppose that the index set $\mathscr{I}$ is infinite. Then there exists a wellordering $\leq$ of $\mathscr{I}$ under which $\mathscr{I}$ has no last element. (See Appendix A for terminology.) For each $i \in \mathscr{I}$, we define $I_i = \bigcap_{i<j} M_j$. Then $\{I_i\}$ forms a chain of proper ideals of R. By hypothesis, we can select an element $a_i \in M_i$ such that $1 - a_i \in I_i - M_i$. Now the ideal $I = \cup I_i$, is also a proper ideal of R, since $1 \neq I$. By our choice of the I_i, I is not contained in any maximal ideal of R. Indeed, suppose that there does exist an index i for which $I \subseteq M_i$, then,

$$1 - a_i \in I_i \subseteq I \subseteq M_i,$$

yielding the contradiction $1 \in M_i$. But it is known that every proper ideal of R is contained in a maximal ideal of R. From this contradiction we conclude that $\mathscr{I}$ must be finite. □

Let us now turn to a consideration of another radical which plays an essential role in ring theory, to. wit, the prime radical. Its definition may also be framed in terms of the intersection of certain ideals.

4.2 Radical of the ring

Definition 4.2.1. The prime radical of a ring R, denoted by Rad R (in contrast with rad R), is the set

$$\text{Rad } R = \cap\{P | P \text{ is a prime ideal of } R\}.$$

If Rad $R = \{0\}$, we say that the ring R is without prime radical or has zero prime radical.

Lemma 4.2.2. *Let I be an ideal of the ring R. Further, assume that the subset $S \subseteq R$ is closed under multiplication and disjoint from I. Then there exists an ideal P which is maximal in the set of ideals which contain I and do not meet S; any such ideal is necessarily prime.*

Proof. Consider the family $\mathscr{F}$ of all ideals J of R such that $I \subseteq J$ and $J \cap S = \phi$. This family is not empty since I itself satisfies the indicated conditions. Our immediate aim is to show that for any chain of ideals $\{J_i\}$ in $\mathscr{F}$, their union $\cup J_i$ also belongs to $\mathscr{F}$. It has already been established in Theorem 52 that the union of a chain of ideals is again an ideal; moreover, since $I \subseteq J_i$ for each i, we certainly have $I \subseteq \cup J_i$. Finally, observe that

$$(\cup J_i) \cap S = \cup(J_i \cap S) = \cup\phi = \phi.$$

The crux of the matter is that Zorn's Lemma can now be applied to infer that $\mathscr{F}$ has a maximal element P; this is the ideal that we want.

By definition, P is maximal in the set of ideals which contain I but do not meet S. To settle the whole affair there remains simply to show that P is a prime ideal. For this purpose, assume that the product $ab \in P$ but that $a \notin P$ and $b \notin P$. Since it is strictly larger than P, the ideal (P, a) must contain some element r of S; similarly, we can find an element $s \in S$ such that $s \in (P, b)$. This means that

$$rs \in (P, a)(P, b) \subseteq (P, ab) \subseteq P.$$

As S is hypothesized to be closed under multiplication, the product rs also lies in S. But this obviously contradicts the fact that $P \cap S = \phi$. Our argument therefore shows that either a or b is a member of P, which proves that P is a prime ideal. □

Remark 4.2.3. The ideal P need not be a maximal ideal of R, in the usual meaning of the term, but only maximal with respect to exclusion of the set S. To put it another way, if J is any ideal of the ring R which properly contains P, then J must contain elements of S.

Two special cases of this general setting are particularly noteworthy: $S = \{1\}$ and $I = \{0\}$. In the event $S = \{1\}$, the ideal P mentioned in the lemma is actually a maximal ideal (in the usual ideal-theoretic sense); consequently, we have a somewhat different proof of the facts that (i) every proper ideal is contained in a maximal ideal and (ii) each maximal ideal is prime.

The case where I is the zero ideal is the subject of the following corollary, a result which will be utilized on several occasions in the sequel.

Corollary 4.2.4. *Let S be a subset of the ring R which is closed under multiplication and does not contain 0. Then there exists an ideal maximal in the set of ideals disjoint from S; any such ideal is prime.*

As it stands, the preceding lemma is just the opening wedge; we can exploit it rather effectively by now proving.

Theorem 4.2.5. *The intersection of all prime ideals of R which contain a given ideal I is precisely the nil radical of I:*

$$\sqrt{I} = \cap\{P|P \supseteq I;\ P \textit{ is a prime ideal }\}$$

Proof. If the element $a \notin \sqrt{I}$, then the set $S = \{a^n|n \in Z_+\}$ does not intersect I. Since S is closed under multiplication, the preceding lemma insures the existence of some prime ideal P which contains I, but not a; that is, a does not belong to the intersection of prime ideals containing I. This establishes the inclusion

$$\cap\{P|P \supseteq I;\ P \text{ is a prime ideal }\} \subseteq \sqrt{I}$$

The reverse inclusion follows readily upon noting that if there exists a prime ideal which contains I but not a, then $a \notin \sqrt{I}$, since no power of a belongs to P.

As with the case of the Jacobson radical, the prime radical may be characterized by its elements; this is brought out by a result promised earlier. □

Corollary 4.2.6. *The prime radical of a ring R coincides with the nil radical of R; that is, Rad R is simply the ideal of all nilpotent elements of R.*

The assertion is all but obvious upon taking $I = \{0\}$ in Theorem.

An immediate consequence of this last corollary is the potentially powerful statement: every nil ideal of R is contained in the prime radical, not simply contained in the larger Jacobson radical.

Example 4.2.7. For an illustration of above Theorem, let us fall back on the ring $\mathbb{Z}$ of integers. In this setting, the nontrivial prime ideals are the principal ideals (p), where p is a prime number. Given $n > 1$, the ideal $(n) \subseteq (p)$ if and only if p divides n; this being so,

$$\sqrt{(n)} = \bigcap\nolimits_{p_i|n}(p_i)$$

Thus, if we assume that n has the prime power factorization

$$p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}\ (k_i \in Z_+),$$

it follows that

$$\sqrt{(n)} = (p_1) \cap (p_2) \cdots (p_r) = (p_1 p_2 \cdots p_r).$$

Theorem 4.2.8. *An ideal I of the ring R is a semiprime ideal if and only if I is an intersection of prime ideals of R.*

Corollary 4.2.9. *The prime radical Rad R is a semiprime ideal which is contained in every semiprime ideal of R.*

Theorem 4.2.10. *For any ring R, the quotient ring $R/Rad\ R$ is without prime radical.*

Proof. For clarity of exposition, set $I = \text{Rad}\ R$. Suppose that $a + I$ is any nilpotent element of R/I. Then, for some positive integer n,

$$(a + I)^n = a^n + I = I,$$

so that $a^n \in I$. But I consists of all nilpotent elements of R. Thus, we must have $(a^n)^m = 0$ for suitably chosen $m \in Z_+$; this is simply the statement that $a \in I$, and, hence, $a + I$ is the zero element of R/I. Our argument implies that the quotient ring R/I has no nonzero nilpotent elements, which is to say that Rad $(R/I) = \{0\}$.

To round out the picture, two theorems are stated without proof; it will be observed that these take the same form as the corresponding result established for the Jacobson radical. □

Chapter 5

Unit 5

5.1 Partially Ordered sets and Lattices

The most general concept we shall consider in this chapter is that of a partially ordered set. We recall that a binary relation on a set S is a subset R of the product set $S \times S$. We say that a is in the relation R to b and write aRb if and only if $(a, b) \in R$. We now give

Definition 5.1.1. *A partially ordered set is a set S together with a binary relation $a \geq b$ satisfying the following conditions:*

PO1 *If $a \geq a$, then a is (reflexivity).*

PO2 *If $a \geq b$ and $b \geq a$, then $a = b$ (anti-symmetry).*

PO3 *If $a \geq b$ and $b \geq c$, then $a \geq c$ (transitivity).*

If $a \geq b$ and $a \neq b$, then we write $a > b$. Also we write $a \leq b$ as an alternative for $b \geq a$ and $a < b$ for $b > a$. In general we may have neither

$a \geq b$ nor $b \geq a$ for a pair of elements $a, b \in S$. If we do have $a \geq b$ or $b \geq a$ for every pair (a, b), then we call S *totally ordered* (or a *chain*).

We have encountered quite a few examples of partially ordered sets: the set $\mathscr{P}(S)$ of subsets of a set S where $A \geq B$ for subsets A and B means $A \supset B$, the set of subrings of a ring, the set of subgroups of a group, the set of ideals of a ring, and so onall partially ordered by inclusion as defined for subsets. In general, if $S, \geq$ is a partially ordered set, then any subset T of S is partially ordered by the relation $\geq$ of S restricted to T.

Other interesting examples of partial orderings arise in discussing divisibility in monoids and rings. For example, in the multiplicative monoid of positive integers, we can define $a \geq b$ to mean $a \mid b$ (a is a divisor of b). Then PO1PO3 hold. More generally, let S be a commutative monoid satisfying the cancellation law. We say that S is *reduced* if 1 is the only invertible element in S. In this case, $a \mid b$ and $b \mid a$ imply $a = b$. Then S is partially ordered if we define $a \geq b$ by $a \mid b$. If S is not reduced, we obtain a non-trivial congruence relation in S by defining $a \sim b$ if $a = bu$, u invertible. The quotient monoid $\overline{S}$ relative to this congruence relation is reduced and can be partially ordered by the divisibility relation.

In a finite partially ordered set, the relation $>$ can be expressed in terms of a relation of covering. We say that a_1 is a *cover* of a_2 if $a_1 > a_2$ and there exists no u such that $a_1 > u > a_2$. It is clear that

$a > b$ in a finite partially ordered set if and only if there exists a sequence $a = a_1, a_2, \ldots, a_n = b$ such that each a_i is a cover of a_{i+1}. The notion of cover suggests a way of representing a finite partially ordered set S by a diagram. We represent the elements of S by dots. If a_1 is a cover of a_2, then we place a_1 above a_2 and connect the two dots by a straight line. Then $a > b$ if and only if there is a descending broken line connecting a to b. If no line connects a and $b \neq a$, then a and b are not comparable, that is, we have neither $a \geq b$ nor $b \geq a$.

An element u of a partially ordered set S is an *upper bound* of a subset A of S if $u \geq a$ for every $a \in A$. The element u is a *least upper bound* or sup of A if u is an upper bound of A and $u \leq v$ for every upper bound v of A. It is clear from PO2 that if $\sup A$ exists, then it is unique. In similar fashion, one defines lower bounds and greatest lower bounds or infs of a set A. Also, if $\inf A$ exists, then it is unique. We now introduce the following

Definition 5.1.2. *A lattice is a partially ordered set in which any two elements have a least upper bound and a greatest lower bound.*

We denote the least upper bound of a and b by $a \vee b$ ("a cup b" or "a union b") and the greatest lower bound by $a \wedge b$ ("a cap b" or "a meet b"). If a, b, c are elements of a lattice L, then $(a \vee b) \vee c \geq a, b, c$ and if $v \geq a, b, c$, then $v \geq (a \vee b), c$ so $v \geq (a \vee b) \vee c$. Hence, $(a \vee b) \vee c$ is a sup of a, b, c. By induction, one shows that any finite set of elements of a lattice have a sup. Similarly, any finite subset has an inf. We denote the sup and inf of $a_1, a_2, \ldots, a_n$ by $a_1 \vee a_2 \vee \cdots \vee a_n$, $a_1 \wedge a_2 \wedge \cdots \wedge a_n$ respectively.

Any totally ordered set is a lattice. For, if a and b are two elements of such a set, we have either $a \geq b$ or $b \geq a$. In the first case, $a \vee b = a$ and $a \wedge b = b$. If $b \geq a$, then $a \vee b = b$ and $a \wedge b = a$.

A partially ordered set is called a *complete lattice* if every subset $A = \{a_\alpha\}$ has a sup and an inf. We denote these by $\bigvee a_\alpha$ and $\bigwedge a_\alpha$, respectively. If the set $\{a_\alpha\}$ coincides with the underlying set of the lattice L, then $0 \equiv \bigwedge a_\alpha$ is the least element of L and $1 \equiv \bigvee a_\alpha$ is the greatest element of L: $0 \leq a$ and $1 \geq a$ for every $a \in L$. The following is a very useful criterion for recognizing that a given partially ordered set is complete lattice.

Theorem 5.1.3. *A partially ordered set with a greatest element 1 such that every non-vacuous subset $\{a_\alpha\}$ has a greatest lower bound is a complete lattice. Dually, a partially ordered set with a least element 0 such that every non-vacuous subset has a least upper bound is a complete lattice.*

Proof. Assuming the first set of hypotheses, we have to show that any $A = \{a_\alpha\}$ has a sup. Since $1 \geq a_\alpha$, the set B of upper bounds of A is non-vacuous. Let $b = \inf B$. Then it is clear that $b = \sup A$. The second statement follows by symmetry. □

5.2 Baics Examples

1. For any set S, $\mathscr{P}(S)$ is a complete lattice. Here $1 = S$ and $0 = \emptyset$.

2. The set of subgroups of a group G ordered by inclusion. Since G is a subgroup and the intersection of any set of subgroups is a subgroup, the set of subgroups is a complete lattice. The proof of Theorem 8.3 shows that the sup of a set of subgroups is the intersection of all subgroups containing the given set $\{H_\alpha\}$. Clearly, this is the subgroup generated by all the H_α.

   The next four examples are similar to 2. They are complete lattices in which $\geq$ means inclusion.

3. The set of normal subgroups of a group. The sup of a set of normal subgroups is the subgroup they generate.

4. The set of subspaces of a vector space ordered by inclusion. The inf is the set intersection and the sup is the subspace spanned by the given set of subspaces.

5. The set of ideals of a ring R. Inf is the set intersection, sup is the ideal generated. For two ideals I_1, I_2, this is $I_1 + I_2$, the set of sums $b_1 + b_2$, $b_i \in I_i$.

6. The set of left (right) ideals of a ring.

7. The set of positive integers partially ordered by divisibility: $a \geq b \iff a \mid b$. Here, $a \vee b$ is the greatest common divisor of a

and b and $a \wedge b$ is the least common multiple of a and b. This is a lattice but it is not complete.

8. All the diagrams above except the last one represent lattices (necessarily complete since they are finite).

9. The set $\mathbb{Q}$ of rational numbers with $a \geq b$ having the usual significance. This is totally ordered and hence, as we noted above, $\mathbb{Q}$ is a lattice. However, $\mathbb{Q}$ is not complete.

10. Even the subset of $\mathbb{Q}$ of rationals between 0 and 1 is not complete. On the other hand, the real interval $[0, 1]$ (with the usual order) is a complete lattice.

It is useful to sort out the basic properties of the binary compositions $a \wedge b$ and $a \vee b$ in a lattice L. This will lead us to an alternative definition of a lattice in terms of conditions on two binary compositions on a set. We note first that it follows from the definitions that $a \vee b$ and $a \wedge b$ are symmetric in the two arguments. Hence we have the commutative laws: $a \vee b = b \vee a \quad \text{and} \quad a \wedge b = b \wedge a$. Also, we saw that $(a \vee b) \vee c$ is the sup of a, b, and c. Since the sup is a symmetric function of a, b, and c, it follows that: $(a \vee b) \vee c = a \vee (b \vee c)$, and similarly, $(a \wedge b) \wedge c = a \wedge (b \wedge c)$. It is clear that every a is idempotent relative to $\vee$ and $\wedge$: $a \vee a = a \quad \text{and} \quad a \wedge a = a$. Also, it is clear that if $a \geq b$, then $a \vee b = a$ and $a \wedge b = b$. Hence, for any a and b we have: $(a \vee b) \wedge a = a \quad \text{and}$
$(a \wedge b) \vee a = a$.

Conversely, let L be any set in which there are defined two binary compositions $\vee$ and $\wedge$ satisfying the conditions we have noted:

L1. $a \vee b = b \vee a, \quad a \wedge b = b \wedge a.$

L2. $(a \vee b) \vee c = a \vee (b \vee c), \quad (a \wedge b) \wedge c = a \wedge (b \wedge c).$

L3. $a \vee a = a, \quad a \wedge a = a.$

L4. $(a \vee b) \wedge a = a, \quad (a \wedge b) \vee a = a$

We shall show that L is a lattice relative to a suitable definition of $\geq$ and that $a \vee b$ and $a \wedge b$ are the sup and inf of a and b in this lattice.

Before proceeding to the proof, we remark that we have made precisely the same assumptions on the two compositions $\vee$ and $\wedge$. Hence, we have the important *principle ofduality* that states that, if S is a statement which can be deduced from our axioms, then the dual statement S', obtained by interchanging $\vee$ and $\wedge$ throughout S can also be deduced. We note next that, if $a, b \in L$ (satisfying L1-L4), then the conditions $a \vee b = a$ and $a \wedge b = b$ are equivalent. We shall now define a relation $\geq$ in L by specifying that $a \geq b$ means that $a \vee b = a$, hence $a \wedge b = b$. Evidently, in dualizing, a statement $a \geq b$ has to be replaced by $b \geq a$.

We shall now verify that the $\geq$ we have introduced satisfies PO1-PO3. Since $a \vee a = a$ we have $a \geq a$, so PO1 holds. If $a \geq b$ and $b \geq a$, then we have $a \vee b = a$ and $b \vee a = b$. Since $a \vee b = b \vee a$, this gives $a = b$, which proves PO2. Next, assume that $a \geq b$ and $b \geq c$. Then $a \vee b = a$

and $b \vee c = b$. Hence

$$a \vee c = (a \vee b) \vee c = a \vee (b \vee c) = a \vee b = a,$$

which means that $a \geq c$. Hence PO3 is valid. Since $(a \vee b) \wedge a = a$, by L4, $a \vee b \geq a$. Similarly, $a \vee b \geq b$. Now let c be an element such that $c \geq a$ and $c \geq b$. Then $a \vee c = c$ and $b \vee c = c$. Hence

$$(a \vee b) \vee c = a \vee (b \vee c) = a \vee c = c$$

so $c \geq a \vee b$. Thus $a \vee b$ is a sup of a and b in L. By duality, $a \wedge b$ is an inf of a and b. This completes the verification that a set L with binary compositions satisfying L1-L4 is a lattice and $a \vee b$ and $a \wedge b$ are the sup and inf in this lattice. A subset M of a lattice L is called a *sublattice* if it is closed under the compositions $\vee$ and $\wedge$. It is evident that a sublattice is a lattice relative to the induced compositions. On the other hand, a subset of a lattice may be a lattice relative to the partial ordering $\geq$ defined in L without being a sublattice. For example, the lattice of subgroups of a group G is not a sublattice of the set $\mathscr{P}(G)$ since $H_1 \cup H_2$ is generally not a subgroup.

If a is a fixed element of a lattice L, then the subset of elements x such that $x \geq a$ ($x \leq a$) is evidently a sublattice. If $a \leq b$, the subset of elements $x \in L$ such that $a \leq x \leq b$ is a sublattice. We call such a sublattice an interval and we denote it as $I[a, b]$.

The definition of a lattice by means of the axioms L1-L4 makes it natural to define a *homomorphism* of a lattice L into a lattice L' to be a map $a \to a'$ such that: $(a \vee b)' = a' \vee b'$ and $(a \wedge b)' = a' \wedge b'$. In this case, if $a \geq b$, then we have $a \vee b = a$; hence $a' \vee b' = a'$ and $a' \geq b'$. A map between partially ordered sets having this property is called *order preserving*. Thus, we have shown that a lattice homomorphism is order preserving. However, the converse need not hold. A bijective homomorphism of lattices is called an *isomorphism*. These can be characterized by order preserving properties, as we see in the following

Theorem 5.2.1. *A bijective map of a lattice L onto a lattice L' is a lattice isomorphism if and only if it and its inverse are order preserving.*

Proof. We have seen that if $a \to a'$ is a lattice isomorphism, then this map is order preserving. It is clear also that the inverse map is an isomorphism of L' into L, so it is order preserving. Conversely, suppose $a \to a'$ is bijective and it and its inverse are order preserving. This means that $a \geq b$ in L if and only if $a' \geq b'$ in L'. Let $d = a \vee b$. Then $d \geq a, b$, so $d' \geq a', b'$. Let $e' \geq a', b'$ and let e be the inverse image of e'. Then $e \geq a, b$. Hence $e \geq d$ and $e' \geq d'$. Thus we have shown that $d' = a' \vee b'$. In a similar fashion, we can show that $(a \wedge b)' = a' \wedge b'$. □

5.2.1 Distributivity and Modularity

One of the compositions of a lattice may be viewed as the analogue of addition in a ring, and the other can be taken as the analogue of multiplication. Depending on which we use for addition and which for multiplication, we can formulate the following two distributive laws:

$$D \qquad\qquad a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

and its dual

$$D' \qquad\qquad a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c).$$

It is a bit surprising thatas we shall now showthese two conditions are equivalent. Suppose D holds. Then

$$\begin{aligned} (a \vee b) \wedge (a \vee c) &= ((a \vee b) \wedge a) \vee ((a \vee b) \wedge c) \\ &= a \vee ((a \vee b) \wedge c) \\ &= a \vee ((a \wedge c) \vee (b \wedge c)) \\ &= (a \vee (a \wedge c)) \vee (b \wedge c) \\ &= a \vee (b \wedge c), \end{aligned}$$

which is D'. Dually D' implies D. A lattice in which these distributive laws hold is called *distributive.* There are some important examples of this. First, as we showed in the Introduction (p. 4), the lattice $\mathscr{P}(S)$ of subsets of a set S is distributive. Second, we have the following Any totally ordered set is a distributive lattice.

Proof. We wish to establish D for any three elements a, b, and c and we distinguish two cases:

1. $a \geq b$, $a \geq c$

2. $a \leq b$ or $a \leq c$

In (1), we have $a \wedge (b \vee c) = b \vee c \quad \text{and} \quad (a \wedge b) \vee (a \wedge c) = b \vee c$. In (2), we have $a \wedge (b \vee c) = a \quad \text{and} \quad (a \wedge b) \vee (a \wedge c) = a$. Hence, in both cases (D) holds. □

This lemma can be used to show that the set of positive integers ordered by divisibility is a distributive lattice. In this example,
$a \vee b = (a, b)$ the g.c.d. of a and b and $a \wedge b = [a, b]$ the l.c.m. of a and b. Also, if we write $a = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$, $b = p_1^{b_1} p_2^{b_2} \cdots p_k^{b_k}$ where the p_i are distinct primes and the a_i and b_i are non-negative integers, then $(a, b) = \prod p_i^{\min(a_i, b_i)}, \quad [a, b] = \prod p_i^{\max(a_i, b_i)}$. Hence, if $c = p_1^{c_1} p_2^{c_2} \cdots p_k^{c_k}$, c_i non-negative integral, then

$$[a, (b, c)] = \prod p_i^{\max(a_i, \min(b_i, c_i))},$$

and

$$([a, b], [a, c]) = \prod p_i^{\min(\max(a_i, b_i), \max(a_i, c_i))}.$$

Now the set of non-negative integers with the natural order is totally ordered, and $\max(a_i, b_i) = a_i \vee b_i, \quad \min(a_i, b_i) = a_i \wedge b_i$ in this lattice. Hence, the distributive law D' in this lattice gives the relation

$$\max(a_i, \min(b_i, c_i)) = \min(\max(a_i, b_i), \max(a_i, c_i)).$$

Then we have

$$[a, (b, c)] = ([a, b], [a, c])$$

which is D for the lattice of positive integers ordered by divisibility.

The same reasoning applies to any reduced factorial monoid.

Another remark on distributivity which is worth noting is that in any lattice we have $a \wedge (b \vee c) \geq (a \wedge b)$ and $a \wedge (b \vee c) \geq a \wedge c$. Hence

$$a \wedge (b \vee c) \geq (a \wedge b) \vee (a \wedge c).$$

Thus in order to establish distributivity it suffices to establish the reverse inequality

$$(1) \qquad a \wedge (b \vee c) \leq (a \wedge b) \vee (a \wedge c).$$

The most important lattices which occur in algebra (e.g., the lattice of sub-modules of a module, the lattice of normal subgroups of a group) are not distributive. For instance, let $L(V)$ denote the lattice of subspaces of a vector space V over a field F. Assume $\dim V \geq 2$ and let x and y be linearly independent vectors in V. Then
$F(x+y) \cap (Fx+Fy) = F(x+y)$
but $F(x+y) \cap Fx = 0$ and $F(x+y) \cap Fy = 0$
so $F(x+y) \cap (Fx+Fy) \neq (F(x+y) \cap Fx) + (F(x+y) \cap Fy)$. As we shall see in a moment, the lattice $L(V)$ satisfies a weakening of the distributive condition, which was first formulated by Dedekind. This is the condition:

M $\qquad$ If $a \geq b$, then $a \wedge (b \vee c) = b \vee (a \wedge c)$.

Since b = a $\wedge b$, the right-hand side can be replaced by $(a \wedge b) \vee (a \wedge c)$.

Hence, the condition **M** is equivalent to **D** in the special case in which $a \geq b\,(\text{or}\, a \geq c)$.

Condition **M** is called modularity, and a lattice satisfying it is said to be modular. The dual condition M' reads: If a $\leq b$, then $a \vee (b \wedge c) = b \wedge (a \vee c)$.
Clearly, this is the same as **M**. It follows that, as for distributive lattices, the principle of duality is valid in modular lattices.

The importance of modular lattices in algebra stems from the following.

Theorem 5.2.2. *The lattice of normal subgroups of a group is modular. The lattice of submodules of a module is modular.*

Proof. The normal subgroup generated by two normal subgroups H_1 and H_2 of a group G is $H_1H_2 = H_2H_1$. Hence we have to prove that if $H_i, i = 1, 2, 3$, are normal subgroups such that $H_1 \supset H_2$ then

$$H_1 \cap (H_2H_3) = H_2(H_1 \cap H_3).$$

The remark above about the distributive law shows that it is enough to prove that

$$H_1 \cap (H_2H_3) = H_2(H_1 \cap H_3).$$

Suppose $a \in H_1 \cap (H_2H_3)$. Then $a = h_1 = h_2h_3$, $h_i \in H_i$, and $h_3 = h_2^{-1}h_1 \in H_1$, since $H_1 \supset H_2$. Thus $h_3 \in H_1 \cap H_3$ and $a = h_2h_3 \in h_2(H_1 \cap H_3)$. This proves the required inclusion. The argument for modules is similar and simpler so we omit it. □

An alternative definition of modularity which is sometimes useful can be extracted from the following

Theorem 5.2.3. *A lattice L is modular if and only if whenever $a \geq b$ and $a \wedge c = b \wedge c$ and $a \vee c = b \vee c$ for some c in L, then $a = b$.*

Proof. Let L be modular and let a, b, c be elements of L such that $a \geq b$, $a \vee c = b \vee c$, and $a \wedge c = b \wedge c$. Then

$$a = a \wedge (a \vee c) = a \wedge (b \vee c) = b \vee (a \wedge c) = b \vee (b \wedge c) = b.$$

Conversely, suppose that L is any lattice satisfying the condition stated in the theorem. Let $a, b, c \in L$ and $a \geq b$. We know that $a \wedge (b \vee c) \geq b \vee (a \wedge c)$. Also,

$$(a \wedge (b \vee c)) \wedge c = a \wedge ((b \vee c) \wedge c) = a \wedge c.$$

and

$$a \wedge c = (a \wedge c) \wedge c \leq (b \vee (a \wedge c)) \wedge c \leq a \wedge c.$$

Hence

$$(b \vee (a \wedge c)) \wedge c = a \wedge c.$$

since $b \leq a$ the dual of our first relation is

$$(b \vee (a \wedge c)) \vee c = b \vee c.$$

and the dual of the second one is

$$(a \wedge (b \vee c)) \vee c = b \vee c.$$

Thus, we have

$$(a \wedge (b \vee c)) \wedge c = (b \vee (a \wedge c)) \wedge C,$$

$$(a \wedge (b \vee c)) \wedge c = (b \vee (a \wedge c)) \wedge C.$$

Hence the assumed property implies that $a \wedge (b \vee c) = b \vee (a \wedge c)$, which is the modular axiom. □

We shall prove next an analogue for modular lattices of the second isomorphism theorem for groups, namely,

If a and b are elements of a modular lattice, then the map $x \to x \wedge b$ is an isomorphism of the interval $I[a, a \vee b]$ onto $I[a \wedge b, b]$. The inverse isomorphism is $y \to y \vee a$.

Proof. We note first that in any lattice the maps $x \to x \vee a$ and $x \to x \wedge a$ are order preserving. For, we have $x \geq y$ if and only if $x \vee y = x$ and if and only if $x \wedge y = y$. Then $x \vee y = x$ implies $(x \vee a) \vee (y \vee a) = (x \vee y) \vee (a \vee a) = (x \vee y) \vee a = x \vee a$. Hence, $x \geq y$ implies $x \vee a \geq y \vee a$. Similarly, we have $x \wedge a \geq y \wedge a$. Now, if $a \leq x \leq a \vee b$, then $a \wedge b \leq x \wedge b \leq b = (a \vee b) \wedge b$, and if $a \wedge b \leq y \leq b$, then $a = a \vee (a \wedge b) \leq y \vee a \leq a \vee b)$. Hence, $x \to x \wedge b$ and $y \to y \vee a$ map $I[a, a \vee b]$ into $I[a \wedge b, b]$ and $I[a \wedge b, b]$ into $I[a, a \vee b]$, respectively. Since these maps are order preserving, the theorem will follow from above theorem if we can show that they are inverses. Let $x \in I[a, a \vee b]$. Then, since $x \geq a$, by modularity

$$(x \wedge b) \vee a = x \vee (a \vee b),$$

and since $x \leq a \vee b$, this gives $(x \wedge b) \vee a = x$. Dually, we have that if $y \in I[a \wedge b, b]$, then $(y \vee a) \wedge b = y$. This proves the two maps are inverses. □

This theorem leads us to introduce a notion of equivalence for intervals, which in modular lattices is stronger than isomorphism. First, we define the intervals $I[u, v]$ and $I[w, t]$ to be *transposes* if there exist a and b in the lattice such that one of these coincides with $I[a, a \vee b]$ and the other with $I[a \wedge b, b]$. The intervals $I[u, v]$ and $I[w, t]$ are *projective* if there exists a finite sequence

$$I[u, v] = I[u_1, v_1], I[u_2, v_2], \ldots, I[u_n, v_n] = I[w, t]$$

such that consecutive pairs $I[u_k, v_k], I[u_{k+1}, v_{k+1}]$ are transposes. It is immediate that this is an equivalence relation. Also it is clear from Theorem 5.8 that in a modular lattice projective intervals are isomorphic.

5.2.2 The Theorem of Jordan-Hölder-Dedekind

A partially ordered set S is said to be of *finite length* if the lengths (number of distinct terms) of its chains (= totally ordered subsets) are bounded. If a and b are elements of a partially ordered set of finite length and $a > b$, then we can find a finite sequence of elements $a = a_1, a_2, \ldots, a_n = b$ such that each a_i is a cover of a_{i+1}. A sequence of elements having this property is called a *connected chain from a to b.* A desirable property is that any two connected chains from a to b ($a > b$) have the same length. We shall now show that this property is assured for a lattice L of finite length if L is *semi-modular*, in the sense that if $a \vee b$ covers a and b, then a and b cover $a \wedge b$. We have seen that if L is modular, then $I[a \wedge b, a]$ and $I[b, a \vee b]$ are isomorphic. Hence, it is clear that modularity implies semi-modularity. The following theorem is the lattice analogue of the Jordan-Hlder theorem for finite groups.

Theorem 5.2.4. ***THEOREM OF JORDAN-HÖLDER-DEDEKIND.*** *Let L be a semi modular lattice of finite length. Then any two connected chains from a to b, $a > b$, have the same length. Moreover, if L is modular and*

$$a = a_1 > a_2 > \cdots > a_{m+1} = b \tag{2}$$

$$a = a'_1 > a'_2 > \cdots > a'_{m+1} = b \tag{3}$$

are two connected chains from a to b, then the corresponding intervals $I[a_{i+1}, a_i]$ and $I[a'_{j+1}, a'_j]$ can be paired so that the paired ones are projective.

Proof. The proof imitates the proof of the group result. We use induction on n where $n+1$ is the length of one of the connected chains from a to b. If $n=1$, then a is a cover of b and the result is clear. If $a_2 = a_2'$, then we have two connected chains from a_2 to b and the theorem follows by induction on n. Now suppose $a_2 \neq a_2'$. Then a_1 is a cover of a_2 and of $a_2' \neq a_2$, which implies that $a_2 \vee a_2' = a_1$. Then the semi-modularity implies that a_2 and a_2' are covers of $a_3'' \equiv a_2 \wedge a_2'$. Also, $a_3'' \geq b$. If $b = a_3''$.

In this case, $m = n = 2$ and, in the modular case, $I[a_2, a_1]$ and $I[b, a_2']$, and $I[a_2', a_1]$ and $I[b, a_2]$ are transposes. If $a_3'' > b$, then we can find a connected chain $a_3'', a_4'', \ldots, a_{q+1}'' = b$. Then the result follows by induction on n applied to $a_2, a_3, \ldots, a_{n+1} = b$, and $a_2, a_3'', \ldots, a_{q+1}'' = b$ as well as to $a_2', a_3'', \ldots, a_{q+1}'' = b$ (using $q = n$) and $a_2', a_3', \ldots, a_{m+1}' = b$. Also, in the modular case we have to use the fact that $I[a_2, a_1]$ and $I[a_3'', a_2']$ and $I[a_2'', a_1]$ and $I[a_3'', a_2]$ are transposes as in the proof of the group result. The remaining details are left to the reader. □

Assume now that L is modular with a least element 0, and that L is of finite length. If we have a connected chain $a_1 = a, a_2, \ldots, a_{n+1} = b$ from a to b, then we shall call the number n (uniquely determined by a and b) the length of the interval $I[b, a]$. We denote the length of $I[0, a]$ as $d(a)$ and call this the *rank* of a. If $a \geq b$, then it is clear that

$$d(a) = d(b) + \operatorname{length} I[b, a].$$

Hence for any a and b in L we have

$$d(a \vee b) = d(a) + \text{length}\, I[a, a \vee b]$$
$$d(b) = d(a \wedge b) + \text{length}\, I[a \wedge b, b]$$

Since $I[a, a \vee b]$ and $I[a \wedge b, b]$ are isomorphic, they have the same lengths. Hence,

$$d(a \vee b) - d(a) = d(b) - d(a \wedge b)$$

or

$$d(a \vee b) = d(a) + d(b) - d(a \wedge b) \tag{4}$$

which is analogous to the dimensionality formula for the subspaces of a finite dimensional vector space.

5.3 Boolean Algebras

Definition 5.3.1. *A Boolean algebra is a lattice with a greatest element 1 and least element 0 which is distributive and complemented.*

The most important instances of Boolean algebras are the lattices of subsets of any set S. More generally, any *field of subsets* of S, that is, a collection of subsets of S which is closed under union and intersection, contains S and $\emptyset$, and the complement of any set in the collection is a Boolean algebra. The following theorem gives the most important elementary properties of complements in a Boolean algebra.

Theorem 5.3.2. *The complement a' of any element a of a Boolean algebra B is uniquely determined. The map $a \to a'$ is an anti-automorphism of period ≤ 2: $a \to a'$ satisfies*

$$(11) \qquad (a \vee b)' = a' \wedge b', \qquad (a \wedge b)' = a' \vee b',$$

$$(12) \qquad a'' = a.$$

Proof. Let $a \in B$ and let a' and $a - 1$ satisfy $a \vee a' = 1$, $a \wedge a_1 = 0$. Then

$$a_1 = a_1 \wedge 1 = a_1 \wedge (a \vee a') = (a_1 \wedge a) \vee (a_1 \wedge a') = a_1 \wedge a'.$$

Hence, if in addition, $a \vee a_1 = 1$ and $a \wedge a' = 0$, then $a' = a' \wedge a_1$, and so $a' = a_1$. This proves the uniqueness of the complement. It is clear that a is the complement of a'. Hence $a'' \equiv (a')' = a$, and $a \to a'$ is of period one or two; hence bijective. Now, let $a \leq b$. Then $a \wedge b' \leq b \wedge b' = 0$, so

$$b' = b' \wedge 1 = b' \wedge (a \vee a') = (b' \wedge a) \vee (b' \wedge a') = b' \wedge a'.$$

Hence $b' \leq a'$. Since $a \to a'$ is its own inverse and is order inverting, it follows from Theorem 5.1.4 that $a \to a'$ is a lattice anti-isomorphism.

□

Historically, Boolean algebras were the first lattices to be studied. They were introduced by Boole to formalize the calculus of propositions. For a long time it was supposed that the type of algebra represented by these systems was of a different character from that involved in number systems and their generalizations (algebras in the technical sense and rings). However, it was discovered rather late in the day by M. H. Stone that this is not the case. In fact, any Boolean algebra, if properly viewed, becomes a special type of ring. In order to make a ring out of a Boolean algebra B, we introduce the new composition

$$a + b = (a \wedge b') \vee (a' \wedge b),$$

which is called the *symmetric difference* of a and b.

We have

$$\begin{aligned} (a \vee b) \wedge (a \wedge b)' &= (a \vee b) \wedge (a' \vee b') \\ &= ((a \vee b) \wedge a') \vee ((a \vee b) \wedge b') \\ &= ((a \wedge a') \vee (b \wedge a')) \vee ((a \wedge b') \vee (b \wedge b')) \\ &= (b \wedge a') \vee (a \wedge b') \\ &= a + b. \end{aligned} \tag{13}$$

The first formula shows that in the Boolean algebra of subsets of a set, $U + V$ is the set of elements contained in U or in V, but not in both:

We shall now show that B is a ring with $+$ as just defined, the product $ab = a \wedge b$, and 1 as the unit of B.

Evidently $+$ is commutative. To prove associativity we note first that, by (13),

$$(a + b)' = (a \vee b)' \vee (a \wedge b) = (a \wedge b) \vee (a' \wedge b')$$

Hence,

$$(a+b)+c = [((a \wedge b') \vee (a' \wedge b)) \wedge c'] \vee [((a \wedge b) \vee (a' \wedge b')) \wedge c]$$

$$= [(a \wedge b' \wedge c') \vee (a' \wedge b \wedge c')] \vee [(a \wedge b \wedge c) \vee (a' \wedge b' \wedge c)]$$

$$= (a \wedge b' \wedge c') \vee (a' \wedge b \wedge c') \vee (a \wedge b \wedge c) \vee (a' \wedge b' \wedge c)$$

This is symmetric in a, b, and c. In particular, $(a+b)+c = (c+b)+a$. Commutativity therefore, implies the associative law for $+$. Evidently,

$$a + 0 = (a \wedge 1) \vee (a' \wedge 0) = a$$

and

$$a + a = (a \wedge a') \vee (a' \wedge a) = 0.$$

Hence $(B, +, 0)$ is a commutative group.

We know that $\cdot\ (= \wedge)$ is associative and commutative. Also, $a \cdot 1 = 1 \cdot a = a \wedge 1 = a$ for all a in B. It remains to check one of the distributive laws. Now we have

$$\begin{aligned}(a+b)c &= (a \wedge b') \vee (a' \wedge b) \wedge c \\ &= (a \wedge b' \wedge c) \vee (a' \wedge b \wedge c) \\ ac+bc &= ((a \wedge c) \wedge (b \wedge c)') \vee ((a \wedge c)' \wedge (b \wedge c)) \\ &= ((a \wedge c) \wedge (b' \wedge c')) \vee ((a' \vee c') \wedge (b \wedge c)) \\ &= (a \wedge c \wedge b') \vee (a' \wedge b \wedge c).\end{aligned}$$

Comparison shows that $(a+b)c = ac+bc$. Hence $(B, +, \cdot, 0, 1)$ is a ring.

We have noted also that the ring B is commutative and every element is of order ≤ 2 in the additive group. Also every element is idempotent: $a^2 = a \wedge a = a$. These properties of a ring are not independent; for, as we now note, if every element of a ring is idempotent, then the ring is commutative and $2a = 0$ for every a. To prove this we observe that

$$a + b + ab + ba = a^2 + b^2 + ab + ba = (a+b)^2 = a + b.$$

Hence, $ab + ba = 0$. Then $2a = 2a^2 = aa + aa = 0$, and so $a = -a$. Then $ab = -ba = ba$. These considerations lead us to introduce the following

Definition 5.3.3. *A ring called Boolean if all of its elements are idempotent.*

We have seen that such a ring is of characteristic two. We shall prove next that any Boolean ring B defines a Boolean algebra, and that, in fact, these two concepts are equivalent. Suppose $(B, +, \cdot, 0, 1)$ is a Boolean ring. In order to reverse the process we used to go from a Boolean algebra to a Boolean ring, we now define

$$a \vee b = a + b - ab = 1 - (1 - a)(1 - b).$$

The second expression for $a \vee b$ shows that if we introduce the map $\sigma : x \to 1 - x$ in B, then $a \vee b = \sigma^{-1}(\sigma(a)\sigma(b))$, since $\sigma^2 = 1$. It is clear from this and the associative law of multiplication in B that $\vee$ is associative and, of course, this composition is commutative. Also, $a \vee a = 2a - a^2 = -a^2 = a$. We now define $a \wedge b = ab$. Then associativity and commutativity are clear, and $a \wedge a = a$ since every element of B is idempotent. Also we have $(a \vee b) \wedge a = (a + b - ab)a = a$ and $(a \wedge b) \vee a = ab + a - a^2b = a$. Thus the defining conditions $L1L4$ on $\vee$ and $\wedge$ for a lattice hold. It is immediate that the ring 1 and 0 are the greatest and least elements of the lattice $(B, \vee, \wedge)$, and that $1 - a$ is a complement of a, since $a \vee (1 - a) = 1$ and $a \wedge (1 - a) = 0$. The lattice is distributive since

$$\begin{aligned} (a \vee b) \wedge c &= (a + b - ab)c \\ &= ac + bc - abc \\ &= ac + bc - acbc \\ &= (a \wedge c) \vee (b \wedge c). \end{aligned}$$

Thus, $(B, \vee, \wedge, 0, 1,')$ is a Boolean algebra. It remains to show that the process of passing from a Boolean algebra to a ring and the process of passing from a ring to a Boolean algebra are inverses. Thus suppose we begin with a Boolean algebra $(B, \vee, \wedge, 0, 1, ')$. Then we obtain the ring $(B, +, \cdot, 0, 1)$ in which $a + b = (a \wedge b') \vee (a' \wedge b)$ and $ab = a \wedge b$. An application of the second process to this ring gives a Boolean algebra in which $1 = 1$, $0 = 0$, $a' = 1 - a$, and the new $\vee$ and $\wedge$ which we now denote as $\bar{\vee}$ and $\bar{\wedge}$ respectively are $a \bar{\vee} b = a + b - ab = 1 - (1 - a)(1 - b) = (a' \wedge b')' = a \vee b$ and $a \bar{\wedge} b = ab = a \wedge b$. Hence, $\bar{\vee} = \vee$, $\bar{\wedge} = \wedge$, and so we obtain the original Boolean algebra. On the other hand, suppose we start with a Boolean ring $(B, +, \cdot, 0, 1)$ and we obtain the Boolean algebra $(B, \vee, \wedge, 0, 1, ')$ in which $a \vee b = a + b - ab$, $a \wedge b = ab$, $0 = 0$, $1 = 1$, and $a' = 1 - a$. Then, applying the process we gave yields a ring in which the new addition $\oplus$ and multiplication $\odot$ are

$$\begin{aligned}
a \oplus b &= (a \wedge (1 - b)) \vee ((1 - a) \wedge b) \\
&= a(1 - b) \vee (1 - a)b \\
&= (a - ab) \vee (b - ab) \\
&= a - ab + b - ab - (a - ab)(b - ab) \\
&= a - ab + b - ab - ab + ab + ab - ab \\
&= a + b \\
a \odot b &= a \wedge b = ab.
\end{aligned}$$

Also $1 = 1$, $0 = 0$ so we obtain the original ring. We have therefore proved the following theorem, which is due to Stone.

Theorem 5.3.4. *Boolean algebra and Boolean ring.*

There is one more remark worth making. In passing from a Boolean algebra to a Boolean ring we could have used $\vee$ for $\wedge$, $\wedge$ for $\vee$, 1 for 0, and 0 for 1 in the construction. This follows from the principle of duality, which is applicable to Boolean algebras. Our process then leads to a ring B' with the same underlying set B and with the addition

$$a +' b = (a \vee b') \wedge (a' \vee b)$$

and multiplication

$$a \cdot' b = a \vee b.$$

Also the new 0 and 1 are $0' = 1$, $1' = 0$. In terms of the ring B, we have

$$\begin{aligned} a +' b &= (a + (1 - b) - a + ab)(b + (1 - a) - b + ab) \\ &= (1 - b + ab)(1 - a + ab) \\ &= 1 - (a + b), \\ a \cdot b &= a + b - ab. \end{aligned}$$

We define an *ideal* of a Boolean algebra B to be an ideal of the associated Boolean ring $(B, +, \cdot, 0, 1)$. The conditions for a subset I to be an ideal are:

(1) if $u, v \in I$, then $u + v \in I$, and

(2) if a is arbitrary in B, then $ua \in I$.

Since $ua = u \wedge a$ and $ua = a$ if and only if $a \leq u$, the second condition is equivalent to: if $u \in I$, then $b \in I$ for every $b \leq u$. Since $u \vee v = u + v + uv$, $u \vee v \in I$ for every $u, v \in I$.Conversely, let I be a subset of B such that if $u, v \in I$, then $u \vee v \in I$ and if $u \in I$, then every $b \leq u$ is in I. Then $u \wedge v'$ and $v \wedge u' \in I$ (u' and v' the complements of u and v). Hence, $u + v = (u \wedge v') \vee (v \wedge u') \in I$ and so I is an ideal. Thus a subset I of a Boolean algebra is an ideal if and only if it is closed under $\vee$ and contains every $b \leq u$ for any $u \in I$.

An ideal I is called *proper* if $I \neq B$. It is clear that I is proper if and only if $1 \notin I$. If $u \in B$, then $(u) = \{x \in B \mid x \leq u\}$ is an ideal called the *principal ideal* generated by u. An ideal I is *maximal* if I is proper and there is no proper ideal $\bar{I}$ properly containing $I (\bar{I} \supsetneq I)$. We now observe that an ideal I is maximal if and only if I is proper and for every $a \in B$ either a or $a' \in I$. First, suppose I is maximal and let $a \notin I$. Consider the set $\bar{I}$ of elements of the form $u + b$ where $u \in I$ and $b \leq a$. This is an ideal properly containing I, so, by the maximality of I, it coincides with B. Thus, $1 = b + u$ where $b \leq a$ and $u \in I$. Hence, $b' = 1 + b = u \in I$. Since $a' \leq b'$, it is also true that $a' \in I$. Conversely, let I be a proper ideal such that for every $a \in B$, either a or $a' \in I$. Let I be any ideal properly containing I and let $a \in \bar{I}, \notin I$. Then $a' \in \ I$, and so $a' \in \bar{I}$ and $1 = a + a' \in \bar{I}$. Thus $\bar{I} = B$ and I is maximal.

All of this can be dualized by applying the same considerations to the second ring $B' = (B, +', \cdot', 0', 1')$ associated with the Boolean algebra B. Accordingly, we define a *filter (dual ideal)* of B to be an ideal of B'. The foregoing results can be dualized as follows. First, we note that the dual of our criterion for a subset to be an ideal is that a subset F of a Boolean algebra B is a filter if and only if it is closed under $\wedge$ and containing every $b \geq u$ for any $u \in F$. Since $(a \wedge b)' = a' \wedge b'$ and $(a \vee b)' = a' \vee b'$, it is clear that F is a filter if and only if the set F' of complements $a', a \in F$, is an ideal. Condition (1) is equivalent to the finite intersection property: F is closed under finite intersections. A filter is *proper* in the sense that $F \neq B$ if and only if $0 \notin F$. A maximal ideal of B' is called an *ultrafilter* of the Boolean algebra B. A filter F is an ultrafilter if and only if (1) $0 \notin F$, (2) for any $a \in B$, either a or $a' \in F$. If $a \in B$, the subset of elements $x \geq a$ is a filter called the *principal filter* generated by a.

We conclude our brief introduction to Boolean algebras by giving a couple of examples of filters.

Examples:

1. Let $\mathbb{R}$ be the real line endowed with its usual topology and let S denote the collection of non-vacuous open subsets of $\mathbb{R}$. This has the finite intersection property. The set $\bar{S}$ of subsets which contain open subsets of $\mathbb{R}$ is a filter.

2. Let S be any set, $B = \mathscr{P}(S)$ the set of subsets of S. Let I be the set of finite subsets of S. This is an ideal in B; hence the set F of complements of the finite subsets is a filter.