CHAPTER IV

RING THEORY

1. Ring Theory

A *ring* is a set A with two binary operations satisfying the rules given below. Usually one binary operation is denoted '+' and called "addition," and the other is denoted by juxtaposition and is called "multiplication." The rules required of these operations are:

1) A is an abelian group under the operation + (identity denoted 0 and inverse of x denoted $-x$);
2) A is a monoid under the operation of multiplication (i.e., multiplication is associative and there is a two-sided identity usually denoted 1);
3) the distributive laws
$$(x+y)z = xy + xz$$
$$x(y+z) = xy + xz$$

hold for all x, y, and $z \in A$.

Sometimes one does not require that a ring have a multiplicative identity. The word ring may also be used for a system satisfying just conditions (1) and (3) (i.e., where the associative law for multiplication may fail and for which there is no multiplicative identity.) Lie rings are examples of non-associative rings without identities. Almost all interesting associative rings do have identities.

If $1 = 0$, then the ring consists of one element 0; otherwise $1 \neq 0$. In many theorems, it is necessary to specify that rings under consideration are not trivial, i.e. that $1 \neq 0$, but often that hypothesis will not be stated explicitly.

If the multiplicative operation is commutative, we call the ring commutative. *Commutative Algebra* is the study of commutative rings and related structures. It is closely related to algebraic number theory and algebraic geometry.

If A is a ring, an element $x \in A$ is called a *unit* if it has a two-sided inverse y, i.e. $xy = yx = 1$. Clearly the inverse of a unit is also a unit, and it is not hard to see that the product of two units is a unit. Thus, the set $U(A)$ of all units in A is a group under multiplication. ($U(A)$ is also commonly denoted A^*.) If every nonzero element of A is a unit, then A is called a *division ring* (also a skew field.) A commutative division ring is called a field.

Examples:

1) $\mathbf{Z}$ is a commutative ring. $U(\mathbf{Z}) = \{1, -1\}$.
2) The group $\mathbf{Z}/n\mathbf{Z}$ becomes a commutative ring where multiplication is multiplication mod n. $U(\mathbf{Z}/n\mathbf{Z})$ consists of all cosets $i + n\mathbf{Z}$ where i is relatively prime to n.
3) Let F be a field, e.g., $F = \mathbf{R}$ or $\mathbf{C}$. Let $M_n(F)$ denote the set of n by n matrices with entries in F. Add matrices by adding corresponding entries. Multiply matrices by the usual rule for matrix multiplication. The result is a non-commutative ring. $U(M_n(F)) = Gl(n, F) =$ the group of invertible n by n matrices.
4) Let M be any abelian group, and let $\mathrm{End}(M)$ denote the set of endomorphisms of M into itself. For $f, g \in \mathrm{End}(M)$, define addition by $(f+g)(m) = f(m) + g(m)$, and define multiplication as composition of functions. (Note: If M were not abelian we could still define composition because the composition of two endomorphisms is an endomorphism. However, it would not necessarily be true that the sum of two endomorphisms would be an endomorphism. Check this for yourself.)

If A is a ring, a subset B of A is called a *subring* if it is a subgroup under addition, closed under multiplication, and contains the identity. (If A or B does not have an identity, the third requirement would be dropped.)
Examples:

1) $\mathbf{Z}$ does not have any proper subrings.
2) The set of all diagonal matrices is a subring of $M_n(F)$.
3) The set of all n by n matrices which are zero in the last row and the last column is closed under addition and multiplication, and in fact it is a ring in its own right (isomorphic to $M_{n-1}(F)$.) However, it is not a subring since its identity does not agree with the identity of the overring $M_n(F)$.

A function $f : A \to B$ where A and B are rings is called a homomorphism of rings if it is a homomorphism of additive groups, it preserves products: $f(xy) = f(x)f(y)$ for all $x, y \in A$, and finally it preserves the identity: $f(1) = 1$.
Examples: The canonical epimorphism $\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$ is a ring homomorphism. However, the inclusion of $M_{n-1}(F)$ in $M_n(F)$ as suggested in example 3) above is not a ring homomorphism.

A subset $\mathfrak{a}$ is called a *left ideal* of A if it is an additive subgroup and in addition $ax \in \mathfrak{a}$ whenever $a \in A$ and $x \in \mathfrak{a}$. If we require instead that $xa \in \mathfrak{a}$, then $\mathfrak{a}$ is called a *right* ideal. Finally, $\mathfrak{a}$ is called a *two-sided* ideal if it is both a left ideal and a right ideal. Of course, for a commutative ring all these notions are the same.
Examples:

1) It is easy to see that any additive subgroup $n\mathbf{Z}$ is an ideal in $\mathbf{Z}$.
2) Let $A = M_n(F)$. Let L_j be the set of n by n matrices which are zero except possibly in the jth column. It is not hard to see that L_j is a left ideal in A. Can you give an example of a right ideal? There are no two-sided ideals in A. (See the Exercises.)

The following notation is useful. Let $\mathfrak{a}$ and $\mathfrak{b}$ be subsets of A, at least one of which is an additive subgroup of A. Let $\mathfrak{ab}$ denote the set of all sums of elements of the form ab with $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$. It is not hard to see that $\mathfrak{ab}$ is an additive subgroup of A. (Clearly, it is closed under addition. Can you see why it is closed under taking negatives—i.e. additive inverses?) Using this notation, an additive subgroup $\mathfrak{a}$ of A is a left ideal if and only if $A\mathfrak{a} \subseteq \mathfrak{a}$. If $\mathfrak{a}$ is a left ideal of A, and if

$$\mathfrak{a} = Ax_1 + Ax_2 + \cdots + Ax$$

we say that $\mathfrak{a}$ is generated by $x_1, \ldots, x_n$. (Note that all the notation used above has been defined already, at least if we use the shorthand $Ax = A\{x\}$.) If $\mathfrak{a} = Ax$ is generated by a single element, we call it a *principal* left ideal. Similar concepts apply to right ideals. In a commutative ring, of course, we need not distinguish so we just use the terms "principal ideal."

A ring is called a *principal ideal ring* if it is a commutative ring and every ideal is principal. The principal ideal ring which should come immediately to mind is $\mathbf{Z}$. (We saw in an exercise that every subgroup of $\mathbf{Z}$ is of the form $n\mathbf{Z}$ for some non-negative integer n, so the same is true for every ideal. In fact, for $\mathbf{Z}$, every additive subgroup is an ideal.)

We can repeat much of the theory of groups, subgroups, and normal subgroups in the context of rings, subrings, and (two-sided) ideals.

Let $f : A \to B$ be a homomorphism of rings. It is easy to see that the image of the homomorphism is a subring of B. Let $\mathfrak{b}$ be an ideal in B (left, right, or two-sided). Then $f^{-1}(\mathfrak{b}) = \{a \in A | f(a) \in \mathfrak{b}\}$ is an ideal in A (respectively left, right, or two-sided.) For, it is certainly a subgroup, and $a \in A, x \in f^{-1}(\mathfrak{b}) \Rightarrow f(ax) = f(a)f(x) \in \mathfrak{b}$ so $ax \in f^{-1}(\mathfrak{b})$. In particular, $\{0\}$ is a two-sided ideal in B so $\mathrm{Ker} f = f^{-1}(0)$ is a two-sided ideal in A.

Let A be a ring and let $\mathfrak{a}$ be an ideal (always two-sided if not further specified.) Since $\mathfrak{a}$ is a subgroup of A as abelian group, we may construct the factor group $A/\mathfrak{a}$. As usual, it consists of all co-sets $x + \mathfrak{a}$ with $x \in A$. We can in fact make $A/\mathfrak{a}$ into a ring as follows. First note that for x, $y \in A$, the coset $xy + \mathfrak{a}$ depends only on the cosets of x and y. For, $x \equiv x' \bmod \mathfrak{a}$, $y \equiv y' \bmod \mathfrak{a} \Rightarrow x - x'$ and $y - y' \in \mathfrak{a}$. Thus, $xy - x'y' = xy - x'y + x'y - x'y' = (x - x')y + x'(y - y') \in \mathfrak{a}$ since $\mathfrak{a}$ is a two-sided ideal. Hence $xy \equiv x'y' \bmod \mathfrak{a}$ as claimed. It follows that we can define a binary operation on $A/\mathfrak{a}$ by setting

$$(x + \mathfrak{a})(y + \mathfrak{a}) = xy + \mathfrak{a}.$$

We leave it to the student to show that $A/\mathfrak{a}$ becomes a ring with respect to the operations of addition and multiplication inherited from A. (You need to show that the associative law for multiplication holds, $1+\mathfrak{a}$ is a two-sided identity, and the two distributive laws hold. These follow easily from the corresponding facts in A.) $A/\mathfrak{a}$ is called the *factor* ring of A modulo $\mathfrak{a}$. Note that the canonical group epimorphism $A \to A/\mathfrak{a}$ is a ring homomorphism in essence because the product on the right is defined to make it such.

We may now restate the basic isomorphism theorems for groups. To prove them, one need only check that the basic isomorphisms preserve products as well as sums. (That they preserve sums follows because the group theoretic theorems apply already to the additive groups.)

THEOREM. (First Isomorphism Theorem). Let $f : A \to B$ be a homomorphism of groups. Define $\bar{f} : A/\operatorname{Ker} f \to \operatorname{Im} f$ by $\bar{f}\ (a + \operatorname{Ker} f) = f(a)$. Then $\bar{f}$ is a ring isomorphism.

THEOREM. (Second Isomorphism Theorem). Let A' be a subring of A, and let $\mathfrak{a}$ be an ideal in A. Then $A' + \mathfrak{a}$ is a subring of A and $x + A' \cap \mathfrak{a} \rightsquigarrow x + \mathfrak{a}$ defines a ring isomorphism $A'/A' \cap \mathfrak{a} \cong (A' + \mathfrak{a})/\mathfrak{a}$.

THEOREM. (Third Isomorphism Theorem). Let $f : A \to B$ be a ring epimorphism. $\mathfrak{a} \rightsquigarrow f(\mathfrak{a})$ and $\mathfrak{b} \rightsquigarrow f^{-1}(\mathfrak{b})$ provides a one-to-one correspondence between the set of ideals of A containing $\operatorname{Ker} f$ and the ideals of B. In addition, if $\operatorname{Ker} f \subseteq \mathfrak{a}$, then $A/\mathfrak{a} \cong B/f(\mathfrak{a})$.

Note: Applying the Third isomorphism theorem to the canonical epimorphism $A \to A/\mathfrak{b}$ yields for $\mathfrak{a} \subseteq \mathfrak{b}$

$$A/\mathfrak{a} \cong (A/\mathfrak{b})/(\mathfrak{a}/\mathfrak{b}).$$

Exercises.

1. Let $\mathfrak{a}, \mathfrak{b}$, and $\mathfrak{c}$ be additive subgroups of a ring A.
   (a) Prove that $\mathfrak{a}(\mathfrak{b}\mathfrak{c}) = (\mathfrak{a}\mathfrak{b})\mathfrak{c}$.
   (b) Prove $\mathfrak{a}(\mathfrak{b} + \mathfrak{c}) = \mathfrak{a}\mathfrak{b} + \mathfrak{a}\mathfrak{c}$.

2. Fill in the details of the proof of the second isomorphism theorem for rings. In so doing, you may assume the truth of the second isomorphism theorem for groups and that the first isomorphism theorem for rings has been proved. In particular, you may assume that the canonical homomorphism from a ring to the ring modulo a two sided ideal is a ring homomorphism.

3. Let F be a field. Prove that $M_n(F)$ has no nontrivial proper two sided ideals. Hint: Let e_{ij} be the matrix with 1 in the i, j position and zeroes elsewhere. Show that if $a = (a_{ij})$ is an n by n matrix, then $e_{rs}ae_{uv} = a_{su}e_{rv}$.

4. Let A and B be rings. Let $A \times B$ be the direct product of A and B as additive abelian groups and endow it with a multiplicative structure by using componentwise operations. (a) Show that $A \times B$ becomes a ring.
   (b) Let $p : A \times B \to A$ be the projection of the product onto the first factor. Is p a ring homomorphism?
   (c) Let $i : A \to A \times B$ be defined by $i(a) = (a, 0)$. Is i a ring homomorphism?

5. Prove the assertion about $U(\mathbf{Z}/n\mathbf{Z})$ in the text.

2. Maximal Ideals

A left ideal is called *maximal* if it is a proper ideal and it is not contained in any other proper ideals. (Similarly for right ideals or two-sided ideals.) For example, in $\mathbf{Z}$, ideals and subgroups are the same, and since $n\mathbf{Z} \subseteq m\mathbf{Z} \Leftrightarrow m \mid n$, it follows that $n\mathbf{Z}$ is maximal if and only if n is prime.

From the Third Isomorphism Theorem, it follows that a (two-sided) ideal $\mathfrak{a}$ in a ring A is maximal if and only if the factor ring $A/\mathfrak{a}$ has no nontrivial proper (two-sided) ideals. Such a ring is called a *simple* ring. As we shall see in the latter part of this course, the most important examples of simple non-commutative rings are matrix rings $M_n(F)$ over a field F and related rings. Division rings are also clearly simple rings.

In the commutative case, a ring A is simple if and only if it is a field. For, suppose $x \neq 0$ in A. Then Ax is a non-zero ideal and by simplicity, it follows that $Ax = A$. Thus the equation $xy = 1$ has a solution y, and x is invertible.

[The concept of maximal left ideal (or maximal right ideal) is tied to the notion of *simple module* which we shall discuss at length later.]

We want to show that every ring has epimorphic images which are simple. In view of the above discussion, this amounts to showing that every ring has a maximal two-sided ideal. More generally, we shall prove

THEOREM. *Let A be a nontrivial ring and let $\mathfrak{a}$ be a proper left (right, two-sided) ideal of A. Then there exist maximal left (right, two-sided) ideals in A containing $\mathfrak{a}$.*

The proof of this theorem depends on the method of transfinite induction which we discuss in its most commonly used form: Zorn's Lemma. This principle is closely linked to a set of assertions in Set Theory all equivalent to the so-called Axiom of Choice. Naively, the Axiom of Choice asserts that given a collection (perhaps very "large") of sets, one can form a set by choosing one element from each set. That does not sound very profound, but when used in full generality it can produce startling results. (The so-called Banach-Tarski "paradox" in measure theory is one result.) For this reason, many mathematicians are reluctant to use the most general axiom of choice and its consequences. However, it is difficult to do some important mathematics without it, as the above theorem illustrates.

The particular consequence of the axiom of choice which we shall use is Zorn's Lemma. It involves a certain property of some partially ordered sets. (A set S is a partially ordered set if there is a relation $a \leq b$ defined on S satisfying

(i) $a \leq a$ for all $a \in S$,
(ii) $a \leq b$ and $b \leq a \Rightarrow a = b$, and
(iii) $a \leq b$ and $b \leq c \Rightarrow a \leq c$.

A subset C of a partially ordered set S is said to be *linearly ordered* if given any two elements $a, b \in C$ either $a \leq b$ or vice versa. A partially ordered set S is said to be *inductively ordered* if for any linearly ordered subset has an upper bound. An element m of a partially ordered set is called maximal if $m \leq a \Rightarrow m = a$. We may now state

ZORN'S LEMMA. *Any element of a nonempty inductively ordered set is bounded above by a maximal element.*

Clearly, by suitable change of terminology we can deal with minimal elements. Zorn's Lemma is in fact equivalent to the Axiom of Choice so unless you are interested in the logical intricacies, it is just as good to take it as an axiom and proceed from there. That is what we shall do here.

We are now ready to prove the existence of a maximal left ideal in a ring A. Let S be the set of all proper left ideals of A ordered by inclusion. It is nonempty since (0) is an ideal. (In a field, it could be the unique maximal ideal.)

This set is inductively ordered. For, let C be any collection of left ideals in A with the property that for any two ideals in C, one is contained in the other. Form the union $\mathfrak{b}$ of all the ideals in C. It certainly is an upper bound since it contains them all, and moreover it *is* a left ideal. For, first it is a subgroup since if $x, y \in \mathfrak{b}$, each is in some left ideal in C so they are both in the larger of the two, and so is their sum or difference. It is even easier to see that $a \in A, x \in \mathfrak{b} \Rightarrow ax \in \mathfrak{b}$. Finally, it is a proper left ideal since otherwise $1 \in \mathfrak{b}$ which means that 1 is in one of the ideals in C so that one of those ideals is not proper.

By Zorn's Lemma, S contains a maximal element as claimed.

The corresponding arguments for right or two-sided ideals are basically the same.

Exercises.

1. Let F be a field, and let $A = F \times F \times \cdots \times F$ be the n-fold cartesian product of F with itself made into a ring by component-wise operations. Find all the maximal ideals in A.

2. (a) Let I be the set of all non-units in a ring A. Prove that if I is a 2-sided ideal then A/I is a division ring.

(b) Suppose A has the property that for each $x \in A$ either x is a unit or $1 - x$ is a unit. Show that the set of non-units forms a 2-sided ideal.

3. The Chinese Remainder Theorem

The multiplication of additive subgroups of A satisfies the associative and distributive laws:

$$\mathfrak{a}(\mathfrak{b}\mathfrak{c}) = (\mathfrak{a}\mathfrak{b})\mathfrak{c}$$
$$\mathfrak{a}(\mathfrak{b} + \mathfrak{c}) = \mathfrak{a}\mathfrak{c} + \mathfrak{a}\mathfrak{c}$$
$$(\mathfrak{a} + \mathfrak{b})\mathfrak{c} = \mathfrak{a}\mathfrak{c} + \mathfrak{b}\mathfrak{c}.$$

(See the Exercises for Section 1.)

Moreover, it is not hard to see that if $\mathfrak{a}$ is a left ideal then $\mathfrak{a}\mathfrak{b}$ is also a left ideal. Similarly, if $\mathfrak{b}$ is a right ideal then $\mathfrak{a}\mathfrak{b}$ is a right ideal. In particular, if $\mathfrak{a}$ is a left ideal and $\mathfrak{b}$ is a right ideal, then $\mathfrak{a}\mathfrak{b}$ is a two-sided ideal.

Note that in the above formulas, we have used the sum $\mathfrak{a} + \mathfrak{b}$ of two additive subgroups of A. (Since as additive group, A is abelian, the subgroup $\mathfrak{a}+\mathfrak{b}$ in fact consists of all sums $x+y$ where $x \in \mathfrak{a}$ and $y \in \mathfrak{b}$.) If $\mathfrak{a}$ and $\mathfrak{b}$ are left (right, two-sided) ideals then $\mathfrak{a}+\mathfrak{b}$ is a left (right, two-sided) ideal, the smallest such containing $\mathfrak{a}$ and $\mathfrak{b}$. Similarly, we can form the intersection $\mathfrak{a}\cap\mathfrak{b}$ of 2 left (right, two-sided) ideals, and the result is again a left (right, two-sided) ideal. More generally, any arbitrary intersection of left (right, two-sided) ideals is again a left (right, two-sided) ideal: in fact the largest left (right, two-sided) ideal contained in all the ideals.

Note that if $\mathfrak{a}$ and $\mathfrak{b}$ are two-sided ideals, then so is $\mathfrak{a}\mathfrak{b}$ and in addition $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{a} \cap \mathfrak{b}$.

Let A and B be rings. We can form the additive group $A \times B$, and we can define a multiplication operation on it by

$$(a, b)(a', b') = (aa', bb').$$

It is straightforward to check that $A \times B$ becomes a ring with this operation. It is called the *direct* product ring. Consider the subgroup $A \times \{0\} = \{(a, 0) \mid a \in A\}$. A simple calculation shows that it is closed under either left or right multiplication by arbitrary elements of $A \times B$ so it is a two-sided ideal. In fact, it is the kernel of the map $A \times B \to B$ defined by $(a, b) \mapsto b$. (That map is clearly a ring epimorphism.) Similarly, $\{0\} \times B$ is a two-sided ideal in A, the kernel of the projection $A \times B \to A$.

The Chinese Remainder theorem gives us a way of dissecting a ring into a direct product under appropriate circumstances.

THEOREM. *(Chinese Remainder Theorem) Let A be a ring, and suppose $\mathfrak{a}_i, i = 1, 2, \ldots, n$ are two-sided ideals in A such that*

$$\mathfrak{a}_i + \mathfrak{a}_j = A \quad \text{for } i \neq j.$$

Then

$$A/\mathfrak{a} \cong \prod_i A/\mathfrak{a}_i \text{ where } \mathfrak{a} = \cap_i \mathfrak{a}_i.$$

Note: The property $\mathfrak{a} + \mathfrak{b} = A$ for two sided ideals is called *comaximality.*

PROOF. Let $p_i : A \to A/\mathfrak{a}_i$ be the canonical epimorphism, and define $p : A \to \prod A/\mathfrak{a}_i$ by $p(a) = (p_1(a), \ldots, p_n(a))$. Clearly, p is a ring homomorphism, and

$$\operatorname{Ker} p = \{a \mid p_i(a) = 0 \text{ for } i = 1, \ldots, n\} = \cap \mathfrak{a}_i.$$

The theorem will follow from the first isomorphism theorem if we can show that p is an epimorphism.

Let d_i be the element $(0, \ldots, 1, \ldots, 0)$ which is 0 at all components except the ith where it is one. It clearly suffices to show that $d_i = p(a_i)$ for some $a_i \in A$ for each $i = 1, \ldots, n$. (For, in that case $p(aa_i) = p(a)p(a_i) = p(a)d_i = p_i(a)$ and clearly the direct product is generated by elements of this form.)

On the other hand, to say $p(a) = d_i$ amounts to asserting that $p_i(a) = 1$ and $p_j(a) = 0$ for $j \neq i$. This in turn amounts to asserting that $a \equiv 1 \bmod \mathfrak{a}_i$ and $a \equiv 0 \bmod \mathfrak{a}_j$ for $j \neq i$. By hypothesis, $\mathfrak{a}_i + \mathfrak{a}_j = A$. Assume first $n > 2$ and $i > 2$. Then

$$\begin{aligned} A = AA &= (\mathfrak{a}_i + \mathfrak{a}_1)(\mathfrak{a}_i + \mathfrak{a}_2) \\ &= \mathfrak{a}_i^2 + \mathfrak{a}_i\mathfrak{a}_2 + \mathfrak{a}_1\mathfrak{a}_i + \mathfrak{a}_1\mathfrak{a}_2 \\ &\subseteq \mathfrak{a}_i + \mathfrak{a}_1\mathfrak{a}_2 \end{aligned}$$

since the first 3 terms are contained in $\mathfrak{a}_i$. However, A is the whole ring, so

$$A = \mathfrak{a}_i + \mathfrak{a}_1\mathfrak{a}_2.$$

Iterating this argument shows that

$$A = \mathfrak{a}_i + \mathfrak{a}_i'$$

where $\mathfrak{a}_i'$ is the product in order of all the ideals $\mathfrak{a}_1\mathfrak{a}_2 \ldots \mathfrak{a}_n$ with the ith term $\mathfrak{a}_i$ *omitted*. The same argument works with a change of notation if $i = 1$ or 2 and $n > 2$. For $n = 2$, simply take $\mathfrak{a}_1' = \mathfrak{a}_2$ and $\mathfrak{a}_2' = \mathfrak{a}_1$.

It follows in any event that we can write

$$1 = b_i + a_i \text{ where } b_i \in \mathfrak{a}_i \text{ and } a_i \in \mathfrak{a}_i'.$$

We have $a_i \equiv 1 \bmod \mathfrak{a}_i$ and since $a_i \in \mathfrak{a}_i' \subseteq \mathfrak{a}_j$ for $j \neq i$, we also have $a_i \equiv 0 \bmod \mathfrak{a}_j$ for $j \neq i$ as required. □

Example: Let $A = \mathbf{Z}$. Then two ideals $n\mathbf{Z}$ and $m\mathbf{Z}$ are comaximal if and only if $\gcd(n, m) = 1$. For, if $n\mathbf{Z} + m\mathbf{Z} = \mathbf{Z}$, the equation $nx + my = 1$ has integral solutions x, y, and it is easy to see that n and m have to be relatively prime. Conversely, if n and m are relatively prime, since $n\mathbf{Z} + m\mathbf{Z} = d\mathbf{Z}$ for some d, it is easy to see that $d = 1$ so the ideals are comaximal. Also, in this case, it is easy to see that $(n\mathbf{Z}) \cap (m\mathbf{Z}) = nm\mathbf{Z}$ so that the Chinese Remainder Theorem (in its original form) asserts that if $\gcd(n, m) = 1$, then given a and b, we can find x such that $x \equiv a \bmod m$ and $x \equiv b \bmod n$, and such an x is unique $\bmod mn$. We leave it to the student to state the corresponding result for n integers m_i which are relatively prime in pairs.

It is not hard to see that the group of units in a direct product ring is the direct product of the unit groups in the components:

$$U(A \times B) \cong U(A) \times U(B).$$

Hence, if n and m are relatively prime, we get from the Chinese remainder theorem

$$U(\mathbf{Z}/nm\mathbf{Z}) \cong U(\mathbf{Z}/n\mathbf{Z}) \times U(\mathbf{Z}/m\mathbf{Z}).$$

By definition, the Euler ϕ function is given by

$$\begin{aligned} \phi(n) = &\text{ the order of } U(\mathbf{Z}/n\mathbf{Z}) \\ = &\text{ the number of } i, 0 < i < n, \gcd(i, n) = 1. \end{aligned}$$

Hence, if $\gcd(n, m) = 1$, then

$$\phi(nm) = \phi(n)\phi(m).$$

Since for a prime power, $\phi(p^r) = p^r - p^{r-1}$, the above result lets us determine the order of $U(Z/n\mathbf{Z})$ for any n.

Exercises.

1. Let p be a prime and let $n = p^r$. The ring homomorphism $\mathbf{Z}/p^r\mathbf{Z} \to \mathbf{Z}/p\mathbf{Z}$ takes units into units.
   (a) Show that $U(\mathbf{Z}/p^r/\mathbf{Z}) \to U(\mathbf{Z}/p\mathbf{Z})$ is onto. Hint: Show that the kernel is $U_{p,r} = 1 + p\mathbf{Z}/p^r\mathbf{Z}$.
   We will show later in the course that $U(\mathbf{Z}/p\mathbf{Z})$ is cyclic of order $p - 1$. Use this in what follows.
   (b) Show that $U(\mathbf{Z}/p^r\mathbf{Z})$ contains an element of order $p - 1$.
   (c) Show that $U(\mathbf{Z}/p^r\mathbf{Z})$ is the direct product of a cyclic group of order $p - 1$ with $U_{p,r}$.
   (d) Show that $U_{p,r}$ is cyclic. Hint: Use induction on r.
   (e) Conclude that $U(p^r)$ is cyclic if p is odd or if $p = 2$ and $r = 1, 2$ and it is the direct product of a cyclic group of order two with a cyclic 2-group if $p = 2$ and $r > 2$. This is a famous theorem of Gauss in number theory.

4. Domains

A ring A is called a *domain* if it is commutative and for $x, y \in A$, $xy = 0$ implies $x = 0$ or $y = 0$. (If $xy = 0$ without x or $y = 0$, then x and y are called *zero divisors*.)
Examples:

1) Any field is of course a domain.
2) $\mathbf{Z}$ is a domain.
3) The set of all complex numbers of the form $a + bi$ with $a, b \in \mathbf{Z}$ is a domain since it is a subring of the field $\mathbf{C}$. This ring is called the ring of *Gaussian integers* and it is denoted $\mathbf{Z}[i]$.
4) Note that $M_n(F)$ has lots of zero divisors. (Of course, it is also not commutative.) Any direct product of rings will also have lots of zero divisors: multiply elements non-zero in different components.

Clearly, any subring of a field is a domain. Conversely, any domain can be imbedded as a subring of a field as follows. Let E be the set of pairs (a, b) where $a, b \in A$ and $b \neq 0$. Define a relation $\sim$ on E by $(a, b) \sim (c, d)$ if and only if $ad = bc$. It is not difficult to check that this is an equivalence relation. (For example, $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f) \Rightarrow ad = bc$ and $cf = de \Rightarrow adf = bcf = bde \Rightarrow d(af - be) = 0 \Rightarrow af = be \Rightarrow (a, b) \sim (e, f)$. Note the argument uses both the fact that A is commutative and that it is a domain.) Let Q denote the set of equivalence classes of this relation. Denote the equivalence class of (a, b) by a/b. Define operations on Q by

$$a/b + c/d = (ad + bc)/bd$$

and

$$(a/b)(c/d) = (ac)/(bd).$$

Several tedious but routine arguments show that these operations are well defined (i.e., the results depend only on the equivalence classes of the operands), and that Q with these operations is a ring. The 0 element is $0/1$ ($= 0/d$ for any $d \neq 0$ in A), and the identity is $1/1$ ($= d/d$ for every nonzero $d \in A$.) Every nonzero element of Q is a unit; in fact $(a/b)(b/a) = 1/1$ for any nonzero a and b in A. Hence, Q is a field.

Define $i : A \to Q$ by $i(a) = a/1$. It is easy to see that i is a ring homomorphism. It is in fact a monomorphism. For, $i(a) = a/1 = 0/1 \Leftrightarrow a1 = (1)(0) = 0$. Hence, $\operatorname{Im} i$ is a ring isomorphic to A. Moreover, every nonzero element a/b can be written

$$a/b = (a/1)(1/b) = (a/1)(b/1)^{-1}$$

as a quotient in Q of elements in $\operatorname{Im} i$.

The above construction embeds a ring isomorphic to A in a field—which is not exactly what was promised. However, it is easy to use this construction to imbed A itself in a field. Namely, let Q' be the union of A and the complement of $i(A)$ in Q. Define operations on Q' in the obvious way. That is, when an operand or the result of an operation in Q happens to be in $i(A)$, just use the corresponding element of A instead. Q' will then be a field isomorphic to Q and it will contain A.

It is more to the point, however, to consider in general ring monomorphisms $j : A \to P$ where P is a field such that every element of P can be written $i(a)i(b)^{-1}$ for $a, b \neq 0$ in A. We have demonstrated the existence of one such monomorphism. If $j : A \to P$ and $j' : A \to P'$ are two such then it is easy to see that $h : P \to P'$ defined by $h(j(a)j(b)^{-1}) = j'(a)j'(b)^{-1}$ is well defined and a ring isomorphism. Moreover, it makes the diagram below commute

$$\begin{array}{ccc} & & P \\ & j \nearrow & \\ A & & \downarrow h \\ & j' \searrow & \\ & & P' \end{array}$$

Finally, if the diagram commutes, i. e. $h(j(a)) = j'(a)$ then h is clearly the same homomorphism as defined above. Hence, the isomorphism h is unique given that the above diagram commutes.

We call such a field P (more correctly, the monomorphism j) a *quotient* field or *field of fractions* of A, and it is unique up to unique isomorphism in the sense described above. As mentioned earlier, we can in fact assume j is an actual inclusion.

One can generalize the above construction by considering only pairs (a, s) where s is restricted to any appropriate subset of A. For example, the field of fractions of $\mathbf{Z}$ is the field $\mathbf{Q}$ of rational numbers. However, we might consider the subring of $\mathbf{Q}$ of all fractions with denominators relatively prime to some fixed integer. These form a ring called a localization. The concept of localization is extremely important in algebra, and we shall return to it later.

Exercises.

1. Show that any finite domain is a field.

5. Polynomial rings

Let B be a commutative ring, and let A be a subring. If U is a subset of B, we denote by $A[U]$ the smallest subring of B containing A and U. We call it the subring generated by A and U (or just by U where there is only one A under consideration.) If $U = \{u_1, u_2, \ldots, u_n\}$ we use the notation $A[u_1, ..., u_n]$.

If $U = \{u\}$ consists of a single element, then $A[u]$ must contain every *polynomial* $a_0 + a_1u + a_2u^2 + \cdots + a_nu^n$ with coefficients a_i in A. On the other hand, the sum and product of two such polynomials in u are again polynomials in u, so the set of such polynomials is a subring of B containing A and u. Hence, $A[u]$ is the subring of polynomial expressions in u with coefficients in A.

It is clear from the definition that $A[U \cup V] = A[U][V]$. Using that rule, it is not hard to see what $A[u_1, \ldots, u_n]$ looks like.

Two polynomial expressions in an element u in B are of course equal if corresponding coefficients are equal. However, the converse statement may not be true. For example, in $\mathbf{Z}[i]$ (as a subring of $\mathbf{C}$) we have $i + i^2 = -1 + i$. In general, there may be many relations among the powers of u. As in the case of a free group, we are motivated to construct a ring $A[X]$ generated by A and an element X called an *indeterminate* and such that there are no non-trivial relations among the powers of X.

More generally, we proceed as follows. Let X be a set. A *monomial* in the set X is defined to be a function from X to the set of non-negative integers which assumes the value 0 for all but a finite number of elements $x \in X$. We denote such a function in the following special form

$$\prod_{x \in X} x^{n_x}$$

The set $M(X)$ of all such monomials becomes a monoid under the operation defined by

$$\prod x^{n_x} \prod x^{m_x} = \prod x^{n_x + m_x}$$

The monomial 1 is that with all exponents (i.e., values of the function) equal to 0. The values n_x are called the exponents of the monomial. Clearly, they completely determine it.

A *polynomial* in the set of *indeterminates* X is defined to be a function from the monoid $M(X)$ to the ring A which is zero except for a finite number of monomials. Such a function is represented formally

$$\sum a(\mu)\mu.$$

Polynomials are then added and multiplied according to the rules

$$\sum a(\mu)\mu + \sum b(\mu)\mu = \sum (a(\mu) + b(\mu))\mu$$

and

$$(\sum a(\mu)\mu)(\sum b(\nu)\nu) = \sum c(\sigma)\sigma$$

where $c(\sigma) = \sum a(\mu)b(\nu)$—the sum being taken over all pairs (μ, ν) such that $\mu\nu = \sigma$. Note that there are only finitely many such pairs such that $a(\mu)b(\nu) \neq 0$, so the sum is well defined. It is routine but very tedious to show that with these operations, the set of all such polynomials forms a ring which we denote by $A[X]$.

We use the common notation $f(X)$ to denote a polynomial in the indeterminates X. If $X = \{x\}$, we just write $f(x)$. The values of the polynomial $a(\mu)$ are called its coefficients, and since it has just been defined as a function on monomials, clearly the polynomial is completely determined by its coefficients.

As above, we use appropriate notation if X is finite or if it consists of a single element. Note that if B is any ring containing $A[X]$ as a subring, then the subring generated by A and X is just $A[X]$, so the new notation is consistent with the previous notation. But it is certainly possible to get confused when using this notation. One must distinguish the case in which the set X is a set of indeterminates (as just discussed) rather than some arbitrary subset of B.

Let A be a commutative ring, and let X be a set. We may define a ring homomorphism $i : A \to A[X]$ by $i(a) =$ the polynomial all of whose coefficients $a(\mu) = 0$ except for $a(1) = a$. We call $i(a)$ a constant polynomial. Clearly, i is a ring homomorphism, and its image - the subring of constant polynomials - is isomorphic to A. In most cases, we oversimplify and just identify A with $i(A)$ as a subring of $A[X]$. Similarly, we usually identify X and $M(X)$ with subsets of $A[X]$.

More generally, if A is a commutative ring, we call a ring homomorphism $j : A \to B$ into a commutative ring B an A-algebra. (Using the term loosely, we may just call B an A-algebra.) A similar concept may be defined if B is non-commutative, but it is a bit more involved, so we shall save the general concept for later. The polynomial ring plays the same role for A-algebras that the free group plays for groups.

THEOREM. *Let A and B be commutative rings, and let $j : A \to B$ be a ring homomorphism. Let X be a set, and suppose there is a function $\alpha : X \to B$. Then there is a unique ring homomorphism*

$$J : A[X] \to B$$

such that $J(a) = j(a)$ for $a \in A$ and $J(x) = \alpha(x)$ for $x \in X$.

PROOF. For any such J, it is clear that

$$J\left(\prod x^{n_x}\right) = \prod \alpha\left(x\right)^{n_x}$$

where the product on the right is calculated in the ring B. Also, for any polynomial $f(X) = \sum a(\mu)\mu$, we must have

$$J(f(X)) = \sum j(a(\mu))J(\mu)$$

where $J(\mu)$ is as above. On the other hand, it is not hard to check that J defined by these formulas is in fact a ring homomorphism with the right properties. □

Note that since a polynomial is completely determined by its coefficients, it is easier to define this mapping than was true for the analogous case for free groups.

Let X and Y be disjoint sets. Using the universal mapping property described above, one may define appropriate homomorphisms

$$A[X][Y] \to A[X \cup Y] \text{ and } A[X \cup Y] \to A[X][Y]$$

such that both compositions are the identity. Hence,

$$A[X][Y] \cong A[X \cup Y] \cong A[Y][X].$$

Exercises.

1. (a) Let $A = \mathbf{Q}[X]$ where $\mathbf{Q}$ is the field of rational numbers. Show that $A/A(X^2 - 1) \cong \mathbf{Q} \times \mathbf{Q}$.
   (b) Show that the natural homomorphism $\mathbf{Z}[X]/\mathbf{Z}[X](X^2 - 1) \to \mathbf{Z} \times \mathbf{Z}$ is not an epimorphism.
2. Let $A = k[X, Y]$ where k is a commutative ring.
   (a) Show that $U(A) = U(k)$.
   (b) Show that $A/AX \cong k[Y]$.

6. Unique Factorization

Let A be a domain. If $a, b \in A$, we say $a \mid b$ if $\exists c \in A$ such that $b = ac$. If $a \mid b$ and $b \mid a$, we say that a and b are *associates*. That amounts to saying that a is a unit times b or vice-versa.

An element $p \neq 0$ of A is called *irreducible* provided

$$a \mid p \Rightarrow a \text{ is a unit or } p \mid a.$$

In the special case, $A = \mathbf{Z}$, we use the term *prime* for a positive irreducible element.

A domain A is called a *unique factorization domain* (UFD) provided

(UF1) Every nonzero non-unit in A is a product of irreducible elements.
(UF2) Given two factorizations of the same element, there is a one-to-one correspondence between the two sets of irreducible factors such that corresponding factors are associates.

Examples:

1) As you are undoubtedly aware, the Fundamental Theorem of Arithmetic asserts that $\mathbf{Z}$ is a UFD.
2) Every field is trivially a UFD since there are no irreducible elements.
3) The following ring is not a UFD. Let A be the subring of $\mathbf{C}$ given by $A = \mathbf{Z}[\alpha]$ where α is one of the two roots of $\alpha^2 = -5$. Note that every element of A may be written uniquely $a + b\alpha$ where a and $b \in \mathbf{Z}$. Define
$$N(a + b\alpha) = (a + b\alpha)(a - b\alpha) = a^2 + b^2.$$
By explicit calculation, it is easy to see that $N(xy) = N(x)N(y)$ for x and $y \in A$. Since N is multiplicative on A, it necessarily takes $U(A)$ into $U(\mathbf{Z}) = \{1, -1\}$. Conversely, if $N(x) = \pm 1$, it follows that x must be a unit. Consider now the equations
$$6 = (3)(2) = (1 - \alpha)(1 + \alpha).$$
By means of appropriate arguments using $N(x)$ it is possible to show that $2, 3, 1 + \alpha$, and $1 - \alpha$ are all irreducible, and that 2 and 3 are not associates of either $1 + \alpha$ or $1 - \alpha$. (Details left to you.) It follows that A cannot be a UFD since UF2 fails.

Principal Ideal Domains. A commutative ring A is called a *principal ideal* domain (PID) if it is a domain and every ideal is of the form Ax for some $x \in A$.

Examples:

1) We have remarked previously that every ideal in $\mathbf{Z}$ is principal. The crucial step (done in an exercise on subgroups of $\mathbf{Z}$) used the division algorithm.
2) Let k be a field. Then $A = k[X, Y]$ where X and Y are indeterminates is not a PID. For, the ideal $AX + AY$ is not principal since if it were X and Y would have to be multiples of the same element of A, but X and Y are clearly non-associate irreducible elements.

THEOREM. *Every PID A is also a UFD.*

PROOF. First we demonstrate UF1. Let a be a nontrivial non-unit. If a cannot be written as a product of irreducible elements, we may by a sequence of proper factorizations (of previously found factors) produce

a sequence $a = a_1, a_2 \mid a_1, \ldots, a_{n+1} \mid a_n, \ldots$ where successive factors are not associates. It follows that the sequence of ideals

$$Aa_1 \subseteq Aa_2 \subseteq \cdots \subseteq Aa_n \subseteq \ldots$$

forms a strictly increasing chain. (Otherwise, two successive generators would have to divide one another.) Let $\mathfrak{a}$ be the union of all of the ideals in that chain. Because it is an increasing chain, it follows that $\mathfrak{a}$ is again an ideal. (See the proof of the existence of maximal ideals using Zorn's Lemma.) Hence, $\mathfrak{a} = Aa$ for some $a \in A$ since every ideal is principal. However, we must have $a \in Aa_n$ for some n so a is divisible by a_n and hence by every a_m with $m > n$. On the other hand, $a_m \in Aa = \mathfrak{a}$ so a also divides a_m. It follows that a and a_m are associates for every $m > n$—which is nonsense.

To prove UF2, we derive instead the following statement:

(UF2′) If p is irreducible and $p \mid ab$ then $p \mid a$ or $p \mid b$.

UF2′ implies UF2 because if we had two factorizations of the same element into irreducible elements, we could use UF2′ and the fact that we are operating in a domain to strike off associate factors from both factorizations (possibly producing units in their place because the stricken factors need only be associates) until we run out of factors.

To prove UF2′ suppose p does not divide a, and consider the ideal $Ap + Aa$ which they generate. Let $Ap + Aa = Av$. Since $v \mid p$ and since p is irreducible, either v is a unit or it is an associate of p. In the latter case, since $v \mid a$, we would have $p \mid a$; hence we may assume v is a unit. In that case $Ap + Aa = A$, and we may write $1 = xp + ya$ for appropriate $x, y \in A$. Hence, $b = bxp + bya$, and since both terms are divisible by p (since $p \mid ab$), it follows that $p \mid b$ as required. □

Note that UF1 and UF2 together imply UF2′. For, if p is irreducible and divides xy then up to an associate it must appear in the unique factorization of xy. However, it is clear that the unique factorization of xy is the product of the unique factorizations of x and y. Hence, p must occur (up to an associate) in one of those.

We saw earlier that the reason why we know that $\mathbf{Z}$ is a PID is because of the division algorithm. More generally, we define a *Euclidean domain* to be a domain A together with a function

$$\phi : A - \{0\} \to \text{ the set of non-negative integers}$$

such that

(E1) $$a \mid b \Rightarrow \phi(a) < \phi(b)$$

and

(E2) for each $a, b \neq 0 \in A, \exists q, r \in A$ such that $a = bq + r$ and $\phi(a) < \phi(b)$ or $r = 0$.

Examples:

1) For $\mathbf{Z}$ take $\phi(n) = |n|$.
2) For $\mathbf{Z}[i]$ take $\phi(x) = |x|^2$.
3) Let k be a field and let $A = k[X]$ be the polynomial ring over k in a *single* indeterminate X. Take $\phi(f(X)) = \deg f(X) =$ the highest power of X occurring with non-zero coefficient. We have $\deg(f(X)g(X)) = \deg f(X) + \deg g(X)$. Using ordinary long division of polynomials yields a division algorithm as required.
4) There are PID's which are not Euclidean but showing that a ring is not Euclidean is tricky.

THEOREM. *Every Euclidean domain is a PID.*

PROOF. Left to the student.

Exercises.

1. Let k be a field. Show that $k[X]$ is a Euclidean domain using $N(f(X)) = \deg f(X)$.
2. Show that every Euclidean domain is a PID.

7. Unique factorization in polynomial rings and Gauss's Lemma

We shall prove the following important theorem.

THEOREM. *If A is a UFD, then the polynomial ring $A[X]$ in a single indeterminate is a UFD.*

By induction, this gives

COROLLARY. *If A is a UFD, then the polynomial ring $A[X_1, \ldots, X_n]$ in n indeterminates is a UFD.*

In this way we may construct UFD's by starting for example with some PID such as a field k or $\mathbf{Z}$. The rings constructed, however, may not be PID's. For example, $\mathbf{Z}[X]$ is not a PID. To see this, consider the ideal (p, X) generated by p and X where p is a prime. If it were principal, its generators would have to be divisible by the same element which is not possible since they are clearly non-associate irreducible elements of the ring. Similarly, as mentioned earlier, $k[X, Y]$ is not a PID (using basically the same reasoning), but it is a UFD.

To prove the theorem, we need to introduce some concepts. First note that in a UFD A, we may unambiguously define the concept of greatest common divisor (GCD) at least up to associates: given two nontrivial non-units, examination of their unique factorizations shows that there is a factor of both of them which is divisible by all other factors common to both. For $f(X) \in A[X]$, define $c(f)$—called the *content* of $f(X)$—to be the GCD of all its coefficients. A polynomial is called *primitive* if its content is a unit.

LEMMA. *Let $f(X)$ and $g(X)$ be non-constant polynomials in $A[X]$ where A is a UFD. Then $c(fg)$ and $c(f)c(g)$ are associates in A. In particular, the product of primitive polynomials is primitive.*

PROOF. Let $f(X) = c(f)f_1(X)$ where $f_1(X)$ is primitive. Up to associates, we have $c(dg(X)) = dc(g)$ for any primitive polynomial $g(X)$. So we need only prove that the product of primitive polynomials is primitive. Let the first polynomial have coefficients $a_0, \ldots, a_n$ and let the second polynomial have coefficients $b_0, \ldots, b_m$. The coefficients of the product are given by

$$c_k = \sum_{i+j=k} a_i b_j$$

Suppose p is an irreducible element of A which divides every c_k. Suppose $p \mid a_0, \ldots, a_{s-1}$ but p does *not* divide a_s. (There is such an s since the content is a unit.) Similarly, suppose $p \mid b_0, \ldots, b_{t-1}$ but p does *not* divide b_t. Then p divides every term in c_{s+t} except possibly $a_s b_t$. On the other hand, since p divides c_{s+t}, it must also divide $a_s b_t$ and since it is irreducible it must divide a_s or b_t—a contradiction. Hence, the coefficients of the product do not have a common irreducible factor so its content is a unit. □

To prove the theorem, we use the lemma to reduce the general case to the case to the case in which A is a field. In that case, we know the polynomial ring $A[X]$ is a UFD since it is a Euclidean domain (hence PID, hence UFD).

Let $f(X) \in A[X]$. As above, we can factor it into its content times a primitive polynomial. The content has a unique factorization in A by hypothesis. Hence, we may assume $f(X)$ is primitive and has degree greater than zero. Let Q be the field of fractions of A. As before, we may view A imbedded as a subring of Q, and similarly we may view $A[X]$ imbedded as a subring of $Q[X]$. By the field case, we may factor $f(X)$ in $Q[X]$ into a product of factors irreducible in $Q[X]$. For each such factor, since the coefficients are in Q, there is a common denominator d such that d times the factor is in $A[X]$. Since any non-zero element of A (or even Q) is a unit in $Q[X]$, the new polynomial is also irreducible in $Q[X]$. (Note that any polynomial in $A[X]$ which is irreducible in $Q[X]$ is necessarily irreducible in $A[X]$ since a factorization in the smaller ring $A[X]$ would also be a factorization in $Q[X]$.) Form $Df(X)$ where D is the product of the common denominators for all the factors just discussed. It is clearly the product of irreducible polynomials in $A[X]$. Compare contents. The content of $Df(X)$ is just D since $f(X)$ is primitive. The content of the product is the product of the contents of the factors by the lemma. If we remove the content of each factor from the equation, we get a factorization of $f(X)$ as a product of polynomials in $A[X]$ which are irreducible in $Q[X]$—hence also in $A[X]$.

Suppose we had two factorizations of $f(X)$ into irreducible factors in $A[X]$. Comparing contents as above, we can reduce the question of uniqueness to the case that $f(X)$ is primitive in which case any irreducible factors in $A[X]$ are also primitive - as above. In $Q[X]$ two factorizations would have to be the same except for multiplication by units. However, it is not hard to see that the only units in $Q[X]$ are the non-zero constants. (Use the fact that the degree of a product is the sum of the degrees.) Since two primitive polynomials in $A[X]$ differing by a factor in Q must in fact differ by a unit in A (—use Gauss's Lemma and the fact that any element in Q is a quotient of elements of A—) we are done. □

Gauss's Lemma. The lemma is often called Gauss's Lemma. But, if you carefully examine the proof the the theorem, you will implicit in it the following result.

If A is a UFD, Q is its field of fractions, and $f(X) \in A[X]$ is a non-constant polynomial, then any factorization $h(X) = F(X)G(X)$ into non-constant polynomials in $Q[X]$ yields a factorization $h(x) = f(x)g(x)$ in $A[X]$ where $f(X)$ is a constant multiple of $F(X)$ and $g(x)$ is a constant multiple of $G(X)$.

This result is intimately related to the previous lemma and is usually included under the rubric Gauss's Lemma.

We saw earlier that an ideal $\mathfrak{a}$ in a commutative ring A is maximal if and only if $A/\mathfrak{a}$ is a field. An ideal $\mathfrak{p}$ is called *prime* if and only if $A/\mathfrak{p}$ is a domain—that is $A/\mathfrak{p}$ has no zero divisors. Thus, $\mathfrak{p}$ is prime if and only if $xy \in \mathfrak{p} \Rightarrow x \in \mathfrak{p}$ or $y \in \mathfrak{p}$.

If A is a domain and Ap is prime, then p is necessarily irreducible. For, suppose $p = ab$. Since $ab \in Ap$, either $a \in Ap$ or $b \in Ap$, so that either a is an associate of p or b is an associate of p and the other is a unit. Conversely, we have

PROPOSITION. *If A is a UFD and p is irreducible, then Ap is prime.*

PROOF. This is just a restatement of Uf2′: $p \mid xy \Rightarrow p \mid x$ or $p \mid y$. □

Note that a maximal ideal is necessarily prime since every field is a domain. The converse is not in general true.

Example:

Let $A = k[X, Y]$ be a polynomial ring in two indeterminates with coefficients in a field k. By the above proposition, AX is prime since X is irreducible. However, $AX \subseteq AX + AY$ and they are not equal, so AX is not maximal. ($AX + AY$ is maximal. Why?)

THEOREM. *A domain A is a PID if and only if it is a UFD and every prime ideal is maximal.*

PROOF. Suppose A is a PID. We already know it is a UFD. Let Ap be a prime ideal so p is irreducible. If $Ap \subseteq Aq$ for some $q \in A$, then $q \mid p$ so q is a unit or q is an associate of p. In the former case $Aq = A$ and in the latter case $Aq = Ap$. Hence Ap is maximal.

Conversely, suppose the two conditions in the statement of the theorem hold. Let $\mathfrak{a}$ be an ideal in A. If $\mathfrak{a} = A$, it is principal so assume $\mathfrak{a}$ does not contain a unit. For each element $a \in \mathfrak{a}$, consider its unique factorization $a = p_1 \dots p_r$ into irreducibles. Since a is not a unit, $r \geq 1$. Let s be the minimum value attained by r for all elements $a \in \mathfrak{a}$. s is a function of the ideal $\mathfrak{a}$. We proceed by induction to show that all ideals with given s are principal. Let

$$a = p_1 \dots p_s \text{ with } p_i \text{ all irreducible}$$

be an element of $\mathfrak{a}$ for which the minimum is attained. Consider

$$\mathfrak{a}' = \{x \in A \mid p_1 x \in \mathfrak{a}\}.$$

$\mathfrak{a}'$ is an ideal since $p_1x, p_1y \in \mathfrak{a} \Rightarrow p_1(x + y) \in \mathfrak{a}$ and similarly $p_1x \in \mathfrak{a}$ and $z \in A \Rightarrow p_1zx = zp_1x \in \mathfrak{a}$. Since $p_2 \dots p_s \in \mathfrak{a}'$, there are elements of $\mathfrak{a}'$ which may be represented as products of fewer than s irreducible elements. Thus, by induction we may assume $\mathfrak{a}'$ is principal—$\mathfrak{a}' = Aq$. We have $Aqp_1 = \mathfrak{a}'p_1 \subseteq \mathfrak{a}$ by definition. We shall show they are equal. Since p_1 is irreducible, Ap_1 is prime and hence by hypothesis, it is maximal. Let $y \in A$, y *not* in Ap_1. Since A/Ap_1 is a field, we can solve the congruence

$$xy \equiv 1 \bmod Ap_1,$$

that is we can find x and $z \in A$ such that

$$xy = 1 + zp_1$$

or

$$1 = xy - zp_1.$$

Multiply this equation by $p_2 \dots p_s$. We get

$$p_2 \dots p_s = p_2 \dots p_s xy - zp_1 p_2 \dots p_s.$$

Since $a = p_1 \dots p_s$, the right hand side is in $\mathfrak{a}$ if $y \in \mathfrak{a}$. Suppose now that $y \in \mathfrak{a} - \mathfrak{a}'p_1$. Then y is *not* in Ap_1. For, if it were of the form up_1, since it is in $\mathfrak{a}$, u would have to be in $\mathfrak{a}'$ so that $y = up_1$ would be in $\mathfrak{a}'p_1$. We may now use the above equation. Since the left hand side has been written with fewer than s irreducibles, this contradicts the minimality of s for $\mathfrak{a}$. Hence, $\mathfrak{a} = \mathfrak{a}'p_1 = Aqp_1$ as claimed. □

1. Prove the second form of Gauss's Lemma stated in the text: if A is a UFD, Q is its field of fractions, and $f(X) \in A[X]$ is a non-constant polynomial, then any factorization $h(X) = F(X)G(X)$ into non-constant polynomials in $Q[X]$ yields a factorization $h(x) = f(x)g(x)$ in $A[X]$ where $f(X)$ is a constant multiple of $F(X)$ and $g(x)$ is a constant multiple of $G(X)$. Moreover, show that if $f(X)$ is primitive, then so are $f(x)$ and $g(x)$.