F1.3YE2/F1.3YK3

ALGEBRA AND ANALYSIS

Part 2: ALGEBRA.

RINGS AND FIELDS

LECTURE NOTES AND EXERCISES

Contents

Chapter 1

Revision of Group Theory

1.1 Introduction

The Algebra section of this course is about certain types of algebraic structure that generalise – and include as examples – many such structures with which we are already familiar.

For example, given two natural numbers a, b, we can add and multiply them to get new natural numbers $a + b$ and ab. We can also subtract one from the other, but the result $a - b$ is not always a natural number. (It may be a negative integer.)

If we allow a, b to be arbitrary integers, we can add, multiply and subtract them and the result will also be an integer. We can also divide a by b (provided $b \neq 0$), but the result will not always be an integer.

If a, b are arbitrary rational numbers (or real numbers, or complex numbers) then we can add, multiply and subtract to get a new number of the same type. We can also divide a by b if $b \neq 0$.

We will be interested in properties of $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ with respect to the algebraic operations of addition, subtraction, multiplication and division, but we will also be interested in similar algebraic operations on other objects.

For example, we know that we can add or subtract two vectors in $\mathbb{R}^n$ to get a new vector in $\mathbb{R}^n$. If A, B are $n \times n$ matrices, we can add, subtract and multiply to get new $n \times n$ matrices $A + B$, $A - B$ and AB. These operations share many of the familiar properties of arithmetic of numbers – *but not all of them.*

For example, if a, b are numbers then we know that $ab = ba$. But there are examples of 2×2 matrices A, B such that $AB \neq BA$. We are interested in developing an abstract theory that will apply to a wide variety of different algebraic situations.

All the above examples have a common feature: they are abelian groups with respect to addition. The purpose of this chapter is to revise the features of group theory that are relevant to our later studies.

1.2 Binary Operations

A *binary operation* $*$ on a set A is a map $A \times A \to A$, written $(a, b) \mapsto a * b$.

Examples include most of the standard arithmetic operations on the real or complex numbers, such as addition $(a + b)$, multiplication $(a \times b)$, subtraction $(a - b)$. Other examples of binary operations (on suitably defined sets) are exponentiation a^b (on the set of positive reals, for example), composition of functions, matrix addition and multiplication, subtraction, vector addition, vector product of 3-dimensional vectors, and so on.

Definition A binary operation $*$ on a set A is *commutative* if $a * b = b * a \ \forall a, b \in A$.

Addition and multiplication of numbers is commutative, as is addition of matrices or vectors, union and intersection of sets, etc. Subtraction of numbers is not commutative, nor is matrix multiplication.

Definition A binary operation $*$ on a set A is *associative* if $a * (b * c) = (a * b) * c$ $\forall a, b, c \in A$.

Addition and multiplication (of numbers and matrices) are associative. Examples of nonassociative binary operations are subtraction (of anything), exponentiation of positive reals, and vector product.

Definition An *identity* for a binary operation $*$ on a set A is an element $e \in A$ such that $e * a = a = a * e \ \forall a \in A$.

Examples are 0 for addition of numbers, 1 for multiplication of numbers, the identity $n \times n$ matrix for matrix multiplication. Not all binary operations have identities, however: an example is subtraction of numbers.

Definition Let $*$ be a binary operation on a set A and let $a \in A$. An *inverse* for a (with respect to $*$) is an element $b \in A$ such that $a * b$ and $b * a$ are identities for $*$.

Thus for example -5 is an inverse for 5 with respect to addition of integers; $\frac{2}{3}$ is an inverse for $\frac{3}{2}$ with respect to multiplication of positive real numbers. Other examples are matrix inverses (matrix multiplication) and appropriately defined inverse functions (function composition).

Lemma 1.1 *If a binary operation on a set has an identity, then this identity is unique.*

Proof. Suppose that e and f are both identities for a binary operation $*$ on a set A. Then $e = e * f = f$. The first equality holds because f is an identity. The second holds because e is an identity.

Lemma 1.2 *If $a \in A$ has an inverse with respect to an associative binary operation $*$ on A, then the inverse is unique.*

Proof. Suppose that b and c are both inverses for a. Then $b = b*(a*c) = (b*a)*c = c$. The first equality holds because $a * c$ is an identity, the second because $*$ is associative, and the third because $b * a$ is an identity.

In the last result, the associativity of $*$ is definitely used in the proof. In fact the result is not in general true for nonassociative binary operations.

1.3 Groups

Definition A *group* $(G, *)$ is a set G together with a binary operation $*$ on G such that

1. $*$ is associative;

2. there is an identity $e \in G$ for $*$;

3. every element of G has an inverse with respect to $*$.

If the binary operation $*$ is also commutative, then G is called a *commutative*, or *abelian* group (after a 19th century Norwegian mathematician Niels Abel[1]).

Examples

1. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are all abelian groups with respect to addition. In each case 0 is the identity and the inverse of x is $-x$.

2. Any vector space V is a group with respect to vector addition. The identity is the zero vector, and the inverse of $v \in V$ is $-v$.

3. The set $\mathbb{Q}^*$ of nonzero rational numbers is a group with respect to multiplication. The identity is 1, and the inverse of $\frac{a}{b}$ is $\frac{b}{a}$. Similarly the sets $\mathbb{R}^*$ and $\mathbb{C}^*$ of nonzero real and complex numbers, respectively, are groups with respect to multiplication.

4. The set $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ of complex numbers of modulus 1 is a group with respect to multiplication of complex numbers.

5. The set of invertible $n \times n$ matrices forms a group with respect to matrix multiplication. The identity element is the $n \times n$ identity matrix I_n.

6. Let X be a set. Then the set $\mathcal{S}(X)$ of all *permutations* of X, that is, bijective maps $X \to X$, forms a group with respect to composition of maps. The identity map $X \to X$ is the identity element. This group is called the *symmetric group on* X. In the particular case where X is the set $\{1, 2, \ldots, n\}$, this group is denoted S_n, and called the *symmetric group of degree* n.

[1] `http://www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Abel.html`

7. Let $n > 0$ be an integer, and let $\mathbb{Z}_n$ denote the set $\{0, 1, \ldots, n-1\}$. Define a binary operation $*$ on $\mathbb{Z}_n$ by $a * b = a + b$ if $a + b < n$, and $a * b = a + b - n$ otherwise. Then $\mathbb{Z}_n$ is an abelian group with respect to $*$, with identity 0. The inverse of $a > 0$ in $\mathbb{Z}_n$ is $n - a$ (the inverse of 0 is 0). This group is called the *cyclic* group of order n. The binary operation $*$ is usually denoted $+$, and referred to as *addition modulo* n.

1.4 Cayley tables

One way of describing a binary operation $*$ on a set G (provided G is not too big) is to form a grid with rows and columns labelled by the elements of G, and enter the element $a * b$ in the cell in row a and column b (for all $a, b \in G$). This is called a *multiplication table* or a *Cayley table* or (in the case where $(G, *)$ is a group) a *group table*.

Example

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

This is the Cayley table for $\mathbb{Z}_4$, the cyclic group of order 4.

Example

*	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

This describes a binary operation on the set $G = \{e, a, b, c\}$ with respect to which G is a group.

Two groups G and H are said to be *isomorphic* if they have Cayley tables which are identical, except for relabelling of the elements. For example $G = \{1, i, -1, -i\}$ is a group with respect to multiplication of complex numbers. It is isomorphic to $\mathbb{Z}_4$, because its Cayley table

×	1	i	-1	-i
1	1	i	-1	-i
i	i	-1	-i	1
-1	-1	-i	1	i
-i	-i	1	i	-1

is identical to that of $\mathbb{Z}_4$, if we relabel elements of $\mathbb{Z}_4$ by the rule $0 \mapsto 1$, $1 \mapsto i$, $2 \mapsto -1$, $3 \mapsto -i$ (in other words, $k \mapsto i^k$, $k = 0, 1, 2, 3$).

Question Is the group $G = \{e, a, b, c\}$ in the second example above isomorphic to $\mathbb{Z}_4$?

If G is a group containing only two elements, then G is isomorphic to $\mathbb{Z}_2$. To see this, note that one of the elements of G is the identity e. Let g be the other element of G. The $e * g = g = g * e$, where $*$ is the binary operation in G. What is g^{-1}? Since $g^{-1} * g = e \neq g = e * g$, $g^{-1} \neq e$, so $g^{-1} = g$. Hence $g * g = e$, and this determines the Cayley table of G as

$*$	e	g
e	e	g
g	g	e

Clearly this is the same as that of $\mathbb{Z}_2$, using the relabelling $0 \mapsto e$, $1 \mapsto g$.

Exercise Show that any group containing exactly three elements is isomorphic to $\mathbb{Z}_3$.

1.5 Subgroups

A *subgroup* of a group G is a subset $H \subseteq G$ that is also a group with respect to the same binary operation as G. Examples include $\mathbb{Z}$ as a subgroup of $\mathbb{R}$ (with respect to addition), $\mathbb{R}^*$ and S^1 as subgroups of $\mathbb{C}^*$ with respect to multiplication.

It is important to recognise when a subset of a group G is actually a subgroup of G. The following result gives a useful criterion.

Theorem 1.3 *(The subgroup test) Let G be a group with respect to a binary operation $*$, and let H be a subset of G. Then H is a subgroup of G if and only if the following three conditions are satisfied:*

1. *Closure: $x * y \in H$ $\forall x, y \in H$.*
2. *Identity: $e_G \in H$, where e_G is the identity element of G.*
3. *Inverse: $x^{-1} \in H$ $\forall x \in H$, where x^{-1} is the inverse of x in G (with respect to $*$).*

Proof. Suppose first that H is a subgroup with respect to $*$. Then in particular $*$ is a binary operation on H, in other words a function $H \times H \to H$. Thus $x * y \in H$ whenever $x, y \in H$, giving the closure property.

Being a group, H has an identity e_H, say. Thus $e_H * e_H = e_H$ in G, so

$$e_G = e_H * e_H^{-1} = (e_H * e_H) * e_H^{-1} = e_H * (e_H * e_H^{-1}) = e_H * e_G = e_H \in H$$

(where e_H^{-1} denotes the inverse of e_H in G, and all the calculations are carried out in G.

Finally, let $x \in H$ and let $\overline{x}$ denote the inverse of x in H. Then

$$x^{-1} = x^{-1} * e_G = x^{-1} * e_H = x^{-1} * (x * \overline{x}) = (x^{-1} * x) * \overline{x} = e_G * \overline{x} = \overline{x} \in H.$$

Conversely, suppose that H is a subset of G satisfying the three listed properties.
Since $e_G \in H$, H is nonempty.
By the closure property, $*$ defines a binary operation on H.
This binary operation is associative, since it is already associative on the bigger set G:

$$(x * y) * z = x * (y * z) \ \forall \ x, y, z \in G \Rightarrow (x * y) * z = x * (y * z) \ \forall \ x, y, z \in H.$$

Finally, for any $x \in H$, since $x^{-1} \in H$, x has an inverse in H.
Hence $(H, *)$ is a group. In other words, H is a subgroup of G.

Examples

1. Let $n > 0$ be an integer and let $n\mathbb{Z} = \{nx : x \in \mathbb{Z}\}$, the set of integers divisible by n. Then $n\mathbb{Z}$ is a subgroup of $\mathbb{Z}$ with respect to addition. For the closure property, note that $nx + ny = n(x + y)$. The identity element is $0 = n.0 \in n\mathbb{Z}$. The inverse in $\mathbb{Z}$ of $nx \in n\mathbb{Z}$ is $-(nx) = n(-x) \in n\mathbb{Z}$.

2. Let G be the group of $n \times n$ invertible matrices with respect to matrix multiplication, and let H be the set of matrices $A \in G$ such that $\det(A) = 1$. Then H is a subgroup of G. The closure property follows since $\det(AB) = \det(A)\det(B)$; $I_n \in H$ since $\det(I_n) = 1$; and finally if $\det(A) = 1$ then $\det(A^{-1}) = 1$.

3. Let $G = S_n$ and let $H = \{\sigma \in G : \sigma(1) = 1\}$. Then H is a subgroup of G. To check the closure property, if $\sigma, \tau \in H$ then $\sigma(1) = \tau(1) = 1$, so $(\sigma \circ \tau)(1) = \sigma(\tau(1)) = \sigma(1) = 1$, and so $\sigma \circ \tau \in H$. Clearly the identity map sends 1 to 1, so belongs to H. Finally, if $\sigma \in H$ then $\sigma(1) = 1$, so $\sigma^{-1}(1) = 1$ and $\sigma^{-1} \in H$.

Definition The *order* of a group G is the number of elements in G (finite or infinite). It is denoted $|G|$. Note that $|G| \geq 1$, since every group contains at least one element, namely the identity. The *order* of an element $g \in G$ is the least positive integer k such that g^k is equal to the identity element of G, where g^k denotes $g * g * \ldots * g$ (k times). If no such positive integer k exists, then g has *infinite order*.

Lemma 1.4 *Let G be a group and $g \in G$. Then the set $\langle g \rangle = \{g^k : k \in \mathbb{Z}\}$ is a subgroup of G, and its order is equal to that of g.*

(Here $g^0 = e_G$, g^{-1} is the inverse of g in G, and if $k > 0$ then g^{-k} denotes $(g^{-1})^k$, which is the inverse of g^k.)

Proof. The set $\langle g \rangle$ is clearly nonempty and closed under $*$. It contains $e_G = g^0$, and the inverse g^{-k} of each of its elements g^k. Hence $\langle g \rangle$ is a subgroup of G, by the subgroup test.

Suppose that there are integers $j < k$ with $g^j = g^k$. Then $g^{k-j} = e_G$, so g has finite order $n \leq k - j$. It follows that the n elements $g^0 = e_G, g^1 = g, g^2, \ldots, g^{n-1}$ of G are pairwise distinct elements of $\langle g \rangle$, so $|\langle g \rangle| \geq n$.

On the other hand, every integer k can be expressed in the form $k = an + b$ for some $k \in \mathbb{Z}$ and some $b = 0, 1, \ldots, n-1$, so

$$g^k = (g^n)^a g^b = (e_G)^a g^b = g^b \in \{g^0, g^1, \ldots, g^{n-1}\},$$

and $|\langle g \rangle| \leq n$.

Hence $|\langle g \rangle| = n$.

If $g^j \neq g^k$ whenever $j \neq k$, then in particular $g^k \neq e_G$ whenever $k > 0$, so g has infinite order. In this case $\langle g \rangle$ also has infinite order, since the elements g^k, $k \in \mathbb{Z}$, are pairwise distinct.

Theorem 1.5 (Lagrange's Theorem) *Let G be a finite group and H a subgroup of G. Then the order of H divides that of G.*

Corollary 1.6 *Let G be a group and $g \in G$. Then the order of g divides that of G.*

Definition Let H be a subgroup of G, and $g \in G$. Then the *left coset* of H in G represented by g is the subset $gH = \{gh,\ h \in H\}$ of G, and the *right coset* of H in G represented by g is the subset $Hg = \{hg,\ h \in H\}$ of G.

Exercise Show that:

1. the left cosets form a partition of G;
2. the map $h \mapsto gh$ is a bijection from H to gH.

Hence prove Lagrange's Theorem.

Definition A subgroup $N \subset G$ is *normal* if each left coset is also a right coset. (That is $gN = Ng$ for all $g \in G$.) Equivalently, $gng^{-1} \in N$ for all $g \in G$ and all $n \in N$.

1.6 Homomorphisms

Definition Let $(G, *)$ and $(H, \dagger)$ be groups. A map $f : G \to H$ is a *homomorphism* if

$$f(x * y) = f(x) \dagger f(y) \qquad \forall\ x, y \in G.$$

Example The exponential map from $(\mathbb{R}, +)$ to $(\mathbb{R}_+, \times)$ is a homomorphism (since $\exp(x + y) = \exp(x) \exp(y)$).

Lemma 1.7 *Let $f : G \to H$ be a homomorphism. Then the* image *of f,*

$$\mathrm{Im}(f) := \{f(x),\ x \in G\} \subset H,$$

is a subgroup of H, and the kernel *of f,*

$$\mathrm{Ker}(f) := \{x \in G,\ f(x) = e_H\} \subset G,$$

is a normal subgroup of G.

Proof. I will prove the first part, and leave the second as an exercise – see the example sheet at the end of this chapter.

We use the subgroup test to check that $I := \mathrm{Im}(f)$ is a subgroup of H.

Firstly, the closure property. If $x, y \in I$, then there are elements $a, b \in G$ such that $f(a) = x$ and $f(b) = y$ (by definition of $\mathrm{Im}(f)$). Then, since f is a homomorphism, we have

$$x \dagger y = f(a) \dagger f(b) = f(a * b) \in \mathrm{Im}(f) = I.$$

Next, the identity property. Let e_H be the identity element of H, and e_G the identity element of G. Then

$$f(e_G) \dagger f(e_G) = f(e_G * e_G) = f(e_G),$$

so

$$f(e_G) = \left(f(e_G)^{-1} \dagger f(e_G)\right) \dagger f(e_G) = f(e_G)^{-1} \dagger \left(f(e_G) \dagger f(e_G)\right) = f(e_G)^{-1} \dagger f(e_G) = e_H.$$

In particular, $e_H = f(e_G) \in \mathrm{Im}(f) = I$.

Finally, the inverse property. If $x = f(a) \in \mathrm{Im}(f) = I$, then

$$x \dagger f(a^{-1}) = f(a) \dagger f(a^{-1}) = f(a * a^{-1}) = f(e_G) = e_H,$$

, so $x^{-1} = f(a^{-1}) \in \mathrm{Im}(f) = I$.

Hence $\mathrm{Im}(f)$ is a subgroup of H, as claimed.

Clearly, a homomorphism $f : G \to H$ is surjective iff $\mathrm{Im}(f) = H$. Less obviously (but easily checked), it is injective iff $\mathrm{Ker}(f) = \{e_G\}$, the trivial subgroup of G. A homomorphism that is both injective and surjective (ie bijective) is called an *isomorphism*, and two groups G, H are *isomorphic* (denoted $G \cong H$) if there is an isomorphism $G \to H$.
(Exercise: understand why this is the same thing as the less formal definition of isomorphic groups given earlier.)
(Another exercise: check that $\cong$ is an equivalence relation between groups.)

Example The exponential map from $(\mathbb{R}, +)$ to $(\mathbb{R}_+, \times)$ is an isomorphism. It is a homomorphism, and bijective (with inverse $\ln : \mathbb{R}_+ \to \mathbb{R}$). Hence the groups $(\mathbb{R}, +)$ and $(\mathbb{R}_+, \times)$ are isomorphic.

1.7 Quotient Groups

Let G be a group, and N a normal subgroup. The *quotient group* (or *factor group*) of G by N, denoted G/N or $\frac{G}{N}$, is defined to be the set of left cosets gN for all $g \in G$. (Since N is normal, this is *the same* as the set of right cosets Ng.) The binary operation on G/N is defined by

$$(xN)(yN) := (xy)N \quad \forall\, x, y \in G.$$

Of course, one needs to check some things - firstly that the definition does not depend on the choices x, y of representatives of the two cosets, and then that the resulting binary operation on G/N satisfies all the axioms for a group.

Let us check the first of these. The second will be an exercise: see the example sheet at the end of the chapter.

Suppose that x', y' are different choices of coset representatives. In other words, $x'N = xN$ and $y'N = yN$. Then $x' = xn_1$ and $y' = yn_2$ for some $n_1, n_2 \in N$. We must show that $(x'y')N = (xy)N$.

Now

$$(x'y')N = xn_1yn_2N = xn_1(yN) = xn_1(Ny) = x(n_1N)y = x(Ny) = x(yN) = (xy)N.$$

More specifically, since $n_1y \in Ny = yN$, there is an element $n_3 \in N$ with $n_1y = yn_3$. Then

$$xn_1yn_2 = (xy)(n_3n_2) \in (xy)N.$$

Example The set $2\mathbb{Z}$ of even integers is a normal subgroup of the group $(\mathbb{Z}, +)$. The quotient group $\mathbb{Z}/2\mathbb{Z}$ has two elements: $0 + 2\mathbb{Z} = 2\mathbb{Z}$ (the set of all even integers and $1 + 2\mathbb{Z}$ (the set of all odd integers). The Cayley table of this group is:

$+$	$0+2\mathbb{Z}$	$1+2\mathbb{Z}$
$0+2\mathbb{Z}$	$0+2\mathbb{Z}$	$1+2\mathbb{Z}$
$1+2\mathbb{Z}$	$1+2\mathbb{Z}$	$0+2\mathbb{Z}$

Note that this is just like addition modulo 2. In fact $\mathbb{Z}/2\mathbb{Z} \cong \mathbb{Z}_2$. This is a special case of an important theorem.

Theorem 1.8 (First Isomorphism Theorem) *Let $f : G \to H$ be a homomorphism. Then*

$$\frac{G}{\mathrm{Ker}(f)} \cong \mathrm{Im}(f).$$

Proof. Let $K = \mathrm{Ker}(f)$ and $I = \mathrm{Im}(f)$. We must define a map $\theta : G/K \to I$ and prove that it is an isomorphism.

There is only one natural way to define θ: namely $\theta(gK) := f(g) \in I$.

First, we should check that this is well-defined. In other words, given a different choice of coset representative $g' \in gK$, the definition gives the same element of I for $\theta(g'K) = \theta(gK)$. We can write $g' = gk$ for some $k \in K$. Then

$$f(g') = f(gk) = f(g)f(k) = f(g)e_H = f(g).$$

Hence θ is indeed well-defined.

Next, we should check that θ is a homomorphism. But

$$\theta(g_1K.g_2K) = \theta((g_1g_2)K) = f(g_1g_2) = f(g_1)f(g_2) = \theta(g_1K)\theta(g_2K).$$

Next, that θ is surjective. But if $x \in I$ then $x = f(a)$ for some $a \in G$. Then $x = \theta(aK)$.

Finally, that θ is injective. Suppose that $\theta(gK) = \theta(g'K)$. Then $f(g) = f(g')$, so

$$f(g^{-1}g') = f(g^{-1})f(g') = f(g)^{-1}f(g) = e_H,$$

so $g^{-1}g' \in \mathrm{Ker}(f) = K$, so $g'K = gK$.

1.8 Finitely generated abelian groups

An abelian group $(A, +)$ is said to be *generated* by a finite set $S = \{s_1, \ldots, s_k\}$ if every element of A can be expressed as $a = n_1s_1 + \cdots + n_ks_k$ for some integers $n_1, \ldots, n_k$. Here, for $n \in \mathbb{Z}$ and $s \in A$, we define $ns = (-n)(-s) \in A$ by induction on $|n|$ by $0s = 0_A$ (the identity element of A, and $(n+1)s = ns + s$ for $n \geq 0$.

Finitely generated abelian groups are completely understood, in the sense that we have the following structure theorem for them.

Theorem 1.9 *If A is a finitely generated abelian group, there are integers $r, s \geq 0$ and $m(1), \ldots, m(s) \geq 2$, which are uniquely determined by A, such that $m(i)$ divides $m(i+1)$ for $1 \leq i \leq s-1$, and*

$$A \cong \mathbb{Z}^r \times \mathbb{Z}_{m(1)} \times \cdots \times \mathbb{Z}_{m(s)}.$$

Corollary 1.10 *If A is a finite abelian group, then there are integers $s \geq 0$ and $m(1), \ldots, m(s) \geq 2$, which are uniquely determined by A, such that $m(i)$ divides $m(i+1)$ for $1 \leq i \leq s-1$, and*

$$A \cong \mathbb{Z}_{m(1)} \times \cdots \times \mathbb{Z}_{m(s)}.$$

Exercises on group theory

1. Define a binary operation $*$ on the set $G = \mathbb{R} \setminus \{0\}$ by $a * b = 5ab$ for all $a, b \in G$. Show that $(G, *)$ is a group, and determine whether or not it is an abelian group.

2. Let $M = M_n(\mathbb{R})$ be the set of 3×3 matrices with real coefficients, and let $S \subset M$ be the set of symmetric matrices (that is, matrices A such that $A = A^T$). Use the subgroup test to show that S is a subgroup of $(M, +)$ (where $+$ denotes matrix addition).

3. Find all subgroups of the group $\mathbb{Z}_8$.

4. Let $\phi: G \to G'$ be a homomorphism of groups. Prove that $\mathrm{Ker}(\phi)$ is a normal subgroup of G.

5. Let H be a normal subgroup of $(G, *)$. Define an operation $\oplus$ on cosets by

$$(a * H) \oplus (b * H) = (a * b) * H$$

   Assuming this is well defined (as shown in the notes) show that this gives $(G/H, \oplus)$ the structure of a group.

6. Describe the elements in each of the following quotient groups (i.e. describe the appropriate cosets).

   (i) $\mathbb{C}/\mathbb{R}$;

   (ii) $\mathbb{Z}/3\mathbb{Z}$;

   Can these quotient groups be described in more familiar terms i.e. are they isomorphic to other groups you know about?

7. Let $S = \{x \in \mathbb{R} : x \neq -1\}$. If the binary operation $*$ is defined by $a*b = a+b+ab$, show that $(S, *)$ is a group. Prove that $(S, *)$ is isomorphic to $(\mathbb{R} \setminus \{0\}, \cdot)$.

8. Let $n \in \mathbb{N}$. Use the First Isomorphism Theorem for Groups to show that $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$. [Hint: Define an appropriate function $\phi: \mathbb{Z} \to \mathbb{Z}_n$].

Chapter 2

Rings, fields and integral domains

2.1 Rings

Definition A *ring* $(R, +, \cdot)$ is a set R together with two binary operations $+$ (called *addition*) and $\cdot$ (called *multiplication*) that satisfy the following axioms.

1. $(R, +)$ is an abelian group. In other words, $+$ is associative and commutative, with an identity (which we denote 0 or 0_R), and each element $x \in R$ has an inverse with respect to $+$ (which we denote $-x$).

2. $\cdot$ is associative. That is, $x(yz) = (xy)z \ \forall \ x, y, z \in R$.

3. $\cdot$ is *left distributive* and *right distributive* over $+$. That is,

   (a) $x(y + z) = (xy) + (xz) \ \forall \ x, y, z \in R$.

   (b) $(y + z)x = (yx) + (zx) \ \forall \ x, y, z \in R$.

Remarks

1. Multiplication $\cdot$ need not be commutative. If it is, then we say that R is a *commutative ring*. (In this case, the left and right distributive conditions are equivalent.)

2. R need not have an identity for $\cdot$. If it does, then we often denote it by 1 or 1_R. We then say that R is a *ring with identity*, or *ring with unity*.

3. If R has an identity 1, then elements of R do not in general have inverses with respect to multiplication. Those elements that do have inverses are called *invertible elements* or *units* of R. If R is a ring with identity, then the units of R form a group $U(R)$.

Examples

1. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are commutative rings with identity, with the usual operations of addition and multiplication.

2. Let $S = \{m + n\sqrt{2};\ m, n \in \mathbb{Z}\} \subset \mathbb{R}$, and let $+, \cdot$ be the usual addition and multiplication of real numbers. Then $(S, +, \cdot)$ is a commutative ring with identity.

   To see this, note first that S is closed with respect to addition and additive inverses: $(m_1 + n_1\sqrt{2}) + (m_2 + n_2\sqrt{2}) = (m_1 + m_2) + (n_1 + n_2)\sqrt{2} \in S$ and $-(m + n\sqrt{2}) = (-m) + (-n)\sqrt{2} \in S$, and contains the identity $0 = 0 + 0\sqrt{2}$ of $(\mathbb{R}, +)$, so is a subgroup of $(\mathbb{R}, +)$ by the subgroup test. Addition in S is commutative, since it is commutative in the larger group $\mathbb{R}$. Hence S is an abelian group.

   Also, S is closed under multiplication, since $(m_1+n_1\sqrt{2})(m_2+n_2\sqrt{2}) = (m_1m_2+2n_1n_2) + (m_1n_2 + m_2n_1)\sqrt{2} \in S$. Hence multiplication in $\mathbb{R}$ gives a binary operation $S \times S \to S$ on S. The associative and distributive properties of $\cdot$ in S hold because they hold in the larger set $\mathbb{R}$.

3. For similar reasons, the set $\mathbb{Z}[i] := \{m+ni;\ m, n \in \mathbb{Z}\} \subset \mathbb{C}$ is a ring with respect to addition and multiplication of complex numbers, where $i \in \mathbb{C}$ denotes a square root of -1.

   The elements of $\mathbb{Z}[i]$ are called *Gaussian integers*.

4. Let $n \geq 2$ be an integer. Then the set $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ is a commutative ring with identity, with respect to addition and multiplication modulo n:

   ab mod n is the integer $r \in \mathbb{Z}_n$ such that $ab = qn + r$ in $\mathbb{Z}$ for some $q \in \mathbb{Z}$.

5. Let $M_n(\mathbb{R})$ denote the set of all $n \times n$ matrices with real entries. Then $M_n(\mathbb{R})$ is a ring with respect to addition and multiplication of matrices. It has an identity, namely the $n \times n$ identity matrix $I = I_n$. It is not commutative when $n \geq 2$, since, for example,

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

   Similarly, the sets $M_n(\mathbb{Z})$, $M_n(\mathbb{Q})$, $M_n(\mathbb{C})$ of $n \times n$ matrices with integer, rational and complex entries, respectively, are noncommutative rings with identity, under addition and multiplication of matrices.

6. Let X be any set, and let $\mathbb{R}^X$ denote the set of all functions $X \to \mathbb{R}$. Define addition and multiplication pointwise on $\mathbb{R}^X$, that is:

$$(f + g)(x) := f(x) + g(x)\ \forall\ x \in X; \quad (fg)(x) := f(x)g(x)\ \forall\ x \in X.$$

Then $\mathbb{R}^X$ is a commutative ring with identity. The additive identity is the constant function 0, and the multiplicative identity is the constant function 1.

7. Define $\mathbb{R}[x]$ to be the set of all *polynomials* with real coefficients in the *variable* x. Elements of $\mathbb{R}[x]$ are formal sums

$$p(x) = \sum_{k=0}^{m} a_k x^k = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0,$$

where $m \in \mathbb{N}$ and $a_0, a_1, \ldots, a_m \in \mathbb{R}$. We can also think of a polynomial $p(x)$ as a function $p : \mathbb{R} \to \mathbb{R}$ defined by the above formula.

There is a natural way to define addition and multiplication on $\mathbb{R}[x]$. As functions, $p(x) + q(x)$ and $p(x)q(x)$ are defined pointwise, as in the previous example. As formal sums, we can define these as follows.

$$\sum_{k=0}^{m} a_k x^k + \sum_{k=0}^{n} b_k x^k = \sum_{k=0}^{N} c_k x^k,$$

where $N = \max(m, n)$, $c_k = a_k + b_k$ for $0 \le k \le \min(m, n)$, $c_k = a_k$ if $n < k \le m$, and $c_k = b_k$ if $m < k \le n$.

$$\left(\sum_{k=0}^{m} a_k x^k \right) \left(\sum_{k=0}^{n} b_k x^k \right) = \sum_{k=0}^{m+n} d_k x^k,$$

where $d_k = \sum \{ a_i b_j ;\ 0 \le i \le m,\ 0 \le j \le n,\ i + j = k \}$.

A polynomial $p(x) = a_0$ (with $m = 0$) is *constant*. The constant polynomial 0 is an additive identity, and the constant polynomial 1 is a multiplicative identity. A nonconstant polynomial can be uniquely written as $p(x) = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0$ with $m > 0$ and $a_m \neq 0$. The integer m is then called the *degree* of the polynomial $p(x)$. Constant polynomials have degree 0.

It is not hard to check that $\mathbb{R}[x]$ is a commutative ring with identity. The units of $\mathbb{R}[x]$ are precisely the nonzero constant polynomials.

In a similar way, we can construct rings $\mathbb{Z}[x]$, $\mathbb{Q}[x]$, $\mathbb{C}[x]$ of polynomials in one variable x with coefficients from $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{C}$ respectively.

2.2 Product rings

Suppose that R, S are two rings. We consider the set $R \times S$ of ordered pairs (r, s) with $r \in R$ and $s \in S$, and define addition and multiplication on this set coordinatewise:

$$(r_1, s_1) + (r_2, s_2) = (r_1 + r_2, s_1 + s_2), \quad (r_1, s_1)(r_2, s_2) = (r_1 r_2, s_1 s_2),$$

where the operations in the first coordinate position are taking place in R and those in the second coordinate position are taking place in S.

It is easy to check that this makes $R \times S$ into a ring. The additive identity is $(0_R, 0_S)$.

If $(a, b) \in R \times S$ is a multiplicative identity, than for any $(r, s) \in R \times S$ we have

$$(ar, bs) = (a, b)(r, s) = (r, s) = (r, s)(a, b) = (ra, sb),$$

so $ar = r = ra$ in R and $bs = s = sb$ in S. In other words, a is a multiplicative identity in R and b is a multiplicative identity in S. Conversely, if each ring R, S has an identity, then $R \times S$ has an identity $(1_R, 1_S)$.

2.3 Some elementary properties

The following result lists some easy facts which are true in any ring.

Lemma 2.1 *Let R be a ring, and let $a, b, c \in R$. Then*

1. *if $a + b = a + c$, then $b = c$;*
2. *$a.0 = 0 = 0.a$;*
3. *$(-a)b = -(ab) = a(-b)$;*
4. *$(-a)(-b) = ab$.*

Proof.

1. This follows from the fact that $(R, +)$ is a group. In detail:

   $$b = 0 + b = (-a + a) + b = -a + (a + b) = -a + (a + c) = (-a + a) + c = 0 + c = c.$$

2. $0 = 0 + 0$, so

   $$0 + 0a = 0a = (0 + 0)a = 0a + 0a.$$

   By the previous property, $0 = 0a$. Similarly $0 = a0$.

3. $ab + (-a)b = (a - a)b = 0b = 0$, so $(-a)b = -(ab)$. Similarly $a(-b) = -(ab)$.

4. By the previous property, $ab = -(-ab) = -(a(-b)) = (-a)(-b)$.

Remark The first property in the above lemma says that addition is *cancellative* – we can cancel an equal term from each side of an equation involving addition. The corresponding property is not true in general for multiplication. In other words, it is possible that $ab = ac$ but $b \neq c$. Easy examples of this come from the second property by taking $a = 0$: $0b = 0 = 0c$ for all $b, c \in R$. But there are also more subtle examples of this phenomenon, such as $a = 2$, $b = 1$, $c = 7$ in $\mathbb{Z}_{12}$: $ab = 2 = ac$ in $\mathbb{Z}_{12}$, but $b \neq c$ in $\mathbb{Z}_{12}$.

2.4 Subrings

Just as we can find many examples of groups as subgroups of other groups, many rings naturally exist as *subrings* of other rings.

Definition A *subring* of a ring R is a subset S of R which is itself a ring with respect to the same addition and multiplication as R.

Examples

1. Each of $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$.

2. $\mathbb{R}[x]$ can be thought of as a subring of $\mathbb{R}^{\mathbb{R}}$ (the ring of all functions $\mathbb{R} \to \mathbb{R}$ with pointwise addition and multiplication). Here we think of polynomials as functions $\mathbb{R} \to \mathbb{R}$ and note that addition and multiplication are defined pointwise.

3. $M_n(\mathbb{Z})$ is a subring of $M_n(\mathbb{R})$, etc.

There is an easy test for a subset of a ring to be a subring, analogous to the subgroup test in group theory.

Lemma 2.2 (The Subring Test) *A subset S of a ring R is a subring if and only if*

1. *S is a subgroup of $(R, +)$ (which can be checked using the subgroup test);*

2. *S is closed under multiplication ($x, y \in S \Rightarrow xy \in S$).*

Proof. Suppose that S is a subring of R. Then it is a ring with respect to the addition and multiplication of R. In particular it is a group with respect to the addition of R, in other words a subgroup of $(R, +)$. It must also be closed with respect to the multiplication of R, since multiplication on R gives a binary operation on S.

Conversely, suppose that the two conditions of the lemma hold. In particular, the addition and multiplication of R give binary operations on S, and $(S, +)$ is a group, being a subgroup of $(R, +)$. Indeed, since $(R, +)$ is an abelian group, so is any of its subgroups, so $(S, +)$ is an abelian group. Multiplication on S is associative, since it is associative on the larger set R. For the same reason, multiplication on S is both left and right distributive over $+$. Hence $(S, +, \cdot)$ is a ring, so S is a subring of R.

Remark If R is a commutative ring, then every subring of R is also commutative. If R has identity, then a subring of R may have identity, but need not.

Examples

1. The set $2\mathbb{Z}$ of all *even* integers is a subring of $\mathbb{Z}$. It is a subgroup of $(\mathbb{Z}, +)$ which is closed with respect to multiplication. This subring is of course commutative, but has no identity (since, for example, $2n \neq 2$ for any even integer n).

2. The set S of 2×2 matrices

$$S = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} : \ a, b, d \in \mathbb{R} \right\}$$

is a subring of $M_2(\mathbb{R})$. Clearly S is a subgroup of $(M_2(\mathbb{R}), +)$, and it is easy to check that the product of two matrices in S also belongs to S. This ring S is not commutative. (See this by writing down a formula for the product of two matrices in S, for example.) It contains the identity matrix I_2, which is clearly an identity for S also.

3. The set T of 2×2 matrices

$$S = \left\{ \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} : \ a \in \mathbb{R} \right\}$$

is a commutative subring of $M_2(\mathbb{R})$. It does not contain the identity matrix of $M_2(\mathbb{R})$. Nevertheless, the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

belongs to T and acts as an identity for T.

2.5 Fields

A ring is an algebraic object in which we can add, subtract, and multiply, but not necessarily divide (since elements do not need to have inverses).

In many of our nicest rings (such as $\mathbb{Q}$ or $\mathbb{C}$, we can divide by any element *except* 0, since every nonzero element has an inverse. Rings like this are called *fields*.

Definition A *field* is a commutative ring with identity element $1 \neq 0$, in which every nonzero element $x \neq 0$ has an inverse x^{-1}.

Remark It is important to allow 0 to be an exceptional element with no inverse. Suppose that R is a ring with identity, in which 0 has an inverse. Then $1 = 0.0^{-1} = 0$, and so, for each $r \in R$, we have $r = 1.r = 0.r = 0$. In other words, $R = \{0\}$, which is a ring but not a very interesting one.

Examples

1. $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are fields with respect to the usual addition and multiplication.

2. The subring $\mathbb{Q}[i] := \{a + bi \in \mathbb{C}; \ a, b \in \mathbb{Q}\}$ of $\mathbb{C}$ is a field, called the *field of Gaussian rationals*.

To check that it is a subring, we use the subring test: $\mathbb{Q}[i]$ is closed under addition and multiplication ($(a+bi)+(c+di) = (a+c)+(b+d)i \in \mathbb{Q}[i]$, $(a+bi)(c+di) = (ac-bd)+(ad+bc)i \in \mathbb{Q}[i]$); contains the zero element $0+0i$, and the additive inverse $-a-bi$ of each of its elements $a+bi$.

$\mathbb{Q}[i]$ is certainly commutative, being a subring of $\mathbb{C}$, and it contains the identity $1 = 1+0i$ of $\mathbb{C}$. Finally, if $a+bi \neq 0$ in $\mathbb{Q}[i]$, then its multiplicative inverse is $\dfrac{a-bi}{a^2+b^2} \in \mathbb{Q}[i]$.

3. $\mathbb{Z}_p$ is a field if p is a prime number.

   We already know that $\mathbb{Z}_p$ is a commutative ring with identity, so we need only check that every nonzero element has an inverse. Suppose $x \neq 0$ in $\mathbb{Z}_p$. Consider the map $f : \mathbb{Z}_p \to \mathbb{Z}_p$ given by $f(k) = xk \bmod p$. If $f(j) = f(k)$ for some j, k with $0 \leq j < k < p$, then $x(k-j) = xk - xj = f(k) - f(j) = 0$ in $\mathbb{Z}_p$, so the integer $x(k-j)$ is divisible by the prime number p, and hence one of $x, k-j$ is divisible by p (since p is a prime). But this is a contradiction, since $0 < x < p$ and $0 < k-j < p$ by hypothesis.

   Hence f is injective, and so by the pigeonhole principle it is also surjective. Hence there is a (unique) $y \in \mathbb{Z}_p$ with $xy = f(y) = 1$.

4. Let F denote the set $\{0, 1, x, y\}$ of four elements, and let addition and multiplication on F be defined by the Cayley tables:

$+$	0	1	x	y
0	0	1	x	y
1	1	0	y	x
x	x	y	0	1
y	y	x	1	0

$\times$	0	1	x	y
0	0	0	0	0
1	0	1	x	y
x	0	x	y	1
y	0	y	1	x

   Then $(F, +, \times)$ is a field. (**Exercise**: check the field axioms for F.)

5. Let $\mathbb{H}$ denote the subset of $M_2(\mathbb{C})$ consisting of matrices $A = \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix}$ where $a, b \in \mathbb{C}$ and $\bar{\cdot}$ denotes the complex conjugate. Then $\mathbb{H}$ is a subring of $M_2(\mathbb{C})$ which contains the identity matrix I_2. Moreover, if $A \neq 0$ in $\mathbb{H}$, then $\det(A) = |a|^2 + |b|^2 > 0$, so A is invertible, and its inverse $A^{-1} = \dfrac{1}{\det(A)} \begin{pmatrix} \overline{a} & -b \\ \overline{b} & a \end{pmatrix} \in \mathbb{H}$. Hence every nonzero element of $\mathbb{H}$ has an inverse.

   However, $\mathbb{H}$ is not a field, since it is not commutative.

   The ring $\mathbb{H}$ is known as the *Quaternions*, or *Hamiltonians*. It is an example of a *division-ring*, or *skew-field* – a ring which is not necessarily commutative, in which every nonzero element is a unit.

Remark In a field F, the group of units is $U(F) = F \setminus \{0\}$. When F is finite, the Cayley table for $U(F)$ can be found from the multiplication table for F by stripping off the row and column marked 0.

For the 4-element field in the example above, the unit group is clearly isomorphic to the cyclic group of order 3. For the field $\mathbb{Z}_5$ we have, for example, $2^2 = 4$, $2^3 = 3$, $2^4 = 1$, so 2 has order 4 in $U(\mathbb{Z}_5)$, and $U(\mathbb{Z}_5)$ is isomorphic to the cyclic group of order 4.

2.6 Integral domains

Definition A *zero-divisor*, or *divisor of zero* in a ring R is an element $x \in R$ such that $x \neq 0$, but $xy = 0$ in R for some $y \in R$ with $y \neq 0$.

Example. In $\mathbb{Z}_6$ we have $2 \times 3 = 0 = 3 \times 2$, so 2 and 3 are zero-divisors.

In a field F, there are no zero-divisors. For example, if $xy = 0$ with $x \neq 0$, then $y = 1y = (x^{-1}x)y = x^{-1}(xy) = x^{-1}0 = 0$.

It follows that the same property holds for subrings of fields. If F is a field, and R a subring of F, then any zero-divisor in R would also be a zero-divisor in F.

In particular, $\mathbb{Z}$ has no zero-divisors. The notion of *integral domain* is intended to mean a ring with properties resembling those of $\mathbb{Z}$.

Definition An *integral domain* is a commutative ring with identity, in which there are no zero-divisors.

Examples

1. Any field is an integral domain. In particular, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are integral domains, as is $\mathbb{Z}_p$ for any prime number p.

2. Any subring of a field, which contains the identity, is an integral domain. In particular the ring of integers $\mathbb{Z}$ and the ring of Gaussian integers $\mathbb{Z}[i]$ are integral domains.

3. If R is an integral domain, then so is the ring $R[x]$ of polynomials with coefficients in R. Certainly $R[x]$ is a commutative ring with identity. To see that $R[x]$ has no zero-divisors, suppose that $p(x) = a_m x^m + \ldots$ and $q(x) = b_n x^n + \ldots$ are nonzero polynomials in $R[x]$ of degrees m, n respectively. Then $a_m \neq 0 \neq b_n$ in R. Since R is an integral domain, $a_m b_n \neq 0$, so

$$p(x)q(x) = a_m b_n x^{m+n} + \ldots$$

   is a nonzero polynomial of degree $m + n$ in $R[x]$.

A useful property of integral domains is that multiplication by nonzero elements is cancellative.

Lemma 2.3 *Let R be an integral domain, and $a, b, c \in R$ such that $ac = bc$ and $c \neq 0$. Then $a = b$.*

Proof. By hypothesis, $(a - b)c = ac - bc = 0$. But $c \neq 0$ and R has no zero divisors, so $a - b = 0$. Thus $a = b$ as claimed.

The following result is a generalisation of the fact that $\mathbb{Z}_p$ is a field for any prime p. It is proved in exactly the same way.

Theorem 2.4 *Every* finite *integral domain is a field.*

Proof. Let R be a finite integral domain. Then R is a commutative ring with identity. To prove that R is a field, we need only check that every nonzero element has an inverse.

Suppose that $x \in R$ with $x \neq 0$. Then, by the Lemma above, $xy \neq xz$ whenever $y \neq z$ in R. Thus the map $f : R \to R$ defined by $f(y) = xy$ is injective. Since R is finite, the pigeonhole principle implies that f is also surjective. In particular, $\exists\ y \in R$ with $xy = f(y) = 1_R$. Hence x has an inverse, as required.

Exercises on rings, fields and integral domains

1. Which of the following are rings?

   (a) $(S, +, \cdot)$, where $S = \{2k + 1 : k \in \mathbb{Z}\} \cup \{0\}$ and $+$ and $\cdot$ denote the usual addition and multiplication of real numbers.

   (b) $(S, +, \cdot)$, where S denotes the family of all functions from $\mathbb{R}$ to $\mathbb{R}$, and $+$ and $\cdot$ denote the usual (pointwise) addition and multiplication of functions: $(f + g)(x) = f(x) + g(x)$, $(fg)(x) = f(x)g(x)$.

   (c) $(S, +, \circ)$, where S denotes the family of all functions from $\mathbb{R}$ to $\mathbb{R}$, and $+$ denotes pointwise addition and $\circ$ denotes composition of functions.

   (d) $(S, +, \cdot)$, where S is the family of all subsets of a given set E and $+$ and $\cdot$ are defined by

   $$A + B = (A \setminus B) \cup (B \setminus A)\,; \qquad A \cdot B = A \cap B\,.$$

2. Find all the units in the rings $\mathbb{Z}_4$, $\mathbb{Z}_6$, and $\mathbb{Z}_3 \times \mathbb{Z}_3$.

3. In the ring $\mathbb{Z}_{48}$, find all elements x satisfying $x^2 = 0$, and all elements y satisfying $y^3 = 0$.

4. In the ring $\mathbb{Z}_7 \times \mathbb{Z}_7$, find all elements x satisfying $x^2 = (-1, 1)$, and all elements y satisfying $y^3 = (1, 0)$.

5. Let $(R, +, \cdot)$ be a ring such that $x^2 = x$ for all $x \in R$. Show that for all $x, y \in R$, $xy = -yx$. [Hint: consider $(x + y)^2$.] Deduce that R is commutative.

6. Let S denote the set of all rational numbers of the form $a/2^n$, where $a, n \in \mathbb{Z}$ and $n \geq 0$. Show that S is a subring of $\mathbb{Q}$.

7. Show that the set R of complex numbers of the form $a + 2bi$, with $a, b \in \mathbb{Z}$, is a subring of $\mathbb{C}$.

8. Which of the following are subrings of $\mathbb{C}$?

   (a) $\{z \in \mathbb{C} : \ |z| = 1\}$.

   (b) $\{0 + iy;\ y \in \mathbb{R}\}$.

   (c) $\{(a + ib\sqrt{3});\ a, b \in \mathbb{Z}\}$.

   (d) $\{(a + ib\sqrt{3})/2;\ a, b \in \mathbb{Z},\ a + b \text{ even}\}$.

9. Show that $(\mathbb{R}, \oplus, \otimes)$ is a field where

   $$a \oplus b = a + b + 1\,; \qquad a \otimes b = a + b + ab\,.$$

10. Let $(F, +, \cdot)$ be a field and let $x, y \in F$. Is it necessarily true that $x^3 = y^3$ implies $x = y$?

11. Let p be a prime number. Find all elements in the field $(\mathbb{Z}_p, +, \cdot)$ which are their own multiplicative inverses.

12. Let R and S be two rings, each with more than one element. Show that there are nonzero elements x and y in $R \times S$ such that $xy = 0$. In the case where $R = \mathbb{Z}_2$ and $S = \mathbb{Z}_3$, find all such pairs $x, y \in R \times S$.

13. Determine the units and the divisors of zero of $(\mathbb{Z}_{12}, +, \cdot)$. Write down the group table of the group of units of $\mathbb{Z}_{12}$.

14. Show that $\mathbb{Z} \times \mathbb{Z}$ is not an integral domain.

15. Show that a subring $R \neq \{0\}$ of a field F is an integral domain if and only if $1_F \in R$.

16. Let $(R, +, \cdot)$ be a commutative ring with identity. Prove that R is an integral domain if for $a, b, c \in R$ with $a \neq 0$ the relation $ab = ac$ implies that $b = c$.

Chapter 3

Homomorphisms, ideals, and quotient rings

In group theory, we construct the quotient group G/N of a group G over a normal subgroup N. The normal subgroups are precisely the kernels of homomorphisms between groups. In this chapter we follow the entirely analogous story in ring theory. As in group theory, a ring homomorphism is a map which respects the binary operations. The analogue of a normal subgroup is something called an *ideal* in a ring. Given an ideal, we can construct a quotient ring.

3.1 Homomorphisms

Definition A *homomorphism* from a ring R to a ring S is a map $f : R \to S$ such that

$$f(x+y) = f(x) + f(y) \text{ and } f(xy) = f(x)f(y) \ \forall \ x, y \in R.$$

(In each case the operation of $+$ or $\cdot$ on the left side of the equation takes place in R, while that on the right side takes place in S.)

If the homomorphism $f : R \to S$ is bijective, then it is called an *isomorphism*, and we say that the rings R and S are *isomorphic* (denoted $R \cong S$). The relation $\cong$ is an equivalence relation between rings. We regard isomorphic rings as being 'the same'.

If $f : R \to S$ is a homomorphism of rings, then in particular f is a homomorphism of groups from $(R, +)$ to $(S, +)$. From this we have an immediate list of properties of f:

1. $f(0_R) = 0_S$;
2. $f(-x) = -f(x) \ \forall \ x \in R$;
3. $\mathrm{Im}(f)$ is a subgroup of $(S, +)$;

4. $\mathrm{Ker}(f)$ is a subgroup of $(R,+)$. (Indeed, $\mathrm{Ker}(f)$ is a normal subgroup of $(R,+)$. But in any case, since $(R,+)$ is an abelian group, *all* its subgroups are normal.)

It is not true in general that $f(1_R) = 1_S$, even when both rings R, S have identities. For example, the map $f : \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ given by $f(r) = (r, 0)$ is a homomorphism, but $f(1) \neq (1, 1)$. On the other hand, suppose that $f : R \to S$ is an isomorphism Then S has an identity if and only if R has an identity, and in this case $f(1_R) = 1_S$. To see this, suppose that R has an identity, and let $s \in S$. Then

$$f(1_R)s = f(1_R)f(f^{-1}(s)) = f(1_R f^{-1}(s)) = f(f^{-1}(s)) = s$$

and similarly $sf(1_R) = s$, so $f(1_R)$ is an identity for S. Applying the same argument to the isomorphism $f^{-1} : S \to R$ gives the converse.

Example The rings $\mathbb{R}$ and $\mathbb{C}$ are not isomorphic. The reason is that $\mathbb{C}$ contains a square root of -1, but $\mathbb{R}$ does not. In detail, suppose that $f : \mathbb{C} \to \mathbb{R}$ is an isomorphism. Then $f(0) = 0$, so $f(1) \neq 0$, since f is injective. But $f(1) = f(1^2) = f(1)^2$, so $f(1) \in \mathbb{R}$ is a nonzero solution of the equation $x^2 = x$. There is only one such solution, namely $x = 1$, so $f(1) = 1$, and hence also $f(-1) = -f(1) = -1$. Finally, $f(i)^2 = f(i^2) = f(-1) = -1$, so $f(i)$ is a square root of -1 in $\mathbb{R}$, a contradiction.

Remark This last example depends in a crucial way on the difference between the multiplicative structures of $\mathbb{R}$ and $\mathbb{C}$. Indeed, as groups, $(\mathbb{R},+)$ and $(\mathbb{C},+)$ are isomorphic. (They are vector spaces of equal infinite dimension over $\mathbb{Q}$.)

Theorem 3.1 *If $f : R \to S$ is a ring homomorphism, then* $\mathrm{Im}(f) = \{s \in S : \exists\ r \in R \text{ with } f(r) = s\}$ *is a subring of S.*

Proof. We have already noted that $\mathrm{Im}(f)$ is a subgroup of $(S,+)$ (since f is a group homomorphism from $(R,+)$ to $(S,+)$).

It only remains to show that $\mathrm{Im}(f)$ is closed under multiplication. So suppose that $x, y \in \mathrm{Im}(f)$. Then there are elements $a, b \in R$ such that $f(a) = x$ and $f(b) = y$ (by definition of $\mathrm{Im}(f)$). Since f is a homomorphism, we have

$$xy = f(a)f(b) = f(ab) \in \mathrm{Im}(f).$$

Examples

1. The map $f : \mathbb{Z} \to \mathbb{Z}_n$ given by $f(k) = k \bmod n$ is a homomorphism, by the definition of addition and multiplication in $\mathbb{Z}_n$.

2. The map $f : \mathbb{C} \to \mathbb{C}$ given by $f(z) = \overline{z}$ (the complex conjugate of z), is a homomorphism, since $\overline{(w+z)} = \overline{w} + \overline{z}$ and $\overline{(wz)} = \overline{w} \cdot \overline{z}$ for all $w, z \in \mathbb{C}$. Moreover, since f is bijective, it is an isomorphism.

3. The map $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = -x$ is not a ring homomorphism. It is true that $f(x + y) = f(x) + f(y)$, so f is a group homomorphism from $(\mathbb{R}, +)$ to $(\mathbb{R}, +)$. But $f(xy) = -xy$, while $f(x)f(y) = (-x)(-y) = xy$, so if $x \neq 0 \neq y$ then $f(xy) \neq f(x)f(y)$.

4. If R is a ring with identity $1_R \in R$, we can define $2_R := 1_R+1_R$, $3_R := 1_R+1_R+1_R$, etc. (Inductively, if $n \in \mathbb{N}$ and we have defined $n_R \in R$, then we define $(n+1)_R := n_R + 1_R \in R$.) We can also define $(-n)_R := -(n_R)$. Then $f : \mathbb{Z} \to R$, $f(n) = n_R$, is a ring homomorphism: $(m + n)_R = m_R + n_R$, $(mn)_R = m_R n_R$. (These can easily be checked by induction on $|m| + |n|$.)

5. If R is a ring with identity $1_R \in R$, and $r \in R$, then we can extend the homomorphism $\mathbb{Z} \to R$, $n \mapsto n_R$, to a homomorphism $\phi_r : \mathbb{Z}[x] \to R$, called the *evaluation* homomorphism at r, by $\phi_r(p(x)) = p(r)$. Here we have to interpret integers as belonging to R via $n \mapsto n_R$, substitute $x = r$ in the polynomial $p(x)$, and evaluate the result in R. For example, if $p(x) = x^2 + 3x - 4$, $R = \mathbb{C}$, and $r = i = \sqrt{-1}$, then $\phi_r(p(x)) = p(i) = i^2 + 3i - 4 = -5 + 3i \in \mathbb{C}$.

3.2 Ideals

In this section we study ideals, which are the analogues of normal subgroups in group theory. Thus ideals should be the objects that occur as kernels of ring homomorphisms.

Suppose then that R, S are rings, and that $f : R \to S$ is a ring homomorphism. We know that $\mathrm{Ker}(f)$ is a subgroup of $(R, +)$. It is not difficult to see that in fact $\mathrm{Ker}(f)$ is closed with respect to multiplication, so is a subring of R:

$$x, y \in \mathrm{Ker}(f) \Rightarrow f(xy) = f(x)f(y) = 0_S 0_S = 0_S \Rightarrow xy \in \mathrm{Ker}(f).$$

However, a stronger property holds. In the above equation, in order for $f(x)f(y)$ to be 0 in S, we do not need both $f(x)$, $f(y)$ to be 0. It is sufficient for any one of them to be 0 in S.

This suggests the following definition.

Definition An *ideal* in a ring R is a subset $I \subset R$ such that

1. I is a subgroup of $(R, +)$

2. $(\forall\ x \in I)\ (\forall\ r \in R)\ xr \in I$ and $rx \in I$.

Thus to check that a given subset $I \subset R$ is an ideal of R, we check the above two properties. For the first, we can use the subgroup test, or we may already know for other reasons that I is a subgroup of $(R, +)$. The key property to check is usually the second property $rx, xr \in I$, which I will refer to as the *ideal property*. In practice, most

of the rings that we consider will be commutative, in which case the two statements $xr \in I$ and $rx \in I$ are equivalent (since $rx = xr$), so we need only check one of them.

Examples

1. For $n \in \mathbb{N}$, the set $n\mathbb{Z} = \{nk;\ k \in \mathbb{Z}\}$ is an ideal of $\mathbb{Z}$.

   We already know that $n\mathbb{Z}$ is a subgroup of $(\mathbb{Z}, +)$. To check the ideal property, suppose that $x = nk \in n\mathbb{Z}$ and $r \in \mathbb{Z}$. Then $rx = xr = (nk)r = n(kr) \in n\mathbb{Z}$.

2. The set $I = \{p(x) \in \mathbb{R}[x];\ p(0) = 0\}$ of polynomials with constant term 0 is an ideal in $\mathbb{R}[x]$.

   We first check that I is a subgroup of $(\mathbb{R}[x], +)$, using the subgroup test. If $p(x), q(x) \in I$ then $p(0)+q(0) = 0+0 = 0$, so $p(x)+q(x) \in I$; and $-p(0) = -0 = 0$ so $-p(x) \in I$. Clearly $0 \in I$, so I is a subgroup of $(\mathbb{R}[x], +)$, as required.

   Now if $p(x) \in I$ and $r(x) \in \mathbb{R}[x]$, then $p(0)r(0) = 0r(0) = 0$, so $p(x)r(x) \in I$. Since $\mathbb{R}[x]$ is commutative, we also have $r(x)p(x) = p(x)r(x) \in I$.

3. In any ring R, the sets $\{0\}$ and R are ideals in R.

   These are clearly subgroups of $(R, +)$. The ideal property $rx, xr \in I$ for all $x \in I$ and $r \in R$ is trivially true for $I = R$, and for $I = \{0\}$ it reduces to the property we observed earlier, that $0r = 0 = r0$ for all $r \in R$.

4. If $f : R \to S$ is a ring homomorphism, then $\mathrm{Ker}(f)$ is an ideal in R.

   Since f is a group isomorphism from $(R, +)$ to $(S, +)$, $\mathrm{Ker}(f)$ is a subgroup of $(R, +)$. For the ideal property, suppose that $x \in \mathrm{Ker}(f)$ and $r \in R$. Then

   $$f(rx) = f(r)f(x) = f(r)0 = 0 = 0f(r) = f(x)f(r) = f(xr),$$

   so $rx, xr \in \mathrm{Ker}(f)$, as required.

5. If R is a commutative ring, and $x \in R$, then the set $xR = \{xr;\ r \in R\}$ is an ideal in R, called the *principal ideal* generated by x.

   This generalises the example of $n\mathbb{Z} \subset \mathbb{Z}$ above, and can be checked in a similar way. Specifically, we show that xR is a subgroup of $(R, +)$ using the subgroup test: $xr + xs = x(r + s) \in xR$ by the distributive law, $0 = x0 \in xR$ and $-xr = x(-r) \in xR$. For the ideal property, if $r, s \in R$ then $s(xr) = (xr)s = x(rs) \in xR$.

6. If R and S are rings, I is an ideal in R and J is an ideal in S, then $I \times J$ is an ideal in $R \times S$.

   Suppose that $w, x \in I$, $y, z \in J$, $r \in R$ and $s \in S$. Then $(w, y) + (x, z) = (w + x, y + z) \in I \times J$, $(0, 0) \in I \times J$, $-(x, z) = (-x, -z) \in I \times J$, $(r, s)(x, z) = (rx, sz) \in I \times J$, and $(x, z)(r, s) = (xr, zs) \in I \times J$.

7. If F is a field, then the only ideals in F are $\{0\}$ and F.

   Suppose that $I \neq \{0\}$ is an ideal in F. Since $0 \in I$, there must be a nonzero element $x \neq 0$ in I. If $\overline{x}$ is the inverse of x in F, and $y \in F$, then $y = 1y = (x\overline{x})y = x(\overline{x}y) \in I$. Hence $I = F$.

3.3 Quotient Rings

Let R be a ring and I an ideal of R. Then in particular I is a subgroup of $(R, +)$. Since $+$ is commutative, I is in fact a normal subgroup of $(R, +)$, so we can form the quotient group R/I. The elements of R/I are the cosets $r + I$ for $r \in R$. We will denote the binary operation in this group as $+$, and we recall that this is defined by the rule $(r + I) + (s + I) := (r + s) + I$.

We would like to make $(R/I, +)$ into a ring, so we need to define a multiplication on R/I. We do so in the obvious way, namely: $(r + I) \times (s + I) := (rs) + I$.

Lemma 3.2 *With $+$ and $\times$ defined as above, $(R/I, +, \times)$ is a ring.*

Proof. We already know that $(R/I, +)$ is a group. Indeed, it is an abelian group since $+$ is commutative in R:

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

We need to check that $\times$ is well-defined on R/I. Suppose that $r' \in r + I$ and $s' \in s + I$. Say $r' = r + x$, $s' = s + y$, with $x, y \in I$. Then $r's' = (r + x)(s + y) = rs + z$, where $z = xs + ry + xy$. But $xs, ry, xy \in I$ since $x, y \in I$ and $r, s \in R$, by the definition of ideal. Hence also $z = xs + ry + xy \in I$, by the definition of ideal. Thus $(r's') + I = (rs) + I$, so $\times$ is indeed well-defined on R/I.

We need to check that $\times$ is associative on R/I. But

$$(r + I)[(s + I)(t + I)] = (r + I)(st + I) = r(st) + I$$

$$= (rs)t + I = (rs + I)(t + I) = [(r + I)(s + I)](t + I),$$

using the fact that multiplication in R is associative.

We need to check that $\times$ is distributive over $+$ in R/I. But

$$(r + I)[(s + I) + (t + I)] = (r + I)(s + t + I) = r(s + t) + I = (rs + rt) + I$$

$$= (rs + I) + (rt + I) = (r + I)(s + I) + (r + I)(t + I),$$

using the fact that multiplication in R is (left) distributive over $+$. Thus $\times$ is left distributive over $+$ in R/I. The right distributive property is verified in a similar way.

We have checked all the ring axioms for $(R/I, +, \times)$, so the proof is complete.

Examples

1. The quotient ring $\mathbb{Z}/n\mathbb{Z}$ has n elements $0+n\mathbb{Z}, 1+n\mathbb{Z}, \ldots (n-1)+n\mathbb{Z}$. The sum or product of two cosets $a+n\mathbb{Z}$ and $b+n\mathbb{Z}$ is the coset containing $a+b$ or ab, respectively. Thus addition and multiplication in $\mathbb{Z}/n\mathbb{Z}$ are the same as in the ring $\mathbb{Z}_n$, if we identify $k \in \mathbb{Z}_n$ with the coset $k+n\mathbb{Z}$ in $\mathbb{Z}/n\mathbb{Z}$. In other words the quotient ring $\mathbb{Z}/n\mathbb{Z}$ is isomorphic to $\mathbb{Z}_n$.

2. Let I be the principal ideal $(x^2+1)\mathbb{R}[x]$ in $\mathbb{R}[x]$. Then for any $m \geq 2$ the coset x^m+I is the same as the coset $-x^{m-2}+I$, since $x^m-(-x^{m-2}) = (x^2+1)x^{m-2} \in I$. If $p(x) = a_m x^m + \ldots + a_1 x + a_0$, then

$$p(x) + I = (a_m x^m + I) + \ldots + (a_1 x + I) + (a_0 + I) = (a + bx) + I,$$

   where $a = a_0 - a_2 + a_4 - \ldots$ and $b = a_1 - a_3 + a_5 - \ldots$. Hence every element of $\mathbb{R}[x]/I$ can be (uniquely) expressed in the form $(a+bx)+I$ with $a, b \in \mathbb{R}$.

   Addition and multiplication in $\mathbb{R}[x]/I$ are defined by

$$(a + bx + I) + (c + dx + I) = (a + c) + (b + d)x + I,$$

   and

$$(a+bx+I)(c+dx+I) = ac+(ad+bc)x+bdx^2+I = (ac-bd)+(ad+bc)x+I.$$

   These are similar to the rules for adding and multiplying complex numbers, and indeed the quotient ring $\mathbb{R}[x]/I$ is isomorphic to $\mathbb{C}$ via the map $(a+bx)+I \mapsto a+bi$.

3. Let $I \subset \mathbb{R}[x]$ be the principal ideal $x^2\mathbb{R}[x]$. As in the previous example, each coset in $\mathbb{R}[x]/I$ can be uniquely expressed as $a+bx+I$ with $a, b \in \mathbb{R}$. As an additive group, $\mathbb{R}[x]/I \cong \mathbb{R}^2 \cong \mathbb{C}$, but the multiplication rule in $\mathbb{R}[x]/I$ is different from that in $\mathbb{C}$. We can regard $\mathbb{R}[x]/I$ as $\mathbb{R}^2$ with multiplication given by $(a,b)(c,d) = (ac, ad+bc)$. Note that $(0,1)^2 = (0,0)$ – or, equivalently, $(x+I)^2 = x^2 + I = I$, so that $\mathbb{R}[x]/I$ has zero-divisors, and so cannot be a field.

Theorem 3.3 (First Isomorphism Theorem for Rings)
Let $f : R \to S$ be a ring homomorphism. Then

$$\frac{R}{\mathrm{Ker}(f)} \cong \mathrm{Im}(f).$$

Proof. We must exhibit an isomorphism from R/K to S, where $K = \mathrm{Ker}(f)$ and $S = \mathrm{Im}(f)$. Define $\theta : R/K \to S$ by $\theta(r+K) = f(r)$. The proof of the First Isomorphism Theorem for groups shows that θ is a well-defined isomorphism of groups. In particular it is bijective. In order to show that it is a ring isomorphism, we need only check that it preserves multiplication. But

$$\theta((r+K)(s+K)) = \theta((rs)+K) = f(rs) = f(r)f(s) = \theta(r+K)\theta(s+K),$$

as required.

Examples

1. Define $f : \mathbb{Z} \to \mathbb{Z}_n$ by $f(k) = k \bmod n$. Then f is a surjective homomorphism of rings, and $\mathrm{Ker}(f) = n\mathbb{Z}$, confirming that $\mathbb{Z}/n\mathbb{Z} \cong \mathrm{Im}(f) = \mathbb{Z}_n$.

2. Let $f : \mathbb{R}[x] \to \mathbb{C}$ be the evaluation homomorphism $f(p(x)) := p(i)$. Let $K = \mathrm{Ker}(f)$. Then $x^2 + 1 \in K$, since $f(x^2 + 1) = i^2 + 1 = 0$. By the ideal property, $(x^2 + 1)p(x) \in K$ for every polynomial $p(x) \in \mathbb{R}[x]$, so K contains the principal ideal $I = (x^2+1)\mathbb{R}[x]$. As described in a previous example, every coset in $\mathbb{R}[x]/I$ can be uniquely expressed as $a + bx + I$ for $a, b \in \mathbb{R}$. But $f(a + bx) = a + bi = 0$ only if $a = b = 0$, so the only coset of I contained in K is I itself, from which it follows that $K = I = (x^2 + 1)\mathbb{R}[x]$. The homomorphism f is clearly surjective: given $z = a+bi \in \mathbb{C}$, we have $z = f(a+bx)$. This confirms our earlier observation that $\mathbb{R}[x]/I \cong \mathbb{C}$.

3. Let $f : \mathbb{R}[x] \to \mathbb{R}$ be the evaluation homomorphism $f(p(x)) = p(0)$. Then f is surjective, since for any $a \in \mathbb{R}$ we have $f(a) = a$. The kernel of f is the set of polynomials with constant term 0. But $a_m x^m + \ldots + a_2 x^2 + a_1 x + 0 = x(a_m x^{m-1} + \ldots + a_2 x + a_1)$, so $\mathrm{Ker}(f)$ is the principal ideal $x\mathbb{R}[x]$. Hence $\mathbb{R}[x]/x\mathbb{R}[x] \cong \mathbb{R}$.

4. Suppose $\alpha \in \mathbb{C}$. Let $f : \mathbb{Q}[x] \to \mathbb{C}$ be the evaluation homomorphism $f(p(x)) = p(\alpha) \in \mathbb{C}$. We write $\mathbb{Q}[i]$ for $\mathrm{Im}(f)$. This is a subring of $\mathbb{C}$ containing $\mathbb{Q}$ and α, and indeed it is the smallest such subring. The First Isomorphism Theorem shows that $\mathbb{Q}[\alpha] \cong \mathbb{Q}[x]/I$ for some ideal $I \subset \mathbb{Q}[x]$. For example, when $\alpha = i$ we get the field $\mathbb{Q}[i]$ of Gaussian rationals, and I turns out to be the principal ideal generated by $x^2 + 1$. If, on the other hand, we take $\alpha = \cos(\pi/5) + i\sin(\pi/5)$, then $\mathbb{Q}[\alpha]$ turns out to be a subfield of $\mathbb{C}$, and I turns out to be the principal ideal generated by $x^4 - x^3 + x^2 - x + 1$. We will learn more about examples like this later in the course.

3.4 More Isomorphism Theorems

The First Isomorphism Theorem has two easy corollaries, known as the Second and Third Isomorphism Theorems.

Theorem 3.4 (Second Isomorphism Theorem for Rings)
Let R be a ring, S a subring of R, and I an ideal of R. Then the set $S + I = \{s + x;\ s \in S,\ x \in I\}$ is a subring of R that contains I, $S \cap I$ is an ideal of S, and

$$\frac{S}{S \cap I} \cong \frac{S + I}{I}.$$

Proof. It is not difficult to check that $S + I$ is a subring of R.

Define a ring homomorphism $f : S \to R/I$ by $f(s) = s + I$. Then clearly $\text{Im}(f) = (S + I)/I$, while $\text{Ker}(f) = S \cap I$. The result follows from the First Isomorphism Theorem.

Example Let $R = \mathbb{R}[x]$, $S = \mathbb{Z}[x]$, and let I be the principal ideal $x\mathbb{R}[x]$ in $\mathbb{R}[x]$ – in other words, the ideal consisting of all polynomials with real coefficients and constant term zero. Then $S + I$ is the set of polynomials in $\mathbb{R}[x]$ whose constant term is an integer. Using the subring test, we can easily check that this is a subring of $\mathbb{R}[x]$. The ideal $S \cap I$ of $S = \mathbb{Z}[x]$ consists of all polynomials with integer coefficients, whose constant term is zero. Hence $S \cap I$ is the principal ideal $x\mathbb{Z}[x]$. Finally

$$\frac{\mathbb{Z}[x] + x\mathbb{R}[x]}{x\mathbb{R}[x]} \cong \mathbb{Z} \cong \frac{\mathbb{Z}[x]}{x\mathbb{Z}[x]} = \frac{\mathbb{Z}[x]}{Z[x] \cap x\mathbb{R}[x]}.$$

Theorem 3.5 (Third Isomorphism Theorem for Rings)

Let R be a ring, and I, J ideals of R such that $J \subset I$. then I/J is an ideal in R/J, and

$$\frac{R/J}{I/J} \cong \frac{R}{I}.$$

Proof. Define a ring homomorphism $f : R/J \to R/I$ by $f(r + J) = r + I$, and apply the First Isomorphism Theorem.

Example In $\mathbb{Z}_4$ the set $I = \{0, 2\}$ is the prinicpal ideal $2\mathbb{Z}_4$. But $\mathbb{Z}_4 \cong \mathbb{Z}/4\mathbb{Z}$, and this ideal I corresponds to the ideal $2\mathbb{Z}/4\mathbb{Z}$ in $\mathbb{Z}/4\mathbb{Z}$. Thus

$$\frac{\mathbb{Z}_4}{I} \cong \frac{\mathbb{Z}/4\mathbb{Z}}{2\mathbb{Z}/4\mathbb{Z}} \cong \frac{\mathbb{Z}}{2\mathbb{Z}} \cong \mathbb{Z}_2.$$

Exercises on homomorphisms, ideals and quotient rings

1. Show that $\phi : \mathbb{C} \to M_2(\mathbb{R})$ given by
$$\phi(a+ib) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$
is a ring homomorphism.

2. Show that $\phi : \mathbb{C} \to \mathbb{R}$ defined by $\phi(a+ib) = a$ is not a ring homomorphism.

3. Which of the following are ring homomorphisms?
   (a) $f : \mathbb{Z} \to \mathbb{Z}$, $f(n) = 2n$ for all $n \in \mathbb{Z}$.
   (b) $f : \mathbb{Z} \to M_2(\mathbb{R})$, $f(n) = nI_2$ for all $n \in \mathbb{Z}$.
   (c) $f : M_2(\mathbb{R}) \to \mathbb{R}$, $f(A) = \det(A)$ for all $A \in M_2(\mathbb{R})$.
   (d) $f : \mathbb{R}[x] \to \mathbb{C}$, $f(p(x)) = p(1+i)$ for all $p(x) \in \mathbb{R}[x]$.

4. Let R be a ring with identity and let S be an ideal of R. Prove that
   (i) if $1_R \in S$, then $S = R$;
   (ii) if S contains a unit u of R, then $S = R$.

5. Let R be a ring and let N_1 and N_2 be ideals of R. Prove that $N_1 \cap N_2$ is an ideal of R.

6. Let R, S be rings with 1. Show that every ideal A of the product ring $R \times S$ has the form $I \times J$ for some ideals I of R and J of S.
(Hint: define I to be $\{r \in R : \ (r,0) \in A\}$.)

7. Show that $f : \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$, defined by $f(n) = (n,n)$, is a ring homomorphism. Hence find a subring of $\mathbb{Z} \times \mathbb{Z}$ which is not an ideal.

8. List the elements in the quotient ring $2\mathbb{Z}/8\mathbb{Z}$. Are $2\mathbb{Z}/8\mathbb{Z}$ and $\mathbb{Z}_4$ isomorphic rings?

9. Let S_2 denote the ideal in $\mathbb{Z}[x]$ consisting of all polynomials in which both the constant term and the x term equal zero: $S_2 := \{p(x) \in \mathbb{Z}[x];\ p(0) = p'(0) = 0\}$, where $p'(x)$ is the derivative of $p(x)$. By using the First Isomorphism Theorem show that $\mathbb{Z}[x]/S_2$ is isomorphic to the ring of matrices
$$\left\{ \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} : a, b \in \mathbb{Z} \right\}.$$

Chapter 4

Special types of ideals

Throughout this chapter, and indeed for the rest of the course, we will restirct our attention to rings which are commutative and contain an identity.

In this chapter we study three special kinds of ideals in such rings.

4.1 Principal ideals

Let R be a commutative ring with identity. As defined in the previous chapter, an ideal $I \subset R$ is *principal*, with *generator* $x \in R$ if $I = xR = \{xr : \ r \in R\}$.

Examples

1. The ideals $n\mathbb{Z}$ of $\mathbb{Z}$ are principal.

2. The ideal $(x^2+1)\mathbb{R}[x]$ of $\mathbb{R}[x]$ is principal, with generator x^2+1.

3. Let $I = \{p(x) \in \mathbb{Z}[x]; p(0) \in 2\mathbb{Z}\}$. Then I is an ideal in $\mathbb{Z}[x]$ – for example since it is the kernel of the evaluation homomorphism $f : \mathbb{Z}[x] \to \mathbb{Z}_2$, $f(p(x)) = p(0) \bmod 2$. But I is not principal. To see this, we must show that $I \neq p(x)\mathbb{Z}[x]$ for any $p(x) \in \mathbb{Z}[x]$. Suppose first that $p(x)$ is the constant polynomial a_0 for some $a_0 \in \mathbb{Z}$. Then $a_0 = p(0) \in 2\mathbb{Z}$, so $I = a_0\mathbb{Z}[x] \subset 2\mathbb{Z}[x]$, which is impossible since for example $x \in I$ but $x \notin 2\mathbb{Z}[x]$. On the other hand, if $p(x)$ has degree $m > 0$, then $p(x)q(x)$ has degree m or greater for any nonzero polynomial $q(x)$, so there are no nonzero constant polynomials in $I = p(x)\mathbb{Z}[x]$. But this contradicts the fact that $2 \in \mathbb{Z}[x]$.

The following result is a useful characterisation of the principal ideal xR of a ring R.

Lemma 4.1 *Let R be a commutative ring with identity, and let $x \in R$. Then the principal ideal xR is the smallest ideal in R that contains x.*

Proof. We have already seen that xR is an ideal. Since R has an identity, we see that $x = x1_R \in xR$.

Conversely, suppose that I is an ideal in R and that $x \in R$. Then the ideal property says that $xr \in I$ for all $r \in R$, and so $xR \subset I$.

Definition A *principal ideal domain* (or PID) is an integral domain in which every ideal is principal.

Examples

1. The polynomial ideal $\mathbb{Z}[x]$ is an integral domain, but is not a principal ideal domain, since it contains an ideal which is not principal. (See the example above.)

2. $\mathbb{Z}$ is a principal ideal domain. We already know that $\mathbb{Z}$ is an integral domain. Suppose that $I \subset \mathbb{Z}$ is an ideal. We need to show that $I = n\mathbb{Z}$ for some $n \in \mathbb{Z}$. If $I = \{0\}$ then $I = 0\mathbb{Z}$, so we can suppose that $I \neq \{0\}$. Since $0 \in I$, there must be some $k \in I$ with $k \neq 0$. Since I is an additive subgroup of $\mathbb{Z}$, we also have $-k \in I$, and so I contains a positive integer.

   Let n be the *least* positive integer in I. Then $n \in I$ so $n\mathbb{Z} \subset I$. We will see that, in fact, $I = n\mathbb{Z}$. To see this, let us suppose the statement is false, and derive a contradiction.

   If $I \neq n\mathbb{Z}$, then there is an integer $k \in I$ with $k \notin n\mathbb{Z}$. Then $-k \in I$ with $-k \notin n\mathbb{Z}$, so there is a positive integer in $I \setminus n\mathbb{Z}$.

   Let m be the least such positive integer. Then $m > n$ by choice of n as the least positive integer in I. Then $m - n \in I$ since $m \in I$ and $-n \in n\mathbb{Z} \subset I$. But $0 < m - n < m$, and so by choice of m as the least positive integer in $I \setminus n\mathbb{Z}$ we must have $m - n \in n\mathbb{Z}$. But then $m = (m - n) + n \in n\mathbb{Z}$, since $m - n \in n\mathbb{Z}$ and $n \in n\mathbb{Z}$. This contradicts our choice of $m \notin n\mathbb{Z}$.

   Hence $I = n\mathbb{Z}$ as claimed.

3. Every field is a principal ideal domain. The only two ideals in a field F are $\{0\} = 0F$ and $F = 1F$. Each of these is principal.

4. If F is a field, then the polynomial ring $F[x]$ is a principal ideal domain. The proof is similar to the case of $\mathbb{Z}$. We already know that $F[x]$ is an integral domain. The ideal $\{0\} = 0F[x]$ is principal, so we suppose $I \neq \{0\}$ is an ideal and show that it must be principal.

   Choose a nonzero polynomial $p(x) \in I$ of least degree (m, say). Then $p(x)F[x] \subset I$. We will show that $I = p(x)F[x]$.

   If $I \neq p(x)F[x]$, let $q(x)$ be a polynomial in $I \setminus p(x)F[x]$ of least degree (n, say). Then $n \geq m$, by choice of $p(x)$. Suppose that the leading coefficient of $p(x)$ is a_m, while that of $q(x)$ is b_n. Let $r(x) = a_m q(x) - b_n x^{n-m} p(x)$. Then $r(x) \in I$ by the ideal properties, since $p(x), q(x) \in I$. Moreover, $r(x)$ has degree at most

$n-1$, since the coeficient of x^n in $r(x)$ is $a_m b_n - b_n a_m = 0$. Hence $r(x) \in p(x)F[x]$, by choice of $q(x)$. But then $q(x) = a_m^{-1}(r(x) + b_n x^{n-m} p(x)) \in p(x)F[x]$, a contradiction.

Hence $I = p(x)F[x]$ as claimed.

5. The ring $\mathbb{Z}[i]$ of Gaussian integers is a principal ideal domain. Certainly $\mathbb{Z}[i]$ is an integral domain, since it is a subring of the field $\mathbb{C}$ that contains the identity of $\mathbb{C}$. The ideal $\{0\} = 0\mathbb{Z}[i]$ is principal. Suppose that $I \neq \{0\}$ is an ideal. We must show that I is principal.

   Choose a nonzero element $z = a + ib \in I$ such that $|z|^2 = a^2 + b^2 \in \mathbb{N}$ is least possible. Clearly $z\mathbb{Z}[i] \subset I$. We will show that $I = z\mathbb{Z}[i]$. Suppose that $w = c + id \in I$. Then $v = (w/z) = x + iy \in \mathbb{C}$ is a complex number. If we let e, f be the integers closest to x, y respectively, then $u = e + if \in \mathbb{Z}[i]$, and $|u - v|^2 = |x - e|^2 + |y - f|^2 \leq \frac{1}{2}$. Now $zu - w \in I$, and $|zu - w|^2 = |z|^2|u - v|^2 \leq \frac{1}{2}|z|^2$. By choice of z, we must have $zu - w = 0$, so $w = zu \in z\mathbb{Z}[i]$, as required.

4.2 Maximal ideals

In a given ring R, we can compare ideals using the partial order $\subset$. If $I \subset J$ then we regard J as 'bigger than' I. In this context, the unique 'smallest' ideal is $\{0\}$ and the unique 'biggest' ideal is R itself. An ideal $I \subset R$ is said to be a *proper ideal* if $I \neq R$. In this section we consider proper ideals which are maximal with respect to the partial order $\subset$.

Definition An ideal $I \subset R$ is *maximal* if $I \neq R$, and for any ideal $J \subset R$ with $I \subset J \subset R$, either $J = I$ or $J = R$.

Examples

1. In $\mathbb{Z}$, the ideal $4\mathbb{Z}$ is not maximal, since $4\mathbb{Z} \subset 2\mathbb{Z} \subset \mathbb{Z}$, but $4\mathbb{Z} \neq 2\mathbb{Z} \neq \mathbb{Z}$. However, $2\mathbb{Z}$ is maximal.

   To see this, suppose that $2\mathbb{Z} \subset I \subset \mathbb{Z}$ for some ideal I. If $I \neq 2\mathbb{Z}$ then there exists an odd integer $k \in I$. Then $(k+1)$ is even, so $(k+1) \in 2\mathbb{Z} \subset I$. Since I is an ideal, we can deduce that $1 = (k+1) - k \in I$, so $y = 1y \in I$ for all $y \in \mathbb{Z}$. In other words, $I = \mathbb{Z}$.

2. If F is a field, then $\{0\}$ is a maximal ideal, since the only other ideal of F is F itself.

3. In $\mathbb{R}[x]$, the principal ideal $M = (x^2 + 1)\mathbb{R}[x]$ is maximal.

   To see this, suppose that $M \subset I$ for some ideal I of $\mathbb{R}[x]$. If $I \neq M$ then there exists a polynomial $p(x) \in I$ with $p(x) \notin M$. In particular $p(x) \neq 0$. Among all such $p(x)$, choose one with the least possible degree, d say. Write $p(x) =$

$a_d x^d + \cdots + a_1 x + a_0$. If $d > 1$ then put $q(x) = p(x) - a_d x^{d-2}(x^2+1) \in I$. Then $q(x)$ has degree less than d, so $q(x) \in M$. But then $p(x) = q(x) + a_d x^{d-2}(x^2+1) \in M$, contradicting the choice of $p(x)$. We must therefore have $d \leq 1$, say $p(x) = ax+b$, with $(a,b) \neq (0,0)$ since $p(x) \neq 0$ in $\mathbb{R}[x]$.

Then $a^2 + b^2 = a^2(x^2+1) - (ax-b)p(x) \in I$. Since $a, b \in \mathbb{R}$ with $(a,b) \neq (0,0)$, $c = a^2 + b^2 > 0$, and c is a unit in $\mathbb{R}[x]$. Since I contains a unit, it must be the whole ring. (In other words, $p(x) = c(c^{-1}p(x)) \in I$ for all $p(x) \in \mathbb{R}[x]$.)

The most important property of maximal ideals relates to the corresponding quotient rings.

Theorem 4.2 *Let R be a commutative ring with identity, and let M be an ideal in R. Then M is a maximal ideal if and only if the quotient ring R/M is a field.*

Proof. Suppose first that M is maximal. The ring R/M is commutative, since R is commutative. ($(x+M)(y+M) = xy + M = yx + M = (y+M)(x+M)$.) It also has an identity $1_{R/M} = 1_R + M$. Moreover, $1_{R/M} \neq 0_{R/M}$ since $1_R \notin M$. (If $1_R \in M$, then $M = R$, contadicting the definition of maximal ideal.)

To show that R/M is a field, it remains to prove that every nonzero element $x+M$ in R/M has an inverse. Now $x \notin M$, since $x + M \neq 0 + M$. We will show that $I = M + xR = \{m + xr;\ m \in M,\ r \in R\}$ is an ideal in R. Clearly $M \subset I$, and $M \neq I$ since $x \in I$ and $x \notin M$. Since M is maximal, it follows that $I = R$, and in particular $1_R \in I$ – say $1_R = m + xy$ with $m \in M$ and $y \in R$. Then $(x+M)(y+M) = xy + M = (1_R - m) + M = 1_R + M$, so $x + M$ has an inverse in R/M, as required.

We still have to check that I is an ideal. Let $m_1, m_2 \in M$ and $r_1, r_2, s \in R$. Then $(m_1 + xr_1) + (m_2 + xr_2) = (m_1 + m_2) + x(r_1 + r_2) \in I$, $0 = 0 + x0 \in I$, $-(m_1 + xr_1) = (-m_1) + x(-r_1) \in I$, and $s(m_1 + xr_1) = (m_1 + xr_1)s = m_1 s + x(r_1 s) \in I$. Hence I is an ideal, as claimed.

Conversely, suppose that R/M is a field. Then $M \neq R$, since otherwise $0 = 1$ in R/M, contrary to the definition of a field. Suppose I is an ideal of R with $M \subset I \subset R$. By the third isomorphism theorem, I/M is an ideal in the field $F = R/M$. But a field F has only two ideals F and $\{0\}$. If $I/M = F = R/M$ then $I = R$, while if $I/M = \{0\} = M/M$ then $I = M$. Hence M is maximal, as required.

4.3 Prime ideals

Ideals were originally introduced in number theory, in a (failed) attempt to prove Fermat's Last Theorem, that $x^n + y^n = z^n$ has no nontrivial integer solutions when $n > 2$. The meaning was something like 'ideal number'. One should think of an arbitrary ideal as a generalisation of the principal ideals $n\mathbb{Z}$ of $\mathbb{Z}$, which correspond to the natural numbers n.

In this context, the concept of a *prime ideal* is a generalisation of *prime number*. Properties of prime numbers are defined in terms of divisibility of natural numbers. If (say) n divides x, then $x = ny$ for some y, so $x \in n\mathbb{Z}$. Thus divisibility of numbers corresponds to membership of ideals. This prompts the following definition.

Definition An ideal P of a ring R is *prime* if $P \neq R$ and, whenever $x, y \in R$ such that $xy \in P$, then at least one of $x \in P$, $y \in P$ holds.

Examples

1. If p is a prime number, then $p\mathbb{Z}$ is a prime ideal in $\mathbb{Z}$.

   Clearly $p\mathbb{Z} \neq \mathbb{Z}$, since $p \geq 2$. Suppose that $x, y \in \mathbb{Z}$ and that $xy \in p\mathbb{Z}$. Then the prime number p divides xy, so by the properties of prime numbers, p divides at least one of x, y. In other words, at least one of x, y belongs to $p\mathbb{Z}$.

2. Let R be a commutative ring with identity. Then every maximal ideal of R is prime.

   If M is a maximal ideal of R then $M \neq R$. Moreover, the quotient ring R/M is a field. If $x, y \in R$ with $xy \in M$, then $(x + M)(y + M) = 0 + M$ in the field R/M, which is possible only if one of the elements $x + M, y + M$ is zero in R/M. But, for example, $x + M = 0 + M$ if and only if $x \in M$.

3. The ideal $x\mathbb{Z}[x]$ is prime, but not maximal, in $\mathbb{Z}[x]$.

   Certainly, $x\mathbb{Z}[x]$ is not maximal, since the quotient ring $\mathbb{Z}[x]/x\mathbb{Z}[x] \cong \mathbb{Z}$ is not a field. Also $x\mathbb{Z}[x] \neq \mathbb{Z}[x]$.

   Note that $p(x) \in x\mathbb{Z}[x]$ if and only if $p(0) = 0$. If $p(x), q(x) \in \mathbb{Z}[x]$ with $p(x)q(x) \in x\mathbb{Z}[x]$, then $p(0)q(0) = 0$, so either $p(0) = 0$ or $q(0) = 0$ (or both). But $p(0) = 0$ if and only if $p(x) \in x\mathbb{Z}[x]$, and similarly for $q(x)$.

We characterised maximal ideals in a commutative ring with identity as those for which the corresponding quotient ring is a field. There is a similar characterisation of prime ideals.

Theorem 4.3 *Let R be a commutative ring with identity, and let P be an ideal in R. Then P is a prime ideal if and only if the quotient ring R/P is a non-zero integral domain.*

Proof. Suppose first that P is a prime ideal. Then $P \neq R$, so $1_R \notin P$, and R/P is a commutative ring with identity $1 \neq 0$.

To show that R/P has no zero-divisors, suppose that $x, y \in R$ such that $(x+P)(y+P) = (0 + P) = P$. Then $xy \in P$. Since P is prime, at least one of x, y belongs to P, so at least one of $x + P$, $y + P$ is the zero element $0 + P$ of R/P.

Conversely, suppose that R/P is a non-zero integral domain. Then $P \neq R$ since $R/P \neq \{0\}$. To show that P is prime, suppose that $x, y \in R$ with $xy \in P$. Then

$(x + P)(y + P) = xy + P = P = 0 + P$ in R/P. But R/P has no zero-divisors, so at least one of $x + P$, $y + P$ is equal to $0 + P$. In other words, at least one of x, y belongs to P. Hence P is prime, as claimed.

Exercises on principal ideals, maximal ideals and prime ideals

1. Let m, n be non-negative integers. In the product ring $\mathbb{Z} \times \mathbb{Z}$, show that the ideal $m\mathbb{Z} \times n\mathbb{Z}$ is principal, with generator (m, n). Deduce that every ideal in $\mathbb{Z} \times \mathbb{Z}$ is principal. Why is $\mathbb{Z} \times \mathbb{Z}$ not a principal ideal domain?

2. If $\phi : R \to S$ is a ring homomorphism, and I is an ideal in S, show that

$$\phi^{-1}(I) = \{r \in R;\ \phi(r) \in I\}$$

   is an ideal in R. Deduce that every ideal in $\mathbb{Z}_n$ is principal.

3. In the ring $\mathbb{Z} \times \mathbb{Z}$ show that

   (a) $\mathbb{Z} \times 2\mathbb{Z}$ is a maximal ideal;

   (b) $\mathbb{Z} \times \{0\}$ is a prime ideal which is not a maximal ideal;

   (c) $2\mathbb{Z} \times 2\mathbb{Z}$ is not a prime ideal.

4. Find all the maximal ideals in $\mathbb{Z}_{12}$.

5. Determine whether or not $3\mathbb{Z} \times 5\mathbb{Z} \times 7\mathbb{Z}$ is a prime ideal in $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$.

6. Apply the first isomorphism theorem to the evaluation homomorphism $\mathbb{Z}[x] \to \mathbb{Z}$, $p(x) \mapsto p(1)$, to show that the ideal $J := \{p(x) \in \mathbb{Z}[x];\ p(1) = 0\}$ of the ring $\mathbb{Z}[x]$ is not maximal. Is J prime?

7. Let R be a *finite* commutative ring with identity. Show that every prime ideal of R is a maximal ideal.

Chapter 5

Polynomial Rings

In this chapter we will study rings of the form $R[x]$, the ring of polynomials in one variable x with coefficients from the ring R.

In practice, we will mainly be interested in the case where R is a field, or at least an integral domain. But the definition makes sense for any commutative ring R.

5.1 Polynomials

Let R be a commutative ring, and let $n \geq 0$ be an integer. A *polynomial* in x of *degree* n with coefficients from R is a formal expression

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_0, \ldots, a_n \in R$. If $d > 0$ then we insist that the *leading coefficient* a_n is nonzero.

If R has an identity 1_R, and the leading coefficient of $p(x)$ is 1_R, then $p(x)$ is said to be a *monic* polynomial.

The collection of all polynomials with coefficients from R forms a commutative ring $R[x]$, with addition and multiplication defined just as for polynomials with coefficients in familiar rings such as $\mathbb{Z}$ or $\mathbb{R}$.

In earlier chapters, we have already noted some properties of polynomials and polynomial rings. The following have either been proved earlier or are easy to check.

Lemma 5.1 *1. If $p(x), q(x) \in R[x]$ have degrees m, n respectively, and $m \neq n$, then $p(x) + q(x)$ has degree* $\max(m, n)$. *(If $m = n$, then the degree of $p(x) + q(x)$ is* at most *n.)*

2. If $p(x) \in R[x]$ has degree n, then so does $-p(x)$.

3. If $p(x), q(x) \in R[x]$ have degrees m, n respectively, then the degree of $p(x)q(x)$ is at most $m + n$. If R is an integral domain and $p(x) \neq 0 \neq q(x)$, then the degree of $p(x)q(x)$ is exactly *$m + n$.*

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

5. *If F is a field, then $F[x]$ is a principal ideal domain.*

If S is a ring containing R, $\alpha \in S$, and $p(x) \in R[x]$, then $R[x] \subset S[x]$, so we can think of $p(x)$ as having coefficients in S. We can also substitute α for x in the expression

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

for $p(x)$ to get an expression

$$p(\alpha) = a_n \alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1 \alpha + a_0,$$

which we can evaluate in S.

This gives rise to an *evaluation homomorphism* $\phi_\alpha : R[x] \to S$, $\phi_\alpha(p(x)) := p(\alpha)$.

One can also consider polynomial rings in two or more variables, with appropriate definitions. One way to think of such a ring $R[x_1, \ldots, x_d]$ is as an iterated polynomial ring $R[x_1][x_2]\cdots[x_d]$. In other words, $R[x_1, \ldots, x_d]$ is the ring of polynomials in x_d with coefficients from $R[x_1, \ldots, x_{d-1}]$, which is itself the ring of polynomials in x_{d-1} with coefficients from $R[x_1, \ldots, x_{d-2}]$, etc.

This allows us to use inductive arguments, to prove results such as:

Theorem 5.2 *If R is an integral domain, and $d > 0$, then $R[x_1, \ldots, x_d]$ is also an integral domain.*

In this course, we will however concentrate on polynomial rings in one variable only.

5.2 Polynomials with coefficients in a field

Suppose that F is a field. Then we have seen that the polynomial ring $F[x]$ is a principal ideal domain. So we can associate, to every ideal, a generator $p(x)$ of that ideal. Suppose that $I = p(x)F[x]$ and $J = q(x)F[x]$ are two ideals in $F[x]$. Then we can translate properties of the ideals into properties of their generators.

For example, $J \subset I$ if and only if $q(x) \in I$, since J is the smallest ideal containing $q(x)$. But $q(x) \in I = p(x)F[x]$ if and only if $q(x) = p(x)a(x)$ for some polynomial $a(x) \in F[x]$. If this happens, we say that $q(x)$ is a *multiple* of $p(x)$, or that $p(x)$ *divides* $q(x)$, or that $p(x)$ is a *factor* of $q(x)$.

One obvious question to ask is when two different polynomials generate the same ideal. (We cannot expect, in general, that the generator of an ideal will be unique.)

Lemma 5.3 *Let F be a field and $p(x), q(x) \in F[x]$. Then $p(x)F[x] = q(x)F[x]$ if and only if $q(x) = ap(x)$ for some nonzero constant $a \in F$.*

Proof. Now $p(x)F[x] = q(x)F[x]$ if and only if there are polynomials $a(x)$ and $b(x)$ such that $q(x) = p(x)a(x)$ and $p(x) = q(x)b(x)$. This is true if and only if $p(x)(1 - a(x)b(x)) = 0 = q(x)(1-b(x)a(x))$. Since $F[x]$ is an integral domain, this means either $p(x) = 0 = q(x)$ or $a(x)b(x) = 1$.

In the second case, the degrees of $a(x)$ and $b(x)$ are both 0, since the product $a(x)b(x)$ has degree 0, so a is a nonzero constant in F, with inverse b.

Conversely, if $q(x) = ap(x) \in p(x)F[x]$ with a a nonzero constant, then a is a unit in F, so has an inverse, $b \in F$ say, and then $p(x) = bq(x) \in q(x)F[x]$, and so $p(x)F[x] = q(x)F[x]$.

Corollary 5.4 *Every nonzero ideal I in $F[x]$ has the form $p(x)$ for a* monic *polynomial $p(x)$. Moreover, this choice of monic polynomial is uniquely determined by I.*

Proof. Certainly I is principal, so $I = q(x)F[x]$ for some nonzero polynomial $q(x)$. Since $q(x) \neq 0$, the leading coefficient a_n (say) of $q(x)$ is nonzero. Define $p(x) = a_n^{-1}q(x)$. Then $p(x)$ has leading coefficient $a_n^{-1}a_n = 1$, so is monic. Also, by the lemma, we have $p(x)F[x] = q(x)F[x] = I$.

If $r(x)$ is another monic polynomial that generates I, then the lemma says that $r(x) = ap(x)$ for some constant $a \neq 0$. Comparing leading coefficients, we see that $1 = a.1$, so $a = 1$ and $r(x) = p(x)$, as claimed.

5.3 Long division and the euclidean algorithm

If we are given an ideal $I \neq \{0\}$ in $F[x]$ in some abstract way, how can we find its (monic) generator polynomial? Have another look at the proof that $F[x]$ is a PID to get a hint. $I = p(x)F[x]$ where $p(x)$ is the unique monic polynomial of least degree in I. (This makes sense: suppose that $p(x), q(x) \in I$ are monic polynomials of least possible degree d, then either $p(x) = q(x)$ or $p(x) - q(x) \in I$ is a nonzero polynomial of degree less than d. Some multiple of this will be monic, giving a contradiction.)

Here is a specific example. Suppose that $I = p(x)F[x]$ and $J = q(x)F[x]$ are two nonzero ideals in $F[x]$. Then the set $I + J := \{x + y; x \in I, y \in J\}$ is an ideal in $F[x]$ that contains both I and J. Hence $I + J = h(x)F[x]$ for some polynomial $h(x)$ that divides both $p(x)$ and $q(x)$. Can we find $h(x)$?

Before solving this problem, let us consider the analogous problem in a more familiar PID, namely $\mathbb{Z}$. Given positive integers m, n, the ideal $m\mathbb{Z}+n\mathbb{Z}$ has the form $h\mathbb{Z}$ where h is a common factor of m and n. Indeed h is the *highest common factor* of m, n. There is a well-known method of finding h called the *Euclidean algorithm*: interchange m, n if necessary so that $m \leq n$. Divide n by m to get a remainder r with $0 \leq r < m \leq n$. If $r = 0$ then m is the highest common factor; otherwise replace the pair (m, n) by (r, m) and continue.

Example Find the highest common factor of 63 and 96.

$$\begin{aligned} 96 &= 1 \times 63 + 33. \\ 63 &= 1 \times 33 + 30. \\ 33 &= 1 \times 30 + 3. \\ 30 &= 10 \times 3 + 0. \end{aligned}$$

The last remainder is zero, so the previous remainder, 3 is the highest common factor.

It turns out that the same algorithm works for polynomials with coefficients from a field F, where the measurement we use for the size of a polynomial is its degree. The individual division steps of the algorithm work because of the following result.

Lemma 5.5 *Let F be a field and $a(x), b(x) \in F[x] \setminus \{0\}$. Then there are unique polynomials $q(x), r(x) \in F[x]$ such that*

1. *$b(x) = a(x)q(x) + r(x)$; and*
2. *either $r(x) = 0$ or the degree of $r(x)$ is less than that of $a(x)$.*

Proof. Let I be the principal ideal $a(x)F[x]$. If $b(x) \in I$, then we have $b(x) = a(x)q(x)$ for some $q(x) \in F[x]$ (which is unique, since $F[x]$ is an integral domain), and the result is true with $r(x) = 0$.

Otherwise, choose $r(x)$ to be a polynomial of least possible degree in the coset $b(x) + I$. This degree is less than that of $a(x)$, as the following argument shows. Suppose that $a(x)$ has degree m and leading coefficient α, whereas $r(x)$ has degree $n \geq m$ and leading coefficient ρ. Then $r(x) - \alpha^{-1}\rho x^{n-m}a(x) \in r(x) + I = b(x) + I$ has degree less than n, contrary to the choice of $r(x)$.

It follows that the choice of $r(x)$ is unique – if $r'(x) \in r(x) + I$ also has smaller degree than $a(x)$, then $r(x) - r'(x) \in I = a(x)F[x]$ is a multiple of $a(x)$ but has degree less than that of $a(x)$, so must be zero.

Finally, $b(x) - r(x) \in I = a(x)F[x]$, so $b(x) - r(x) = a(x)q(x)$ for a (unique) polynomial $q(x)$.

To perform a division of polynomials – that is, to find $q(x)$ and $r(x)$ in the notation of the lemma – we can proceed in stages corresponding to the terms of $b(x)$. Suppose that $b(x)$ has degree k and leading coefficient β, while $a(x)$ has degree $m \leq k$ and leading coefficient α, as in the lemma. Then $b_1(x) = b(x) - \alpha^{-1}\beta x^{k-m}a(x)$ has degree less than k. Iterate this process. If $b_1(x) = a(x)q_1(x)+r(x)$, then $b(x) = a(x)q(x)+r(x)$, where $q(x) = \alpha^{-1}\beta x^{k-m} + q_1(x)$.

We can lay this calculation out as a *long division*. For example:

$$
\begin{array}{r}
\qquad\qquad x^2 + \ 2x + 11 \\
x^2 + x + 1 \overline{)x^4 + 3x^3 + 14x^2 - \ 7x - \ 5} \\
\underline{x^4 + \ x^3 + \quad x^2} \\
2x^3 + 13x^2 \\
\underline{2x^3 + \ 2x^2 + \ 2x} \\
11x^2 - \ 9x \\
\underline{11x^2 + 11x + 11} \\
-\ 20x - 16
\end{array}
$$

giving $x^4 + 3x^3 + 14x^2 - 7x - 5 = (x^2 + x + 1))(x^2 + 2x + 11) + (-20x - 16)$.

Example

Find the highest common factor of $x^3 - 3x^2 + x + 2$ and $x^2 - 4x + 4$. A long division shows that $x^3 - 3x^2 + x + 2 = (x^2 - 4x + 4)(x + 1) + (x - 2)$.

A second long division shows that $x^2 - 4x + 4 = (x - 2)(x - 2) + 0$. The remainder is 0, so the previous remainder, $x - 2$ is the highest common divisor.

5.4 Reducible and irreducible polynomials

Given an ideal I in $F[x]$, we know that the quotient ring $F[x]/I$ will be a field if I is maximal, and an integral domain if I is prime. But we also know that I is principal, $I = p(x)F[x]$ for some $p(x)$. (Indeed, we know that we can choose $p(x)$ to be either monic or the zero polynomial.)

Given $p(x)$, can we decide whether or not $I = p(x)F[x]$ is prime or maximal? There is a criterion that we can use.

Definition A polynomial $p(x) \in F[x]$ is *reducible* if it can be expressed as a product of polynomials of lower degrees. A polynomial is *irreducible* if it is not constant and not reducible.

Theorem 5.6 *Let F be a field, and let $p(x)$ be a non-constant polynomial in $F[x]$. Let I denote the principal ideal $I = p(x)F[x]$. Then the following are equivalent:*

(i) I is maximal;

(ii) I is prime;

(iii) $p(x)$ is irreducible.

Proof. (i)⇒(ii). If I is maximal, then it is prime.

(ii)⇒(iii). If $p(x)$ is reducible, say $p(x) = a(x)b(x)$ where $a(x)$ and $b(x)$ each has degree less than that of $p(x)$, then $a(x), b(x) \notin p(x)F[x] = I$, but $a(x)b(x) = p(x) \in I$, so I is not prime.

(iii)⇒(i). It remains to show that, if $p(x)$ is irreducible, then I is maximal. Suppose than that $I \subset J \subset F[x]$ for some ideal J of $F[x]$. Since $F[x]$ is a PID, $J(x) = a(x)F[x]$ for some $a(x)$. Now $p(x) \in I \subset J = a(x)F[x]$, so $p(x) = a(x)b(x)$ for some $b(x)$. Since $p(x)$ is irreducible, it is not possible that both $a(x)$ and $b(x)$ have degrees less than $p(x)$. Hence one of $a(x), b(x)$ has degree equal to that of $p(x)$, and the other is a nonzero constant (and hence a unit in $F[x]$).

There are two cases to consider. If $a(x)$ has the same degree as $p(x)$, then $b = b(x)$ is a unit. Then $a(x) = b^{-1}p(x) \in I$, so $J \subset I$ and hence in fact $I = J$. If, on the other hand, $a = a(x)$ is a unit, then $J = aF[x] = F[x]$. We have therefore shown that I is maximal.

Remark Whether or not a given polynomial is irreducible may depend on the field of coefficients in which we are working, as the following examples show.

Examples

1. $x^2 - 1$ is reducible in $F[x]$ for any field F, since $x^2 - 1 = (x - 1)(x + 1)$.

2. x^2+1 is irreducible in $\mathbb{R}[x]$. Suppose that x^2+1 is the product of two polynomials of degree less than 2 - necessarily each of degree 1. In other words, $x^2 + 1 = (ax + b)(cx + d)$. Then comparing coefficients gives $ac = bd = 1$ and $ad + bc = 0$. Hence
$$0 = ab(ad + bc) = a^2(bd) + b^2(ac) = a^2 + b^2,$$
so $a = b = 0$ and $1 = ac = 0c = 0$, a contradiction.

3. $x^2 + 1$ is reducible in $\mathbb{C}[x]$: $x^2 + 1 = (x + i)(x - i)$.

The ring of polynomials over a field resembles in many ways the ring of integers, with irreducible polynomials playing the part of prime numbers. In particular, there is the following anaologue of the Fundamental Theorem of Arithmetic, whose proof we will omit.

Theorem 5.7 *Let F be a field, and $p(x) \in F[x]$ a nonconstant polynomial. Then there is a unique constant c and a list $\alpha_1(x), \ldots, \alpha_k(x)$ (unique up to order) of monic irreducible polynomials, such that*

$$p(x) = c\alpha_1(x) \cdot \ldots \cdot \alpha_k(x).$$

5.5 Testing for irrecucibility

Given a (nonconstant) polynomial $p(x) \in F[x]$, how can we tell whether $p(x)$ is reducible or irreducible?

This partly depends on the field F, but there are some general rules based on the observation that, if $p(x) = a(x)b(x)$, then $\deg(p(x)) = \deg(a(x)) + \deg(b(x))$, so that at least one of the factors $a(x)$, $b(x)$ has degree less than or equal to $\frac{1}{2}\deg(p(x))$.

Further insight about reducibility comes from the following observation. Suppose that $p(x) \in F$ and $a \in F$. Consider the evaluation homomorphism $\phi_a : F[x] \to F$, $\phi_a(p(x)) := p(a)$. Clearly the linear polynomial $x - a$ belongs to the kernel of ϕ_a. But $x - a$ is irreducible, since it cannot be expressed as a product of constant polynomials. Hence $(x-a)F[x]$ is maximal, and $(x-a)F[x] \subset \mathrm{Ker}(\phi_a) \subset F[x]$. But $\mathrm{Ker}(\phi_a) \neq F[x]$, since for example $\phi_a(1) = 1 \neq 0$. Hence $\mathrm{Ker}(\phi_a) = (x-a)F[x]$. This proves the following result.

Lemma 5.8 *Let F be a field, $a \in F$ and $p(x) \in F[x]$. Then $x - a$ is a factor of $p(x)$ if and only if a is a* root *of $p(x)$ (that is, $p(a) = 0$, or $p(x) \in \mathrm{Ker}(\phi_a)$).*

Tests for irreducibility

1. Linear (i.e. degree 1) polynomials $ax + b$ (with $a \neq 0$) are *always* irreducible.
2. A quadratic (degree 2) or cubic (degree 3) polynomial $p(x) \in F[x]$ is reducible if and only if $p(x)$ has a root in F.
3. A polynomial $p(x)$ of degree 4 or 5 in $F[x]$ is reducible if and only if $p(x)$ either has a root in F or a quadratic factor.

As mentioned above, questions of irreducibility vary according to the field F. In particular, when $F = \mathbb{C}$, everything is completely explained by the remarkable theorem of Gauss:

Theorem 5.9 (Fundamental Theorem of Algebra) *Let $p(x) \in \mathbb{C}[x]$ be a nonconstant polynomial. then $p(x)$ has a root in $\mathbb{C}$.*

Corollary 5.10 *A polynomial $p(x) \in \mathbb{C}[x]$ is irreducible if and only if $p(x)$ is linear.*

Corollary 5.11 *Every poynomial in $\mathbb{R}[x]$ of degree greater than 2 is reducible.*

Proof. Let $p(x) \in \mathbb{R}[x] \subset \mathbb{C}[x]$ be a polynomial of degree greater than 2. Then $p(x)$ has a root $a \in \mathbb{C}$. If $a \in \mathbb{R}$ then $x - a$ is a factor of $p(x)$ in $\mathbb{R}[x]$. Otherwise, the complex conjugate $\overline{a}$ is also a root of $p(x)$, so $x - a$ and $x - \overline{a}$ are both factors of $p(x)$ (in $\mathbb{C}[x]$). Hence so is $(x-a)(x-\overline{a}) = x^2 - (2\mathrm{Re}(a))x + |a|^2 \in \mathbb{R}[x]$. Hence $p(x)$ has a factor of degree at most 2 in $\mathbb{R}[x]$, so is reducible.

In a (small) finite field, it is easy to check for roots (and hence linear factors) of a polynomial by direct calculation.

Examples

1. $x^3 + 2$ has no roots in $\mathbb{Z}_7$, and so is irreducible:

x	0	1	2	3	4	5	6
x^3+2	2	3	3	1	3	1	1

2. $x^4 + x + 1$ has no roots in $\mathbb{Z}_2$:

x	0	1
x^4+x+1	1	1

   Since $x^4 + x + 1$ has degree 4, this does not necessarily mean that $x^4 + x + 1$ is irreducible. We need to check also that $x^4 + x + +1$ has no quadratic factors. In fact, it is enough to check for irreducible quadratic factors, since if a polynomial has a reducible quadratic factor, it has a linear factor and hence a root. But of the four quadratic polyomials $x^2 + ax + b \in \mathbb{Z}_2[x]$, only one is irreducible, namely $x^2 + x + 1$. Since $(x^2 + x + 1)^2 = x^4 + x^2 + 1 \neq x^4 + x + 1$, it follows that $x^4 + x + 1$ is indeed irreducible in $\mathbb{Z}_2[x]$.

We can also do computations in $\mathbb{Q}$ easily to show that certain polynomials are irredcucible.

Example $x^3 + x + 1$ is irreducible in $\mathbb{Q}[x]$.

Otherwise, there would be a root m/n in $\mathbb{Q}$, so an equation

$$\frac{m^3}{n^3} + \frac{m}{n} + 1 = 0.$$

Multiplying both sides by $n^3 \neq 0$, we get $m^3 + mn^2 + n^2 = 0$. Form this equation, we see that, if m has a prime factor p, then $n^3 = 0 \bmod p$, so p is also a prime factor of n. Similarly, if q is a prime factor of n, then $m^3 = 0 \bmod q$, so q is also a prime factor of m. But we can choose m, n to have no common factors, so the only possibility is $m = \pm n = \pm 1$. By direct evaluation, neither 1 nor -1 is a root of $x^3 + x = 1$, so $x^3 + x = 1$ has no rational roots, and so is irreducible.

There are two other results which help us decide questions of reducibility in $\mathbb{Q}[x]$:

Theorem 5.12 (Gauss' Lemma) *Let $p(x) \in \mathbb{Z}[x]$. Then $p(x)$ is reducible in $\mathbb{Q}[x]$ if and only if it is reducible in $\mathbb{Z}[x]$.*

Theorem 5.13 (Eisenstein's criterion) *Let $\alpha(x) = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$. Suppose that p is a prime number such that p divides $a_0, a_1, \ldots, a_{n-1}$ but does not divide a_n, and that p^2 does not divide a_0. Then $\alpha(x)$ is irreducible in $\mathbb{Q}[x]$.*

Examples

1. $x^{300} - 17$ is irreducible in $\mathbb{Q}[x]$ (by Eisenstein's criterion with $p = 17$).

2. $x^4 + x + 1$ is irreducible in $\mathbb{Q}[x]$. For otherwise it is reducible in $\mathbb{Z}[x]$ by Gauss' lemma. If it has a linear factor, $x^4 + x + 1 = (ax + b)(cx^3 + dx^2 + ex + f)$ in $\mathbb{Z}[x]$, then equating coefficients gives $ac = bf = 1$, so $a = \pm b = \pm 1$. This is impossible since ± 1 are not roots of $x^4 + x + 1$. If there are two quadratic factors: $x^4 + x + 1 = (ax^2 + bx + c)(dx^2 + ex + f)$, then again equating coefficients gives $ad = cf = 1$, and also $0 = ae + bd = \pm(e + b)$, and $1 = bf + ce = \pm(b + e)$. This gives a contradiction.

Exercises on polynomial rings

1. Determine whether or not the following polynomials are irreducible in the given polynomial rings:

   (a) $x^2 + x + 1$ in $\mathbb{Z}_2[x]$;

   (b) $x^3 + x^2 + 3x + 5$ in $\mathbb{Z}_7[x]$;

   (c) $x^2 + 5x - 3$ in $\mathbb{R}[x]$;

   (d) $x^3 - 17x^2 + 24x - 1$ in $\mathbb{R}[x]$;

   (e) $x^{345} - 53x^{77} + 1234567x^{22} - 2x + 1$ in $\mathbb{C}[x]$.

   (f) $x^2 + x - 5$ in $\mathbb{Q}[x]$;

   (g) $x^5 + x^2 + x$ in $\mathbb{Z}_2[x]$.

2. In the field $\mathbb{Z}_5$, show that $a^5 = a$ for every $a \in \mathbb{Z}_5$. Deduce that

$$x^5 - x = x(x-1)(x-2)(x-3)(x-4)$$

   in $\mathbb{Z}_5[x]$.

3. Let F be a field and $I \neq F[x]$ an ideal in $F[x]$. Let $g(x) \in F[x]$ be an irreducible polynomial such that $g(x) \in I$. Show that $I = g(x)F[x]$.

4. Use Gauss' Lemma to show that $p(x) = x^4 + x^3 + x^2 + x + 1$ is irreducible in $\mathbb{Q}[x]$.

5. Find all the roots of $x^6 + 1$ in $\mathbb{C}$, and hence factorise $x^6 + 1$ as a product of three irreducible quadratics in $\mathbb{R}[x]$.

6. (a) Explain why $\mathbb{Q}[x]/(x^2 + 3)\mathbb{Q}[x]$ is a field.

   (b) Let $\mathbb{Q}[\sqrt{-3}] = \{a + b\sqrt{-3} \in \mathbb{C} \mid a, b \in \mathbb{Q}\}$. This is a ring under addition and multiplication of complex numbers. Show that there is an isomorphism of rings

$$\mathbb{Q}[x]/(x^2 + 3)\mathbb{Q}[x] \cong \mathbb{Q}[\sqrt{-3}]$$

7. We have seen in lectures that, if F is a field, then $F[x]$ is a principal ideal domain. The converse is also true: if R is a commutative ring such that $R[x]$ is a principal ideal domain, then R is a field. Prove this in steps as follows:

   (a) If $R[x]$ has an identity, then so does R.

   (b) If R has zero-divisors, then so does $R[x]$.

   (c) If R is an integral domain and $r \in R \setminus \{0\}$ has no inverse, then $I := \{p(x) \in R[x];\ p(0) \in rR\}$ is a non-principal ideal in $R[x]$. (Compare the example of $\mathbb{Z}[x]$ in the lecture notes.)

Chapter 6

Field Extensions

6.1 Extending a given field

The techniques we have learnt in previous chapters can be put together to construct new examples of fields. The basic idea is the following. Start from a familiar field F (such as $\mathbb{Q}$ or $\mathbb{Z}_p$ for some prime number p). Find a monic irreducible polynomial $p(x) \in F[x]$. Then the principal ideal $I = p(x)F[x]$ is maximal, and so the quotient ring $K = F[x]/I$ is a field.

Examples

1. Let $F = \mathbb{R}$, and let $p(x) = x^2 + 1$, which has no root in $\mathbb{R}$ and so is irreducible in $\mathbb{R}[x]$. As we have already seen, the resulting field $K = \mathbb{R}[x]/(x^2 + 1)\mathbb{R}[x]$ is isomorphic to the field $\mathbb{C}$ of complex numbers.

2. Let $F = \mathbb{Q}$, and let $p(x) = x^2 + 1$, which is irreducible in $\mathbb{Q}[x]$. Then the quotient field $\mathbb{Q}[x]/(x^2 + 1)\mathbb{Q}[x]$ is isomorphic to the field $\mathbb{Q}[i]$ of Gaussian rationals.

3. Let $F = \mathbb{Z}_2$, and let $p(x) = x^2+x+1$ (the only irreducible quadratic polynomial in $\mathbb{Z}_2[x]$. Then the resulting field $\mathbb{Z}_2[x]/(x^2+x+1)\mathbb{Z}_2[x]$ has four elements $0 = 0+I$, $1 = 1+I$, $a = x+I$, $b = (x+1)+I$, where $I = (x^2+x+1)\mathbb{Z}-2[x]$. The addition and multiplication tables can be deduced from the rule that $(x^2+x+1)+I = 0+I$: for example, $a^2 = x^2 + I = x + 1 + I = b$ in this field. In fact, it is easy to check that this field is isomorphic to the field of four elements we saw in an earlier chapter.

In each of these examples, the first field F is (isomorphic to) a subfield of the resulting field K. Indeed, this is a general feature of our construction. To see this, recall that the units in the polynomial ring $F[x]$ are just the nonzero constant polynomials, in other words the units of F. Since the maximal ideal $I = p(x)F[x]$ is not the whole ring, it cannot contain any units, so the homomorphism $f : F \to K$, $f(r) = r + I$, is injective. (Here we are regarding $r \in F$ as a constant polynomial in $F[x]$.)

It follows from the first isomorphism theorem that $\mathrm{Im}(f)$ is isomorphic to F. We we will always identify the subfield $\mathrm{Im}(f)$ of K with F via the isomorphism f. We have thus constructed a field K containing F as a subfield. Furthermore, the larger field K contains an element $\alpha = x + I$ such that $p(\alpha) = p(x) + I = 0 + I$. In other words, α is a root of $p(x)$ in K.

This is a special case of the following result.

Theorem 6.1 (Kronecker's Theorem) *Let F be a field, and $p(x)$ a non-constant polynomial in $F[x]$. Then there exists a field K, containing F as a subfield, such that $p(x)$ has a root in K.*

Proof. If $p(x)$ is irreducible, then the construction described above produces the desired field K. If $p(x)$ is reducible, then it has at least one irreducible factor, $q(x)$ say. Our construction produces a field $K \supset F$ such that K contains a root α of $q(x)$. Since $p(x)$ is a multiple of $q(x)$, α is also a root of $p(x)$.

Let us take a closer look at the field $K = F[x]/I$ that we have constructed, where $I = p(x)F[x]$ and $p(x)$ is a monic irreducible polynomial in $F[x]$. We know that $F \subset K$ and that K contains a root $\alpha = x + I$ of $p(x)$. What are the other elements of K? Of course, each element of K is a coset $a(x)+I = a(\alpha)$ for some $a(x) \in F[x]$. The element $a(x)$ is not unique, but if two elements $a(x), b(x) \in F[x]$ define the same element of K, then $a(x) - b(x) \in I = p(x)F[x]$. Given $a(x) \in F[x]$, there is a unique representation $a(x) = q(x)p(x) + r(x)$ with $r(x) = 0$ or $\deg(r(x)) < \deg(p(x))$. In other words, there is a unique representative $r(x) \in a(x) + I$ with $r(x) = 0$ or $\deg(r(x)) < \deg(p(x))$.

If $\deg(p(x)) = 1$, then this says that every coset of I is represented by an element of F, so the map $f : F \to F/I = K$ is an isomorphism.

If $\deg(p(x)) = n > 1$, then $p(x) = x^n - c(x)$ for some $c(x) \neq 0$ with $\deg(c(x)) < n$. The elements of K correspond to the polynomials of degree less than n in $F[x]$. These form an n-dimensional vector space over F, with basis $B = \{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$. A typical element has the form

$$d(\alpha) = \lambda_0 + \lambda_1\alpha + \cdots + \lambda_{n-1}\alpha^{n-1} = d(x) + I,$$

where $\lambda_0, \ldots, \lambda_{n-1} \in F$, so $d(x) = \lambda_0 + \lambda_1 x + \cdots + \lambda_{n-1}x^{n-1}$ is a polynomial in $F[x]$ of degree at most $n - 1$.

The rule $(d(x) + I) + (e(x) + I) = (d(x) + e(x)) + I$ for adding in $K = F[x]/I$ tells us that we add two such expressions simply by adding the corresponding coefficients of powers of α. Hence the additive group $(K, +)$ is isomorphic to the vector space $(F^n, +)$ of dimension n over F.

How do we multiply elements of K together? Again, the rule $(d(x)+I)(e(x)+I) = (d(x)e(x)) + I$ tells us. However, the polynomial $d(x)e(x)$ may have degree greater than $n - 1$. We can correct that by using the rule $p(\alpha) = 0$ (or $\alpha^n = c(\alpha)$) to replace high powers of α by F-linear combinations of lower powers.

Examples

1. In $\mathbb{C} = \mathbb{R}[x]/(x^2+1)\mathbb{R}[x]$, we write every element (uniquely) in the form $a+bi$, where $i \in \mathbb{C}$ is a root of x^2+1. The elements $1, i$ form a basis of $\mathbb{C}$ as a vector space over $\mathbb{R}$, giving the familiar rule for adding complex numbers:

$$(a+bi)+(c+di) = (a+c)+(b+d)i.$$

Multiplying two complex numbers $(a+bi), (c+di)$ together yields $ac+(ad+bc)i+bdi^2$. To get this into the canonical form, we need to apply the rule $i^2 = -1$ to get the familiar multiplicaiton rule:

$$(a+bi)(c+di) = (ac-bd)+(ad+bc)i.$$

2. Consider the complex number $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$. This is a cube root of unity in $\mathbb{C}$. It is easy to show that ω^2 is the complex conjugate $\overline{\omega} = -\frac{1}{2} - \frac{\sqrt{3}}{2}i$ of ω, and $\omega^3 = 1$. Indeed, $\omega^2+\omega+1 = 0$ in $\mathbb{C}$, so ω is a root of the irreducible polynomial $x^2+x+1 \in \mathbb{Q}[x]$. Hence the set $\mathbb{Q}[\omega]$ of all complex numbers of the form $a+b\omega$, $a, b \in \mathbb{Q}$ forms a subfield of $\mathbb{C}$ isomorphic to the quotient field $\mathbb{Q}[x]/(x^2+x+1)\mathbb{Q}[x]$.

The elements of $\mathbb{Q}[x]$ have the form $a+b\omega$ and addition rule

$$(a+b\omega)+(c+d\omega) = (a+c)+(b+d)\omega.$$

The multiplication rule in $\mathbb{Q}[\omega]$ is derived in the same way as that for $\mathbb{C}$, but using the rule that $\omega^2+\omega+1 = 0$, or alternatively, $\omega^2 = -1-\omega$. Hence

$$(a+b\omega)+(c+d\omega) = (ac-bd)+(ad+bc-bd)\omega.$$

3. The polynomial $p(x) = x^3+x+1 \in \mathbb{Z}_2[x]$ has no roots in $\mathbb{Z}_2$, since an easy calculation shows that $p(0) = p(1) = 1$. Hence $p(x)$ is irreducible in $\mathbb{Z}_2[x]$, so the ideal $I = p(x)\mathbb{Z}_2[x]$ is maximal, and the quotient ring $K = \mathbb{Z}_2[x]/I$ is a field. Since $p(x)$ has degree 3, the field K has precisely $2^3 = 8$ elements: $a+bx+cx^2$, $a, b, c \in \mathbb{Z}_2$.

Different elements are added by adding the coefficients of $1, x$ and x^2 modulo 2. For example $(1+x)+(1+x^2) = x+x^2$. Elements of K are multiplied by expanding brackets and using the rule $x^3 = 1+x$ to eliminate higher powers of x. For example

$$(1+x^2)(x+x^2) = x+x^2+x^3+x^4 = x+x^2+(1+x)+(x+x^2) = 1+x.$$

6.2 Algebraic number fields

A complex number α is said to be *algebraic* if it is a root of some nonzero polynomial $p(x) \in \mathbb{Q}[x]$. Thus, for example, $\sqrt{2}$, i, $1+i$, and $\frac{1}{2} + \frac{\sqrt{3}}{2}i$ are algebraic numbers, being roots of $x^2 - 2$, $x^2 + 1$, $x^4 + 4$ and $x^3 + 1$ respectively.

Theorem 6.2 *Let $\alpha \in \mathbb{C}$ be an algebraic number. Then there is a unique monic irreducible polynomial $m(x) \in \mathbb{Q}[x]$ such that $m(\alpha) = 0$ in $\mathbb{C}$. The smallest subfield of $\mathbb{C}$ containing α is then isomorphic to $\mathbb{Q}[x]/m(x)\mathbb{Q}[x]$.*

Proof. Let $\phi = \phi_\alpha : \mathbb{Q}[x] \to \mathbb{C}$ be the evaluation homomorphism, $\phi(p(x)) = p(\alpha)$. Then $K = \mathrm{Im}(\phi)$ is a subring of $\mathbb{C}$ containing $\phi(1) = 1$, so K is an integral domain. By the First Isomorphism Theorem, $K \cong \mathbb{Q}[x]/I$, where $I = \mathrm{Ker}(\phi)$. Hence I is a prime ideal. Now $I \neq \{0\}$, since by definition α is the root of at least one nonzero polynomial in $\mathbb{Q}[x]$. Hence $I = p(x)\mathbb{Q}[x]$ for some irreducible polynomial $p(x) \in \mathbb{Q}[x]$. We have seen earlier that $p(x)$ is not unique, but that $p(x)\mathbb{Q}[x] = I = q(x)\mathbb{Q}[x]$ if and only if $q(x) = ap(x)$ for some $a \in \mathbb{Q} \setminus \{0\}$. There is a unique choice of a (namely the inverse of the leading coefficent of $p(x)$, such that $m(x) = ap(x)$ is monic, and hence a unique monic irreducible polynomial $m(x) \in \mathbb{Q}[x]$ such that $I = m(x)\mathbb{Q}[x]$. It follows from this that I is maximal, and hence that K is a field.

By definition, $m(\alpha) = \phi(m(x)) = 0$, since $m(x) \in I = \mathrm{Ker}(\phi)$. If $n(x)$ is another monic irreducible polynomial such that $n(\alpha) = 0$, then $n(x) \in I = m(x)\mathbb{Q}[x]$, so $n(x)$ is a multiple of $m(x)$. Since $n(x)$ is irreducible, $n(x) = am(x)$ for some nonzero constant $a \in \mathbb{Q}$. But $m(x)$ and $n(x)$ are both monic, so $a = 1$ and $n(x) = m(x)$. Hence $m(x)$ is the unique monic irreducible polynomial in $\mathbb{Q}[x]$ for which α is a root.

We have also seen that K is a subfield of $\mathbb{C}$ that is isomorphic to $\mathbb{Q}[x]/I = \mathbb{Q}[x]/m(x)\mathbb{Q}[x]$. It is also clear that K contains $\phi(x) = \alpha$, and that, for any $a \in \mathbb{Q}$, K contains $\phi(a) = a$.

Now suppose that L is another subfield of $\mathbb{C}$ containing α. Then $0, 1 \in L$, and an easy inductive argument shows that $n \in L$ for any non-negative integer n. If m, n are nonnegative integers with $n \neq 0$, then $m, n \in L$, so $n^{-1} \in L$, so $\pm mn^{-1} \in L$. Hence $\mathbb{Q} \subset L$. Since also $\alpha \in L$, it follows that $\phi(p(x)) = p(\alpha) \in L$ for any $p(x) \in \mathbb{Q}[x]$, so $K = \mathrm{Im}(\phi) \subset L$.

Hence K is the smallest subfield of $\mathbb{C}$ containing α, as claimed.

Remarks The monic irreducible polynomial $m(x)$ in this theorem is called the *minimal polynomial* of α.

The subfield $K = \mathrm{Im}(\phi_\alpha)$ is usually denoted $\mathbb{Q}[\alpha]$. An example is the field $\mathbb{Q}[i]$ of Gaussian rationals: the minimal polynomial of i is $x^2 + 1$, and we have already seen that $\mathbb{Q}[i] \cong \mathbb{Q}[x]/(x^2 + 1)\mathbb{Q}[x]$.

A subfield of $\mathbb{C}$ of the form $\mathbb{Q}[\alpha]$ for some algebraic number α is called an *algebraic number field*. The study of these fields, there elements and subrings, is a branch of number theory known as *algebraic number theory*.

Not every complex number is algebraic. Indeed, *most* complex numbers are not algebraic, in the following sense. The set of algebraic numbers can be shown to be *countable*, that is, there is a bijection between that set and the set $\mathbb{N}$ of natural numbers. On the other hand, the set $\mathbb{C}$ of complex numbers can be shown to be *uncountable*, which means that it is strictly bigger than any countable set. Complex numbers that are not algebraic are called *transcendental*. Familiar examples of transcendental numbers are π and the base e of the natural logarithms.

Examples

1. Let $\alpha = \cos(2\pi/5) + i\sin(2\pi/5)$. Then $\alpha^5 = 1$, so α is a root of the polynomial $x^5 - 1 \in \mathbb{Q}[x]$, and so it is an algebraic number. However, the polynomial $x^5 - 1$ is not the minimal polynomial of α, because it is reducible. Indeed 1 is a root of $x^5 - 1$, so $x - 1$ is a factor of $x^5 - 1$.

   If we divide $x^5 - 1$ by $x - 1$ (for example, using long division), we find that $x^5 - 1 = (x - 1)(x^4 + x^3 + x^2 + x + 1)$. Since α is not a root of $x - 1$, it must be a root of the other factor, $x^4 + x^3 + x^2 + x + 1$.

   It is not difficult to check that $x^4 + x^3 + x^2 + x + 1$ is irreducble in $\mathbb{Z}[x]$, and hence also in $\mathbb{Q}[x]$ by Gauss' Lemma. (A linear factor of $x^4 + x^3 + x^2 + x + 1$ in $\mathbb{Z}[x]$ would have to be of the form $\pm x \pm 1$, but ± 1 are not roots of $x^4 + x^3 + x^2 + x + 1$. Similarly, a quadratic factor would have to be of the form $\pm x^2 + ax \pm 1$ for some $a \in \mathbb{Z}$, and comparing coefficients shows that $x^4 + x^3 + x^2 + x + 1$ is not the product of two such quadratic factors.)

   It follows that $m(x) = x^4 + x^3 + x^2 + x + 1$ is the minimal polynomial of α, and so $\mathbb{Q}[\alpha] \cong \mathbb{Q}[x]/I$, where I is the principal ideal generated by $m(x)$.

   The elements of $\mathbb{Q}[\alpha]$ can all be expressed as $\mathbb{Q}$-linear combinations of $1, \alpha, \alpha^2$ and α^3, and added as a $\mathbb{Q}$-vector space. The multiplication table is derived from the rule $\alpha^4 = -1 - \alpha - \alpha^2 - \alpha^3$.

2. Consider the polynomial $p(x) = x^2 - x - 1 \in \mathbb{Q}[x]$. By the quadratic formulae, the roots of $p(x)$ are $(1 \pm \sqrt{5})/2$, neither of which is rational. (One of these roots is the *golden ratio*, the limit of the sequence of ratios of successive Fibonacci numbers: $\frac{1}{2}, \frac{2}{3}, \frac{3}{5}, \ldots$.)

   It follows that $p(x)$ is irreducible, so the minimal polynomial of either of its roots. If α is one of these roots, then the resulting field $\mathbb{Q}[\alpha] \subset \mathbb{R}$ is isomorphic to the quotient ring $\mathbb{Q}[x]/(x^2 - x - 1)\mathbb{Q}[x]$. The elements of $\mathbb{Q}[\alpha]$ have the form $a + b\alpha$ for $a, b \in \mathbb{Q}$. Addition is defined in the obvious way, and multiplication is defined using the rule $\alpha^2 = \alpha + 1$.

   This example differs from the previous example, since the algebraic number α is not a root of unity. Indeed, $|\alpha| \neq 1$, so the powers α^n of α either increase unboundedly as $n \to \infty$, or converge to 0 (depending on which root of $p(x)$

is chosen for α). However, all these powers belong to $\mathbb{Q}[\alpha]$, and they can be computed in the form $a + b\alpha$ using the rule $\alpha^2 = 1 + \alpha$:

$\alpha^2 = 1+\alpha$, $\alpha^3 = \alpha+\alpha^2 = 1+2\alpha$, $\alpha^4 = \alpha+2\alpha^2 = 2+3\alpha$, $\alpha^5 = 2\alpha+3\alpha^2 = 3+5\alpha$, and so on. Can you spot a pattern?

Similarly, we can compute the negative powers of α. To do this, divide the equation $\alpha^2 = 1 + \alpha$ by α and rearrange to get $\alpha^{-1} = -1 + \alpha$. Then iterate:

$\alpha^{-2} = -\alpha^{-1}+1 = 2-\alpha$, $\alpha^{-3} = 2\alpha^{-1}-1 = -3+2\alpha$, $\alpha^{-4} = -3\alpha^{-1}+2 = 5-3\alpha$, and so on. Can you spot another pattern?

3. Let α be a cube root of 2 in $\mathbb{C}$. Then α is a root of $x^3 - 2 \in \mathbb{Q}[x]$, and so is an algebraic number. However, there is no root of $x^3 - 2$ in $\mathbb{Q}$, so $x^3 - 2$ is a monic irreducible, and is therefore the minimal polynomial of α. Hence $\mathbb{Q}[\alpha] = \{a+b\alpha+c\alpha^2 \,;\, a, b, c \in \mathbb{Q}\}$ is a subfield of $\mathbb{C}$ that is isomorphic to $\mathbb{Q}[x]/(x^3-2)\mathbb{Q}[x]$. Addition in $\mathbb{Q}[x]$ is defined in the obvious way, and multiplication is defined using the rule that $\alpha^3 = 2$.

   Notice that in this example there are three different choices for α, since there are three distinct cube roots of 2 in $\mathbb{C}$. One of these is real, so the resulting field $\mathbb{Q}[\alpha]$ is contained in $\mathbb{R}$. For the other two choices, $\mathbb{Q}[\alpha] \not\subset \mathbb{R}$ since $\alpha \notin \mathbb{R}$. Thus we get three distinct subfields F_1, F_2 and F_3 of $\mathbb{C}$. While these subfields are not equal, they are isomorphic to each other: $F_1 \cong F_2 \cong F_3$, since each is isomorphic to $\mathbb{Q}[x]/(x^3 - 2)\mathbb{Q}[x]$.

6.3 Finite fields

We have seen that $\mathbb{Z}_p$ is a field for any prime number p. Given any (monic) irreducible polynomial $m(x) \in \mathbb{Z}_p[x]$, Kronecker's Theorem gives an extension of $\mathbb{Z}_p$ in which $m(x)$ has a root - specifically $K = \mathbb{Z}_p[x]/m(x)\mathbb{Z}_p[x]$. As we have seen, the elements of this field can be naturally expressed in the form

$$a_0 + a_1 x + \cdots + a_{d-1}x^{d-1},$$

where $d = \deg(m(x))$ and each $a_i \in \mathbb{Z}_p$. There are p possible values for each a_i, and hence p^d elements in K.

Indeed, the additive group $(K, +)$ is isomorphic to $\mathbb{Z}_p^d$, the d-dimensional vector space over the field $\mathbb{Z}_p$.

Examples

1. $m(x) = x^2+x+1$ has no roots in $\mathbb{Z}_2$, so it is irreducible in $\mathbb{Z}_2[x]$, so $\mathbb{Z}_2[x]/m(x)\mathbb{Z}_2[x]$ is a field of order 4. Its elements are $0, 1, x, 1+x$, with addition and multiplication tables

$+$	0	1	x	$1+x$
0	0	1	x	$1+x$
1	1	0	$1+x$	x
x	x	$1+x$	0	1
$1+x$	$1+x$	x	1	0

$\times$	0	1	x	$1+x$
0	0	0	0	0
1	0	1	x	$1+x$
x	0	x	$1+x$	1
$1+x$	0	$1+x$	1	x

2. Similarly, $m(x) = x^2 + x + 1$ has no roots in $\mathbb{Z}_5$, as we can easily calculate:

x	0	1	2	3	4
$m(x)$	1	3	2	3	1

Hence $m(x)$ is irreducible in $\mathbb{Z}_5[x]$, and $\mathbb{Z}_5[x]/m(x)\mathbb{Z}_5[x]$ is a field of order $5^2 = 25$. Its elements have the form $a = bx$, $a, b \in \mathbb{Z}_5$, with addition defined modulo 5, and multiplication defined modulo 5 using the rule $x^2 = -1 - x = 4 + 4x$. Thus, for example, we could compute

$$(1 + 2x)(2 + 3x) = 1 + 7x + 6x^2 = 1 + 2x + x^2 = 1 + 2x + (4 + 4x) = 5 + 6x = x.$$

3. $m(x) = x^3+x+1$ has no roots in $\mathbb{Z}_2$, so is irreducible in $\mathbb{Z}_2[x]$. Hence $\mathbb{Z}_2[x]/m(x)\mathbb{Z}_2[x]$ is a field of order $2^3 = 8$. Its elements have the form $a + bx + cx^2$, $a, b, c \in \mathbb{Z}_2$, with addition and multiplication defined modulo 2 using the rule $x^3 = 1+x$. For example, we can comupte the inverse of x in this field by multiplying both sides of the equation $x^3 = 1 + x$ by x^{-1} and rearranging, to get $x^{-1} = 1 + x^2$. To double-check:
$$(1 + x^2)x = x + x^3 = x + (1 + x) = 1.$$

We refer to the number of elements in a finite field F as its *order*, denoted $|F|$. Thus the exmaples that we can construct using Kronecker's Theorem each have order of the form p^d, a prime power. It turns out that this is no acident.

Theorem 6.3 *Let F be a finite field. Then there exist a prime number p and a positive integer d such that F has a subfield isomorphic to $\mathbb{Z}_p$, and F has order p^d.*

Proof. Consider the ring homomorphism $f : \mathbb{Z} \to F$ defined inductively by $f(0) = 0_F$, $f(n + 1) = f(n) + 1_F$ and $f(-n) = -f(n)$. The image of this homomorphism is a subring of F containing 1_F, so it is a finite integral domain. But finite integral domains are fields, so $K = \mathrm{Im}(f)$ is a subfield of F.

By the First Isomorphism Theorem, $K \cong \mathbb{Z}/\mathrm{Ker}(f)$, so $\mathrm{Ker}(f)$ is a maximal ideal of $\mathbb{Z}$, and so of the form $p\mathbb{Z}$ for some prime number p. Thus

$$K = \mathrm{Im}(f) \cong \mathbb{Z}/p\mathbb{Z} \cong \mathbb{Z}_p.$$

Note also that the additive group of F is a vector space of dimension 1 over F, where we define scalar multiplication to be the multiplication of F. The rules for scalar multiplication in a vector space are satisfied because of the associativity and distributivity of multiplication in F:

$$x(yz) = (xy)z; \qquad x(y+z) = xy + xz.$$

Since K is a subfield of F, $(F, +)$ is also a vector space over K. Since F is finite, it must be finite-dimensional over K, of dimension d, say. If $B = \{x_1, \ldots, x_d\}$ is a K-basis for F, then the elements of F can be uniquely expressed in the form $a_1x_1 + \cdots + a_dx_d$ with $a_1, \ldots, a_d \in K$. Since K has order p, there are p possible values for each a_i, so p^d different elements in F. Hence $|F| = p^d$, as required.

The converse of this theorem is also true, in a surprisingly strong form.

Theorem 6.4 *Let p be a prime number, and d a positive integer. Then there exists a field of order p^d. Moreover, this field is unique up to isomorphism: if F_1 and F_2 are fields of order p^d, then $F_1 \cong F_2$.*

I will omit the proof of this theorem, but you can probably imagine how it goes. To prove existence, we do a counting argument to show that there is at least one monic irreducible polynomial in $\mathbb{Z}_p$ of degree d. (For example, in the case $d = 2$ there are p^2 monic quadratics $x^2 + ax + b$, of which $p(p+1)/2$ are products $(x+c)(x+d)$ of two linears.)

To prove uniqueness, we check that (i) every field of order p^d contains an element whose minimal polynomial in $\mathbb{Z}_p[x]$ has degree d; and (ii) if $m(x), n(x) \in \mathbb{Z}_p[x]$ are irreducibles of degree d, then $\mathbb{Z}_p[x]/m(x)\mathbb{Z}_p[x] \cong \mathbb{Z}_p[x]/n(x)\mathbb{Z}_p[x]$.

Example Consider the field $\mathbb{Z}_3$. Of the three elements in $\mathbb{Z}_3$, only two of them, 0 and 1, are squares. Thus $x^2 + 1$ has no root in $\mathbb{Z}_3$, so is irreducible. Let $I = (x^2+1)\mathbb{Z}_3[x]$ and let $K = \mathbb{Z}_3[x]/I$ be the field of order 9. Let us denote the element $x + I$ of K by α. Then the elements of K are $a + b\alpha$ for $a, b \in \mathbb{Z}_3$. The addition and multiplication tables in K are given by

$+$	0	1	2	α	$1+\alpha$	$2+\alpha$	2α	$1+2\alpha$	$2+2\alpha$
0	0	1	2	α	$1+\alpha$	$2+\alpha$	2α	$1+2\alpha$	$2+2\alpha$
1	1	2	0	$1+\alpha$	$2+\alpha$	α	$1+2\alpha$	$2+2\alpha$	2α
2	2	0	1	$2+\alpha$	α	$1+\alpha$	$2+2\alpha$	2α	$1+2\alpha$
α	α	$1+\alpha$	$2+\alpha$	2α	$1+2\alpha$	$2+2\alpha$	0	1	2
$1+\alpha$	$1+\alpha$	$2+\alpha$	α	$1+2\alpha$	$2+2\alpha$	2α	1	2	0
$2+\alpha$	$2+\alpha$	α	$1+\alpha$	$2+2\alpha$	2α	$1+2\alpha$	2	0	1
2α	2α	$1+2\alpha$	$2+2\alpha$	0	1	2	α	$1+\alpha$	$2+\alpha$
$1+2\alpha$	$1+2\alpha$	$2+2\alpha$	2α	1	2	0	$1+\alpha$	$2+\alpha$	α
$2+2\alpha$	$2+2\alpha$	2α	$1+2\alpha$	2	0	1	$2+\alpha$	α	$1+\alpha$

and

$\times$	0	1	2	α	$1+\alpha$	$2+\alpha$	2α	$1+2\alpha$	$2+2\alpha$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	α	$1+\alpha$	$2+\alpha$	2α	$1+2\alpha$	$2+2\alpha$
2	0	2	1	2α	$2+2\alpha$	$1+2\alpha$	α	$2+\alpha$	$1+\alpha$
α	0	α	2α	2	$2+\alpha$	$2+2\alpha$	1	$1+\alpha$	$1+2\alpha$
$1+\alpha$	0	$1+\alpha$	$2+2\alpha$	$2+\alpha$	2α	1	$1+2\alpha$	2	α
$2+\alpha$	0	$2+\alpha$	$1+2\alpha$	$2+2\alpha$	1	α	$1+\alpha$	2α	2
2α	0	2α	α	1	$1+2\alpha$	$1+\alpha$	2	$2+2\alpha$	$2+\alpha$
$1+2\alpha$	0	$1+2\alpha$	$2+\alpha$	$1+\alpha$	2	2α	$2+2\alpha$	α	1
$2+2\alpha$	0	$2+2\alpha$	$1+\alpha$	$1+2\alpha$	α	2	$2+\alpha$	1	2α

respectively. The addition table is self-explanatory, while the multiplication table is calculated using the rule $\alpha^2 = 2$.

Now $x^2 + 2$ is not the only (monic) irreducible quadratic in $\mathbb{Z}_3[x]$. For example, $x^2 + x + 2$ is another monic quadratic with no roots in $\mathbb{Z}_3$, so is also irreducible. Therefore $F = \mathbb{Z}_3[x]/(x^2 + x + 2)\mathbb{Z}_3[x]$ is another example of a field of order 9, and F contains a root β of $x^2 + x + 2$. The theorem classifying finite fields says that F is isomorphic to K. Can we find an explicit isomorphism?

One way to do this is to find a root of $x^2 + x + 2$ in K. The quadratic formula tells us that the roots of $x^2 + x + 2$ are $(-1 \pm \sqrt{1^2 - 4.1.2})/2$, or $1 \pm 2\sqrt{2}$ modulo 3. The square roots of 2 in K are precisely $\pm\alpha$, so for example $1 + \alpha = 1 - 2\alpha$ is a root of $x^2 + x + 2$.

If we consider the evaluation homomorphism $\phi_{1+\alpha} : \mathbb{Z}_3[x] \to K$, defined by

$$\phi_{1+\alpha}(p(x)) = p(1 + \alpha),$$

then $\mathrm{Ker}(\phi_{1+\alpha})$ is precisely the principal ideal $(x^2 + x + 2)\mathbb{Z}_3[x]$, so $\phi_{1+\alpha}$ induces an isomorphism from $F = \mathbb{Z}_3[x]/(x^2 + x + 2)\mathbb{Z}_3[x]$ to K, defined by $a + b\beta \mapsto (a + b) + b\alpha$.

In a similar way, $2 + \beta$ is a square root of 2 in F: $(2 + \beta)^2 = 1 + \beta + \beta^2 = 2$, so the inverse isomorphism $K \to F$ is defined by $a + b\alpha \mapsto (a + 2b) + b\beta$.

In $F \cong K$, the group $U(F) = F \setminus \{0\}$ has order $9 - 1 = 8$. The multiplication table for K tells us that the element β $(= 1 + \alpha)$ has order 8, so that the group $U(F)$ consists precisely of the powers of β: $F = \{1, \beta, \beta^2, \beta^3, \beta^4, \beta^5, \beta^6, \beta^7\}$, with $\beta^8 = 1$, so $U(F)$ is isomorphic to the cyclic group of order 8: $(U(F), \times) \cong (\mathbb{Z}_8, +)$. This is not an accident!

Theorem 6.5 *Let F be a finite field of order $N = p^d$. Then $U(F)$ is a cyclic group of order $N - 1$.*

Proof. The group $U(F)$ is a finite abelian group. There is a classification theorem for finite abelian groups which says that any such group is isomorphic to a direct product of cylic groups $\mathbb{Z}_{m(1)} \times \mathbb{Z}_{m(2)} \times \cdots \times \mathbb{Z}_{m(k)}$ for some $k \geq 1$ and positive integers $m(i)$ such that $m(i+1)$ is a muplitple of $m(i)$ for each i.

In particular, if q is a prime number dividing $m(1)$, and $k > 1$, then there are at least q^2 elements of order dividing q in this group: $(am(1)/q, bm(2)/q, 0, \ldots, 0)$, where $0 \leq a \leq q-1$, $0 \leq b \leq q-1$.

But if $\alpha_1, \ldots, \alpha_{q^2}$ are q-th roots of 1 in F, then $x^q - 1$ has q^2 distinct linear factors $x - \alpha_i$ in $F[x]$, and so is divisible by the degree q^2 polynomial

$$(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_{q^2}).$$

But this contradicts the rules for the degree of a product in $F[x]$, so is impossible.

Hence $k = 1$, and $U(F) \cong \mathbb{Z}_{m(1)}$ is a cylic group. Its order is $N - 1$, since every element of F except for 0 is a unit in F.

Corollary 6.6 *Let F be a finite field of order $N = p^d$. Then there exists an element $\alpha \in U(F) = F \setminus \{0\}$ which has order $N - 1$ in $U(F)$.*

Definition Let F be a finite field, and $\alpha \in F \setminus \{0\}$. The order of α in the multiplicative group $U(F)$ is called the *multiplicative order* of α. If the multiplicative order of α is the order $N - 1$ of $U(F)$, then we say that α is a *primitive element* of F. In this case, the units of F are all powers of α. Indeed, we can list all the elements of F as $F = \{0, 1, \alpha, \alpha^2, \ldots, \alpha^{N-2}\}$.

Examples

1. The polynomial $x^2 + x + 2$ has no roots in $\mathbb{Z}_5$, so is irreducible in $\mathbb{Z}_5[x]$. Hence $K = \mathbb{Z}_5[x]/(x^2 + x + 2)\mathbb{Z}_5[x]$ is a field of order $5^2 = 25$.

   The unit group $U(K)$ has order $25 - 1 = 24$, so its elements have orders dividing 24, by Lagrange's Theorem.

   We can compute various powers of x using the rule $x^2 = 3 + 4x$, to find the multiplicative order of x.

   $x^3 = x(3 + 4x) = 3x + 4x^2 = 3x + (2 + x) = 2 + 4x$.

   $x^4 = x(2 + 4x) = 2x + 4x^2 = 2x + (2 + x) = 2 + 3x$.

   $x^6 = (2 + 4x)^2 = 4 + x + x^2 = 4 + x + (3 + 4x) = 2$.

   $x^8 = (2 + 3x)^2 = 4 + 2x + 4x^2 = 4 + 2x + (2 + x) = 1 + 3x$.

   $x^{12} = 2^2 = 4$.

   Since $x^k \neq 1$ in K for any divisor of 24 less than 24 itself, the multiplicative order of x is 24, so x is primitive.

2. In $\mathbb{Z}_7$, the third power of any element is either 0 or ± 1. In particular, 2 has no cube roots in $\mathbb{Z}_7$, so $x^3 + 5$ is irreducible in $\mathbb{Z}_7[x]$. The resulting field $K = \mathbb{Z}_7[x]/(x^3 + 5)\mathbb{Z}_7[x]$ has order $7^3 = 343$, and so its unit group $U(K)$ has order $342 = 2 \times 9 \times 19$.

   The element x of K cannot be primitive, since by construction $x^3 = 2$ in K, so $x^9 = 2^3 = 1$. Hence x has multiplicative order 9 in K. If we want to find a primitive element, we could try some other elements of K. Note that elements of $\mathbb{Z}_7$ have orders dividing 6, so cannot be primitive in K. Hence also elements of the form ax or ax^2, with $a \in \mathbb{Z}_7$, have orders dividing 18. (For example, $(ax)^{18} = a^{18}x^{18} = 1$, since $a^6 = x^9 = 1$.)

   In fact, there can be at most 18 18-th roots of 1 in K, so the 18 elements ax^b, $a \in U(\mathbb{Z}_7)$, $b = 0, 1, 2$, are the only elements of $U(K)$ of orders dividing 18. The other elements have orders which are multiples of 19.

   The next simplest element to try is $y = 1 + x$. We can compute powers of y using $x^3 = 2$:

   $y^3 = 1 + 3x + 3x^2 + x^3 = 3(1 + x + x^2)$,

   $y^6 = 2(1 + x + x^2)^2 = 2(1 + 2x + 3x^2 + 2x^3 + x^4) = 2(5 + 4x + 3x^2)$,

   $y^9 = 6(1 + x + x^2)(5 + 4x + 3x^2) = 6(5 + 2x + 5x^2 + 3x^4) = 6(5 + x + 5x^2)$,

   $y^{18} = (5 + x + 5x^2)^2 = 4 + 3x + 2x^2 + 3x^3 + 4x^4 = 3 + 4x + 2x^2$,

   $y^{19} = (1 + x)(3 + 4x + 2x^2) = 3 + 6x^2 + 2x^3 = 6x^2$.

   Now $6x^2$ has order dividing 18 but not dividing 6 (since $6x^2 \notin \mathbb{Z}_7$).

   $(6x^2)^9 = 6^9x^{18} = 6$, so $6x^2$ has order 18, and hence y has order $18 \times 19 = 342$. Thus y is a primitive element in K.

Remark The unit groups of finite fields are widely used in cryptography - particularly in the construction of error-correcting codes. The most commonly used fields are those of *characteristic* 2, that is, fields $\mathbb{Z}_2[x]/p(x)\mathbb{Z}_2[x]$ for some irreducible $p(x)$ of degree d. This enables the elements of the field to be stored efficiently on a computer, as binary strings of length d, and the Hamming metric gives a natural distance function between field elements. Messages are encoded using a subset of the field, and messages with errors are corrected to the nearest element of this subset.

All this requires efficient computation in the finite field under consideration. To achieve this, one needs to find a primitive element y of the field, and set up a 'discrete logarithm table'

a	0	1	$\cdots$	$p^d - 2$
y^a	1	y	$\cdots$	y^{-1}

To multiply two field elements α, β quickly, one locates them on the second row of the table, and identifies their *logarithms* a, b from the first row of the table. (This means that $\alpha = y^a$ and $\beta = y^b$.) One then adds $a + b$ (modulo $p^d - 1$) and uses the table to find the *antilogarithm* y^{a+b}, which is the desired product $y^{a+b} = y^a y^b = \alpha\beta$.

Similarly, the discrete logarithm table can be used to carry out fast exponentiation in the finite field: $(y^a)^b = y^{ab}$, so to raise an element to its b-th power one finds its logarithm, multiplies by b modulo $p^d - 1$, and then finds the antilogarithm of the result.

Exercises on field extensions

1. Let F be the field $\mathbb{Q}[x]/(x^3+x+1)\mathbb{Q}[x]$. Express each of the following elements of F as $\mathbb{Q}$-linear combinations of $1, x, x^2$:

   (a) x^4; [(b)] x^6; [(c)] x^{-2}; [(d)] $(1+x^2)(2-x-x^2)$.

2. In the field $F = \mathbb{Q}[x]/(x^4+x^3+x^2+x+1)\mathbb{Q}[x]$, find expressions for each of x^{-1}, x^5 and x^{43} of the form $a+bx+cx^2+dx^3$, $a, b, c, d \in \mathbb{Q}$.

3. Let F be the field $\mathbb{Z}_5[x]/(x^2+x+1)\mathbb{Z}_5[x]$. Express each of the following elements of F as $\mathbb{Z}_5$-linear combinations of $1, x$:

   (a) x^3; [(b)] x^{322}; [(c)] $(1+x)^2$; [(d)] $(1+2x)(2+3x)$.

4. Let F be the field $\mathbb{Z}_2[x]/(x^6+x+1)\mathbb{Z}_2[x]$ of order $2^6 = 64$. Calculate x^7, x^9, $x^{14}(=(x^7)^2)$ and $x^{21}(=(x^7)^3)$ in F as $\mathbb{Z}_2$-linear combinations of $1, \ldots, x^5$. Hence find the order of x in the group of units of F.

5. In the field $F = \mathbb{Z}_2[x]/(x^4+x+1)\mathbb{Z}_2[x]$ of order 16, show that x has multiplicative order 15, and hence find all elements of multiplicative order 3 (expressed as $\mathbb{Z}_2$-linear combinations of $1, x, x^2, x^3$).

6. In the field $F = \mathbb{Z}_7[x]/(x^2+x+3)\mathbb{Z}_7[x]$ of order 49, find the multiplicative order of x. Find an element of multiplicative order 8 in F.

7. If F is a field, and $n \geq 2$ an integer, such that the polynomial $p_n(x) = x^{n-1} + x^{n-2} + \ldots + x + 1$ is irreducible in $F[x]$, show that n is a prime number. [Hint: find a factor for $p_n(x)$ if n is composite.]

8. Use the Euclidean Algorithm to find a greatest common divisor of the given elements in the integral domains indicated

   (a) 49349 and 15555 in $\mathbb{Z}$;

   (b) $2x^3+x^2+2x+1$ and $2x^2+7x+3$ in $\mathbb{R}[x]$;

9. Find the minimal polynomials (in $\mathbb{Q}[x]$) of the following complex numbers:

   (a) $1+i$

   (b) $\sqrt{1+\sqrt{2}}$

   (c) $\dfrac{1}{\sqrt{2}} - \dfrac{i}{\sqrt{2}}$