## Contents

1 Group Theory 3

1 Definition, Examples and Elementary Properties 3
  1.1 Definitions 3
  1.2 Examples – Numbers 4
  1.3 Examples – Matrix Groups 4
  1.4 Examples – Permutation Groups 4
  1.5 Elementary Properties – the Cancellation Laws 5
  1.6 Elementary Properties – Orders of Elements 5
  1.7 Examples – Cyclic Groups 6
  1.8 Isomorphisms and Isomorphic groups 6
  1.9 Direct Products and Groups of Order 4 7
  1.10 Examples – the Dihedral Groups 7
  1.11 Generators and Defining Relations 9

2 Subgroups of Groups 10
  2.1 Definition and Examples 10
  2.2 Cosets and Lagrange’s Theorem 12

3 Normal Subgroups and Quotient Groups 14
  3.1 Normal Subgroups 14
  3.2 Classification of Groups of Orders 6 and 8 14
  3.3 Quotient Groups 16
  3.4 Examples of Quotient Groups 16

4 Homomorphisms and the Isomorphism Theorems 17
  4.1 Homomorphisms 17
  4.2 Kernels and Images 18
  4.3 The Isomorphism Theorems 20
Part I

Group Theory

1 Definition, Examples and Elementary Properties

1.1 Definitions

Definition A group is a set $G$ together with a binary operation $\circ : G \times G \to G$ that satisfies the following properties:

(i) (Closure) For all $g, h \in G$, $g \circ h \in G$;
(ii) (Associativity) For all $g, h, k \in G$, $(g \circ h) \circ k = g \circ (h \circ k)$;
(iii) There exists an element $e \in G$ such that:
   (a) (Identity) for all $g \in G$, $e \circ g = g$; and
   (b) (Inverse) for all $g \in G$ there exists $h \in G$ such that $h \circ g = e$.

(Property (i) does not really need stating, because it is implied by the fact that $\circ : G \times G \to G$ is a binary operation on $G$. But it is traditionally the first of the four group axioms, so we have included it here!)

The number of elements in $G$ is called the order of $G$ and is denoted by $|G|$. This may be finite or infinite.

An element $e \in G$ satisfying (iii) of the definition is called an identity element of $G$, and for $g \in G$, an element $h$ that satisfies (iii)(b) of the definition ($h \circ g = e$) is called an inverse element of $g$.

We shall immediately prove two technical lemmas, which are often included as part of the definition of a group. These two proofs need not be memorised!

Lemma 1.1 Let $G$ be a group, let $e \in G$ be an identity element, and for $g \in G$, let $h \in G$ be an inverse element of $g$. Then $g \circ e = g$ and $g \circ h = e$.

Proof: We have $h \circ (g \circ e) = (h \circ g) \circ e = e \circ e = e = h \circ g$. Now let $h'$ be an inverse of $h$. Then multiplying the left and right sides of this equation on the left by $h'$ and using associativity gives $(h' \circ h) \circ (g \circ e) = (h' \circ h) \circ g$. But $(h' \circ h) \circ (g \circ e) = e \circ (g \circ e) = g \circ e$, and $(h' \circ h) \circ g = e \circ g = g$, so we get $g \circ e = g$.

We have $h \circ (g \circ h) = (h \circ g) \circ h = e \circ h = h$, and multiplying on the left by $h'$ gives $(h' \circ h) \circ (g \circ h) = h' \circ h$. But $(h' \circ h) \circ (g \circ h) = e \circ (g \circ h) = g \circ h$ and $(h' \circ h) = e$, so $g \circ h = e$. □

Lemma 1.2 Let $G$ be a group. Then $G$ has a unique identity element, and any $g \in G$ has a unique inverse.

Proof: Let $e$ and $f$ be two identity elements of $G$. Then, $e \circ f = f$, but by Lemma 1.1, we also have $e \circ f = e$, so $e = f$ and the identity element is unique.

Let $h$ and $h'$ be two inverses for $g$. Then $h \circ g = h' \circ g = e$, but by Lemma 1.1 we also have $g \circ h = e$, so

$$h = e \circ h = (h' \circ g) \circ h = h' \circ (g \circ h) = h' \circ e = h'$$

and the inverse of $g$ is unique. □

Definition A group is called abelian or commutative if it satisfies the additional property:
(Commutativity) For all \( g, h \in G \), \( g \circ h = h \circ g \).

We shall now proceed to change notation!

The groups in this course will either be:

- Multiplicative groups, where we omit the \( \circ \) sign (\( g \circ h \) becomes just \( gh \)), we denote the identity element by 1 rather than by \( e \), and we denote the inverse of \( g \in G \) by \( g^{-1} \); or
- Additive groups, where we replace \( \circ \) by +, we denote the identity element by 0, and we denote the inverse of \( g \) by \( -g \).

If there is more than one group around, and we need to distinguish between the identity elements of \( G \) and \( H \) say, then we will denote them by \( 1_G \) and \( 1_H \) (or \( 0_G \) and \( 0_H \)).

Additive groups will always be commutative, but multiplicative groups may or may not be commutative. The default will be to use the multiplicative notation.

The proof of the next lemma is easy and is left as an exercise. From now on, this result will be used freely and without explicit reference.

**Lemma 1.3** Let \( g, h \) be elements of the multiplicative group \( G \). Then \( (gh)^{-1} = h^{-1}g^{-1} \).

1.2 Examples – Numbers

\( \mathbb{Z} \), \( \mathbb{Q} \), \( \mathbb{R} \), and \( \mathbb{C} \) or indeed the elements of any field form a group under addition. We sometimes denote these by \( (\mathbb{Z}, +) \), \( (\mathbb{Q}, +) \), etc.

Now let \( K \) be any field, such as \( \mathbb{Q} \), \( \mathbb{R} \) or \( \mathbb{C} \), and let \( K^* = K \setminus \{0\} \). Then \( K^* \) is a group under multiplication. But note that \( \mathbb{Z}^* = \mathbb{Z} \setminus \{0\} \) is not a group under multiplication, because most elements do not have multiplicative inverses.

All of these groups are abelian.

1.3 Examples – Matrix Groups

Let \( K \) be a field. Then, for any fixed \( n > 0 \), the set of \( n \times n \) invertible matrices with entries in \( K \) forms a group under multiplication. This group is denoted by \( \text{GL}(n, K) \) or \( \text{GL}_n(K) \).

The \( n \times n \) matrices over \( K \) with determinant 1 also form a group denoted by \( \text{SL}(n, K) \) or \( \text{SL}_n(K) \). This is a subgroup of \( \text{GL}(n, K) \).

Another subgroup of \( \text{GL}(n, K) \) is the group \( O(n, K) \), which consists of the orthogonal \( n \times n \) matrices over \( K \); that is, those matrices \( A \) such that \( A^T = A^{-1} \). (Proof left as exercise.)

These groups are not usually abelian. In fact, except for a couple of examples when \( K \) is a small finite field, they are abelian only when \( n = 1 \).

1.4 Examples – Permutation Groups

Let \( X \) be any set, and let \( \text{Sym}(X) \) denote the set of permutations of \( X \); that is, the bijections from \( X \) to itself. Then \( \text{Sym}(X) \) is a group under composition of maps. It is known as the symmetric group on \( X \).

The proof that \( \text{Sym}(X) \) is a group uses results from Foundations. Note that the composition of two bijections is a bijection, and that composition of any maps obeys the associative law.

The identity element of the group is just the identity map \( X \to X \), and the inverse element of a map is just its inverse map.

Let us recall the cyclic notation for permutations. If \( a_1, \ldots, a_r \) are distinct elements of \( X \), then the cycle \( (a_1, a_2, \ldots, a_r) \) denotes the permutation of \( \phi \in X \) with

\[
(i) \quad \phi(a_i) = a_{i+1} \quad \text{for} \quad 1 \leq i < r.
\]
(ii) \( \phi(a_r) = a_1 \), and
(iii) \( \phi(b) = b \) for \( b \in X \setminus \{a_1, a_2, \ldots, a_r\} \).

When \( X \) is finite, any permutation of \( X \) can be written as a product (= composite) of disjoint cycles. Note that a cycle \((a_1)\) of length 1 means that \( \phi(a_1) = a_1 \), and so this cycle can (and normally is) omitted.

For example, if \( X = \{1, 2, 3, 4, 5, 6, 7, 8\} \) and \( \phi \) maps \( 1, 2, 3, 4, 5, 6, 7, 8 \) to \( 5, 8, 6, 4, 3, 1, 2, 7 \), respectively, then \( \phi = (1, 5, 3, 6)(2, 8, 7) \), where the cycle \((4)\) of length 1 has been omitted. We will denote the identity permutation in cyclic notation by ().

Remember that a composite \( \phi_2 \phi_1 \) of maps means \( \phi_1 \) followed by \( \phi_2 \), so, for example, if \( X = \{1, 2, 3\} \), \( \phi_1 = (1, 2, 3) \) and \( \phi_2 = (1, 2) \), then \( \phi_1 \phi_2 = (1, 3) \), whereas \( \phi_2 \phi_1 = (2, 3) \). This example shows that \( \text{Sym}(X) \) is not in general a commutative group. (In fact it is commutative only when \( |X| \leq 2 \)).

The inverse of a permutation can be calculated easily by just reversing all of the cycles. For example, the inverse of \( (1, 5, 3, 6)(2, 8, 7) \) is \((6, 3, 5, 1)(7, 8, 2)\), which is the same as \((1, 6, 3, 5)(2, 7, 8)\). (The cyclic representation is not unique: \((a_1, a_2, \ldots, a_r) = (a_2, a_3, \ldots, a_r, a_1)\), etc.)

### 1.5 Elementary Properties – the Cancellation Laws

**Proposition 1.4** Let \( G \) be any group, and let \( g, h, k \in G \). Then

(i) \( gh = gk \Rightarrow h = k; \) and
(ii) \( hg = kg \Rightarrow h = k. \)

**Proof:** For (i), we have \( gh = gk \Rightarrow g^{-1}gh = g^{-1}gk \Rightarrow h = k \), and (ii) is proved similarly by multiplying by \( g^{-1} \) on the right.

### 1.6 Elementary Properties – Orders of Elements

First some more notation. In a multiplicative group \( G \), we define \( g^2 = gg \), \( g^3 = ggg \), \( g^4 = gggg \), etc. Formally, for \( n \in \mathbb{N} \), we define \( g^n \) inductively, by \( g^1 = g \) and \( g^{n+1} = gg^n \) for \( n \geq 1 \). We also define \( g^0 \) to be the identity element 1, and \( g^{-n} \) to be the inverse of \( g^n \). Then \( g^{m+n} = g^m g^n \) for all \( m, n \in \mathbb{Z} \).

In an additive group, \( g^n \) becomes \( ng \), where \( 0g = 0 \), and \((-n)g = -(ng)\).

**Definition** Let \( g \in G \). Then the order of \( g \), denoted by \( o(g) \) or by \( |g| \), is the least \( n > 0 \) such that \( g^n = 1 \), if such an \( n \) exists. If there is no such \( n \), then \( g \) has infinite order, and we write \( |g| = \infty \).

Note that if \( g \) has infinite order, then the elements \( g^k \) are distinct for distinct values of \( k \), because if \( g^j = g^k \) with \( j < k \), then \( g^{k-j} = 1 \) and \( g \) has finite order.

Similarly, if \( g \) has finite order \( n \), then the \( n \) elements \( g^0 = 1, g^1 = g, \ldots, g^{n-1} = g^{-1} \) are all distinct, and for any \( k \in \mathbb{Z}, g^k \) is equal to exactly one of these \( n \) elements.

**Lemma 1.5** \( |g| = 1 \iff g = 1 \).

**Lemma 1.6** If \( |g| = n \) then, for \( k \in \mathbb{Z}, g^k = 1 \iff n | k \).

(Recall notation: for integers \( j, k \), \( j | k \) means \( j \) divides \( k \).)
1.7 Examples – Cyclic Groups

**Definition** A group $G$ is called cyclic, if it consists of the integral powers of a single element. In other words, $G$ is cyclic if there exists an element $g$ in $G$ with the property that, for all $h \in G$, there exists $k \in \mathbb{Z}$ with $g^k = h$. The element $g$ is called a generator of $G$.

The most familiar examples of cyclic groups are additive groups rather than multiplicative. The group $(\mathbb{Z}, +)$ of integers under addition is cyclic, because every integer $k$ is equal to $k1$, and so 1 is a generator.

If we fix some integer $n > 0$ and let $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ with the operation of addition modulo $n$, then we get a cyclic group of order $n$, where 1 is once again a generator. (Is 1 the only generator? If not, which elements in $\mathbb{Z}_n$ are generators?)

**Lemma 1.7** In an infinite cyclic group, any generator $g$ has infinite order. In a finite cyclic group of order $n$, any generator $g$ has order $n$.

1.8 Isomorphisms and Isomorphic groups

Later on (Section 4), we shall be considering the more general case of homomorphisms between groups, but for now we just introduce the important special case of isomorphisms.

**Definition** An isomorphism $\phi : G \rightarrow H$ between two groups $G$ and $H$ is a bijection from $G$ to $H$ such that $\phi(g_1g_2) = \phi(g_1)\phi(g_2)$ for all $g_1, g_2 \in G$. Two groups $G$ and $H$ are called isomorphic if there is an isomorphism between them. In this case we write $G \cong H$.

The reason this is so important is that isomorphic groups are considered to be essentially the same group – $H$ can be obtained from $G$ simply by relabelling the elements of $G$.

**Exercise** Show that the relationship between groups of being isomorphic satisfies the conditions of an equivalence relation; that is, $G \cong G$, $G \cong H \Rightarrow H \cong G$, and $G \cong H, H \cong K \Rightarrow G \cong K$.

The following proposition says that there is essentially only one cyclic group of each finite order, and one of infinite order. For this reason it is customary to talk about the cyclic group of order $n$, and the infinite cyclic group.

**Proposition 1.8** Any two infinite cyclic groups are isomorphic. For a positive integer $n$, any two cyclic groups of order $n$ are isomorphic.

**Proof:** If $G$ and $H$ are infinite cyclic groups with generators $g$ and $h$, then $G = \{g^k \mid k \in \mathbb{Z}\}$ and $H = \{h^k \mid k \in \mathbb{Z}\}$. We saw in Subsection 1.7 that the elements $g^k$ of $G$ are distinct for distinct $k \in \mathbb{Z}$, and so the map $\phi : G \rightarrow H$ defined by $\phi(g^k) = h^k$ for all $k \in \mathbb{Z}$ is a bijection, and it is easily checked to be an isomorphism.

If $G$ and $H$ are finite cyclic groups of order $n$, then $G = \{g^k \mid k \in \mathbb{Z}_n\}$ and $H = \{h^k \mid k \in \mathbb{Z}_n\}$ and the map $\phi : G \rightarrow H$ defined by $\phi(g^k) = h^k$ for all $k \in \mathbb{Z}_n$ is an isomorphism. □

We denote a cyclic group of order $m$ by $C_m$. Since any two such groups are isomorphic, this notation is effectively unambiguous.

Here is another easy example of a group isomorphism.

**Proposition 1.9** Let $X$ and $Y$ be two sets with $|X| = |Y|$. Then the groups $\text{Sym}(X)$ and $\text{Sym}(Y)$ of all permutations of $X$ and $Y$ are isomorphic.

The notation $\text{Sym}(n)$ or $S_n$ is standard for the symmetric group on a set $X$ with $|X| = n$. By default, we take $X = \{1, 2, 3, \ldots, n\}$.
The following result is often useful; its proof is left as an exercise.

**Proposition 1.10** If \( \phi : G \to H \) is an isomorphism, then \( |g| = |\phi(g)| \) for all \( g \in G \).

### 1.9 Direct Products and Groups of Order 4

**Definition** Let \( G \) and \( H \) be two (multiplicative) groups. We define the *direct product* \( G \times H \) of \( G \) and \( H \) to be the set \( \{(g, h) \mid g \in G, \ h \in H\} \) of ordered pairs of elements from \( G \) and \( H \), with the obvious component-wise multiplication of elements \( (g_1, h_1)(g_2, h_2) = (g_1g_2, h_1h_2) \) for \( g_1, g_2 \in G \) and \( h_1, h_2 \in H \).

It is straightforward to check that \( G \times H \) is a group under this operation. Note that the identity element is \((1_G, 1_H)\), and the inverse of \((g, h)\) is just \((g^{-1}, h^{-1})\).

If the groups are additive, then it is usually called the direct sum rather than the direct product, and written \( G \oplus H \).

A large part of group theory consists of classifying groups with various properties. This means finding representatives of the isomorphism classes of groups with these properties. As an example of this, let us consider groups of order 4. Let \( G = \{1, a, b, c\} \) be such a group.

Let us consider the orders of the elements \( a, b \) and \( c \). These orders cannot be more than 4, because \( |G| = 4 \), and they cannot be 1 by Lemma 1.5. This leaves 2, 3 and 4 as possibilities.

If an element, \( a \), say, has order 4, then \( G = \{1, a, a^2, a^3\} \) is a cyclic group.

Order 3 is impossible. We shall prove later in Proposition 2.11 that the order of an element must divide the order of the group, but we can see this directly by using the cancellation laws (Proposition 1.4). If \( |a| = 3 \), then we can assume that \( a^2 = b \), and \( a^3 = 1 \), so \( G = \{1, a, a^2, c\} \).

Now what can \( ac \) be? \( ac = a \Rightarrow ac = a1 \Rightarrow c = 1 \): wrong! \( ac = c \Rightarrow ac = 1c \Rightarrow a = 1 \): wrong! \( ac = a^2 \Rightarrow a = c \): wrong! \( ac = 1 \Rightarrow ac = a^3 \Rightarrow c = a^2 \): wrong!

Thus either \( G \) is cyclic, or \( a, b \) and \( c \) all have order 2. In that case, we can use the cancellation laws to prove that the product of any two of \( a, b, c \) must be the third; for example, \( ab = ba = c \), \( ca = b \), etc. In fact we can write its complete *multiplication table*:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>1</td>
<td>c</td>
<td>b</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>c</td>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>b</td>
<td>a</td>
<td>1</td>
</tr>
</tbody>
</table>

Since we have completely determined the multiplication table, we can say immediately that any two groups of order four in which the non-identity elements all have order 2 are isomorphic. For example, the direct product \( C_2 \times C_2 \) of two cyclic groups of order 2 has this property. This group is known as the Klein Four Group. Summing up, we have:

**Proposition 1.11** A group of order 4 is isomorphic either to a cyclic group or to a Klein Four Group.

**Exercise** Use Proposition 1.10 to show that a cyclic group of order 4 is not isomorphic to a Klein Four Group.

### 1.10 Examples – the Dihedral Groups

Let \( n \in \mathbb{N} \) with \( n \geq 2 \) and let \( P \) be a regular \( n \)-sided polygon in the plane. The dihedral group of order \( 2n \) consists of the isometries of \( P \). These consist of
(i) $n$ rotations through the angles $2\pi k/n$ ($0 \leq k < n$) about the centre of $P$; and
(ii) $n$ reflections about lines that pass through the centre of $P$, and either pass through a vertex of $P$ or bisect an edge of $P$ (or both).

Unfortunately, some books denote this group by $D_n$ and others by $D_{2n}$, which can be confusing! We shall use $D_{2n}$.

To analyse these groups, it will be convenient to number the vertices in order 1, 2, ..., $n$, and to regard the group elements as permutations of these vertices. Then the $n$ rotations are the powers $a^k$ for $0 \leq k < n$, where $a = (1, 2, 3, \ldots, n)$ is a rotation through the angle $2\pi/n$.

Let $b$ be the reflection through the bisector of $P$ that passes through the vertex 1. Then $b$ interchanges the vertices 2 and $n$ that are adjacent to 1, and similarly it interchanges 3 and $n-1$, 4 and $n-2$, etc., so we can have $b = (2, n)(3, n-1)(4, n-2)\ldots$. For example, when $n = 5$, $b = (2, 5)(3, 4)$ and when $n = 6$, $b = (2, 6)(3, 5)$. Notice that there is a difference between the odd and even cases. When $n$ is odd, $b$ fixes no vertex other than 1, but when $n$ is even, $b$ fixes one other vertex, namely $(n+2)/2$.

Now we can see, either geometrically or by multiplying permutations, that the $n$ reflections of $P$ are the elements $a^kb$ for $0 \leq k < n$. (Remember that $ab$ means first do $b$ and then do $a$.) Thus we have

$$G = \{a^k \mid 0 \leq k < n\} \cup \{a^kb \mid 0 \leq k < n\}$$

(1).

Again, the odd and even cases are slightly different. When $n = 5$, $b, ab, a^2b, a^3b, a^4b$ are equal to $(2,5)(3,4)$, $(1,2)(3,5)$, $(1,3)(4,5)$, $(1,4)(2,3)$, $(1,5)(2,4)$ respectively, which are the reflections through the bisectors of $P$ through vertices 1, 4, 2, 5, 3. When $n = 6$, we have $b = (2,6)(3,5)$, $a^2b = (1,3)(4,6)$, $a^4b = (1,5)(2,4)$, which are reflections through bisectors of $P$ passing through two vertices, whereas $ab = (1,2)(3,6)(4,5)$, $a^3b = (1,4)(2,3)(5,6)$, $a^5b = (1,6)(2,5)(3,4)$, which are reflections through lines that bisect two edges of $P$.

In all cases, we have $ba = a^{n-1}b = a^{-1}b$; this is the reflection that interchanges vertices $i$ and $n+1-i$ for $1 \leq i \leq n$. Hence we also have $ba^k = a^{n-k}b = a^{-k}b$ for $0 \leq k < n$. These equations, together with $a^n = 1$ and $b^2 = 1$ enable us to calculate the full multiplication table of $G$ with the elements written as they are in the expression (1) above, because they enable us to perform any of the four basic types of products:

(i) $(a^k)(a^l) = a^{k+l}$ ($k + l < n$) or $a^{k+l-n}$ ($k + l \geq n$);
(ii) $(a^k)(a^l) = a^{k+l}$ ($k + l < n$) or $a^{k+l-n}$ ($k + l \geq n$);
(iii) $(a^k)(a^l) = a^{k+n-l}b = a^{k+n-l}b$ ($k < l$) or $a^{k+l}b$ ($k \geq l$);
(iv) $(a^k)(a^l) = a^{k+n-l}b$ ($k < l$) or $a^{k-l}$ ($k \geq l$).

Let us write out the full multiplication table in the case $n = 6$. 

8
1.11 Generators and Defining Relations

**Definition** The elements \( \{g_1, g_2, \ldots, g_n\} \) of a group \( G \) are said to *generate* \( G \) (or to form a set of *generators* for \( G \)) if every element of \( G \) can be obtained by repeated multiplication of the \( g_i \) and their inverses.

This means that every element of \( G \) can be written as an expression like \( g_1^2 g_2^{-1} g_1 g_3 g_1^{-1} g_2^{-1} \) in the \( g_i \) and \( g_i^{-1} \), which is allowed to be as long as you like. Such an expression is also called a *word* in the generators \( g_i \) (and their inverses).

**Examples** 1. A group is cyclic if and only if it can be generated by a single element.

2. The dihedral groups are generated by \( a = (1, 2, \ldots, n) \) and \( b = (2, n)(3, n-1) \).

3. We shall see in Lemma 2.3 below that, when \( X \) is finite, any permutation in the symmetric groups \( \text{Sym}(X) \) can be written as a product of transpositions. So the set of all transpositions generates \( \text{Sym}(X) \). (In fact there are much smaller generating sets for \( \text{Sym}(X) \).)

We shall be needing some techniques to prove isomorphisms between groups satisfying various properties. The following results provide methods of doing this for certain families of groups.

**Proposition 1.12** Let \( G \) be a group of order \( 2n \) generated by two elements \( a \) and \( b \) that satisfy the equations \( a^n = 1 \), \( b^2 = 1 \) and \( ba = a^{-1}b \). Then \( G \cong D_{2n} \).

**Proof:** Since \( G \) is generated by \( a \) and \( b \), any element of \( G \) can be written as a product of the generators \( a, b, a^{-1}, b^{-1} \). Since \( a^n = 1 \) and \( b^2 = 1 \), we can always replace \( a^{-1} \) by \( a^{n-1} \) and \( b^{-1} \) by \( b \), so we can assume that only \( a \) and \( b \) appear in this product. Furthermore, we can use the equation \( ba = a^{-1}b = a^{n-1}b \) to move all occurrences of \( a \) in the product to the left of the expression, and we end up with a word in the form \( a^k b^l \). Using \( a^n = b^2 = 1 \) again, we can assume that \( 0 \leq k < n \) and \( 0 \leq l < 2 \). This leaves us with precisely \( 2n \) different words \( a^k b^l \), and since we are told that \( |G| = 2n \), these words must all represent distinct elements of \( G \). We have now shown that \( G = \{a^k \mid 0 \leq k < n\} \cup \{a^k b \mid 0 \leq k < n\} \), exactly as in \( D_{2n} \).

Using \( ba = a^{-1}b \) twice, we get \( ba^2 = (ba)a = a^{-1}ba = a^{-1}a^{-1}b = a^{-2}b \), and similarly \( ba^k = a^{-k}b \) for all \( k \geq 0 \); and since \( a^{-k} = a^{n-k} \), we have \( ba^k = a^{n-k}b \) for \( 0 \leq k < n \). We saw in Subsection 1.10, that these equations, together with \( a^n = 1 \) and \( b^2 = 1 \) allow us to deduce the whole of the multiplication table of \( D_{2n} \). Hence any group satisfying the hypotheses of this proposition has corresponding elements and the same multiplication as \( D_{2n} \), and so it is isomorphic to \( D_{2n} \). \( \square \)

The equations \( \{a^n = 1, b^2 = 1, ba = a^{-1}b\} \) are called *defining relations* for \( D_{2n} \), which means roughly that \( D_{2n} \) is the largest group generated by two elements \( a \) and \( b \) that satisfy these
equations. We shall not go into the general theory of defining relations in this course, but we shall give two more examples.

Consider the direct product $C_m \times C_n$ of cyclic groups of orders $m$ and $n$, which has order $mn$. Let $c$ and $d$ be generators of $C_m$ and $C_n$, and let $a = (c, 1)$ and $b = (1, d)$. Then we can easily check that $a^m = 1$, $b^n = 1$ and $ab = ba$.

**Proposition 1.13** Let $G$ be a group of order $mn$ generated by two elements $a$ and $b$ that satisfy the equations $a^m = 1$, $b^n = 1$ and $ba = ab$. Then $G \cong C_m \times C_n$.

**Proof:** By a similar argument to that used in the proof of Proposition 1.12, we can show that any element of $G$ can be written as $a^kb^l$ with $0 \leq k < m$, $0 \leq l < n$, and hence deduce that $G = \{a^kb^l \mid 0 \leq k < m, 0 \leq l < n\}$.

From $ba = ab$, we deduce $b^la^k = a^kb^l$ for $k, l \geq 0$, and the defining relations now enable us easily to deduce the complete multiplication table of $G$. The result follows. \(\square\)

As a final example, we consider a specific group of order 8, which is known as the quaternion group, and denoted by $Q_8$. There are several ways to define this. One way is as the subgroup of $GL(2, \mathbb{C})$ consisting of the 8 matrices

$$
1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad a = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad a^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad a^3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},
$$

$$
b = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad ab = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \quad a^2b = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}, \quad a^3b = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.
$$

We have denoted the identity element by 1, as usual, but do not confuse this with the element $1 \in \mathbb{C}$!

We can check directly that $a^4 = 1$, $b^2 = a^2$, and $ba = a^{-1}b$. By a similar argument to that which we used for the dihedral groups, we can deduce the full multiplication table of $Q_8$ from these equations, and thus prove:

**Proposition 1.14** Let $G$ be a group of order 8 generated by two elements $a$ and $b$ that satisfy the equations $a^4 = 1$, $b^2 = a^2$ and $ba = a^{-1}b$. Then $G \cong Q_8$.

The multiplication table of $Q_8$ is as follows. Notice that it has only a single element $(a^2)$ of order 2, and 6 elements of order 4.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>a</th>
<th>$a^2$</th>
<th>$a^3$</th>
<th>b</th>
<th>$ab$</th>
<th>$a^2b$</th>
<th>$a^3b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>a</td>
<td>$a^2$</td>
<td>$a^3$</td>
<td>b</td>
<td>$ab$</td>
<td>$a^2b$</td>
<td>$a^3b$</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>$a^2$</td>
<td>$a^3$</td>
<td>1</td>
<td>$ab$</td>
<td>$a^2b$</td>
<td>$a^3b$</td>
<td>$b$</td>
</tr>
<tr>
<td>$a^2$</td>
<td>$a^2$</td>
<td>$a^3$</td>
<td>1</td>
<td>$a$</td>
<td>$a^2b$</td>
<td>$a^3b$</td>
<td>$b$</td>
<td>$ab$</td>
</tr>
<tr>
<td>$a^3$</td>
<td>$a^3$</td>
<td>1</td>
<td>$a$</td>
<td>$a^2$</td>
<td>$a^3b$</td>
<td>$b$</td>
<td>$ab$</td>
<td>$a^2b$</td>
</tr>
<tr>
<td>b</td>
<td>$ab$</td>
<td>$a^3b$</td>
<td>$a^2b$</td>
<td>$ab$</td>
<td>$a^2$</td>
<td>$a$</td>
<td>$1$</td>
<td>$a^3$</td>
</tr>
<tr>
<td>$a^2$</td>
<td>$a^2$</td>
<td>$a^3$</td>
<td>$a^2$</td>
<td>$a$</td>
<td>$1$</td>
<td>$a^3$</td>
<td>$a^2$</td>
<td>$a$</td>
</tr>
<tr>
<td>$a^3$</td>
<td>$a^3$</td>
<td>$a^2b$</td>
<td>$a$</td>
<td>$1$</td>
<td>$a^3$</td>
<td>$a^2$</td>
<td>$a$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

2 Subgroups of Groups

2.1 Definition and Examples

**Definition** A subset $H$ of a group $G$ is called a **subgroup** of $G$ if it forms a group under the same operation as that of $G$. 


Lemma 2.1 If $H$ is a subgroup of $G$, then the identity element $1_H$ of $H$ is equal to the identity element $1_G$ of $G$.

Proposition 2.2 Let $H$ be a non-empty subset of a group $G$. Then $H$ is a subgroup of $G$, if and only if

(i) $h_1, h_2 \in H \implies h_1h_2 \in H$; and
(ii) $h \in H \implies h^{-1} \in H$.

Proof: $H$ is a subgroup of $G$ if and only if the four group axioms hold in $H$. Two of these, ‘Closure’, and ‘Inverses’, are the conditions (i) and (ii) of the lemma, and so if $H$ is a subgroup, then (i) and (ii) must certainly be true. Conversely, if (i) and (ii) hold, then we need to show that the other two axioms, ‘Associativity’ and ‘Identity’ hold in $H$. Associativity holds because it holds in $G$, and $H$ is a subset of $G$. Since we are assuming that $H$ is non-empty, there exists $h \in H$, and then $h^{-1} \in H$ by (ii), and $hh^{-1} = 1 \in H$ by (i), and so ‘Identity’ holds, and $H$ is a subgroup. \qed

Examples 1. There are two standard subgroups of any group $G$: the whole group $G$ itself, and the trivial subgroup $\{1\}$ consisting of the identity alone. Subgroups other than $G$ are called proper subgroups, and subgroups other than $\{1\}$ are called non-trivial subgroups.

2. The non-zero real numbers $\mathbb{R}^*$ form a subgroup of the multiplicative group of non-zero complex numbers $\mathbb{C}^*$. Another subgroup of $\mathbb{C}^*$ consists of the numbers $z \in \mathbb{C}$ with $|z| = 1$.

3. We have seen some examples in Subsection 1.3: $\text{SL}(n, K)$ and $\text{O}(n, K)$ are subgroups of $\text{GL}(n, K)$.

4. If $g$ is any element of any group $G$, then the set of all powers $\{g^x \mid x \in \mathbb{Z}\}$ forms a subgroup of $G$ called the cyclic subgroup generated by $g$.

Let us look at a few specific examples. If $G = (\mathbb{Z}, +)$, then $5\mathbb{Z}$, which consists of all multiples of 5, is the cyclic subgroup generated by 5. Of course, we can replace 5 by any integer here, but note that the cyclic groups generated by 5 and $-5$ are the same.

If $G = \langle g \rangle$ is a finite cyclic group of order $n$, and $m$ is a positive integer dividing $n$, then the cyclic subgroup generated by $g^m$ has order $n/m$ and consists of the elements $g^{mk}$ for $0 \leq k < n/m$.

Exercise What is the order or the cyclic subgroup generated by $g^m$ for general $m$ (where we drop the assumption that $m|n$)?

In the dihedral groups (Subsection 1.10), the $n$ rotations form a cyclic subgroup generated by $a$.

Exercise Show that $\mathbb{C}^*$ has finite cyclic subgroups of all possible orders.

5. (This is an example that was already introduces in Linear Algebra in connection with the definition of the determinant.) Let $G = \text{Sym}(X)$ be the symmetric group on the finite set $X$ (Subsection 1.4). We saw that every permutation in $G$ can be written as a product of disjoint cycles. A cycle of length two is called a transposition. Since an arbitrary cycle like $(1, 2, 3, \ldots, n)$ can be written as a product of transpositions (for example, $(1, 2, 3, \ldots, n) = (1, 2)(2, 3)(3, 4)\ldots(n-1, n)$), we have:

Lemma 2.3 Any permutation on $X$ can be written as a product of transpositions.

A permutation is called even if it is a product of an even number of transpositions, and odd if it is a product of an odd number of transpositions. In order for this to make much sense, we do need to prove the following:
Proposition 2.4  No permutation is both even and odd.

The proof of this proposition is at the end of Part 1 of the notes. It is not very instructive, and need not be memorised.

It is very easy to see that the set $H$ of even permutations forms a subgroup of $\text{Sym}(X)$. This is known as the alternating group on $X$, and is denoted by $\text{Alt}(X)$. If $(x,y)$ is some fixed transposition on $X$ then, for any $g \in \text{Sym}(X)$, one of the two permutations $g$ and $g(x,y)$ is even and the other is odd. Hence $G = H \cup H(x,y)$, so $|G : H| = 2$.

As with $\text{Sym}(X)$, the isomorphism type of $\text{Alt}(X)$ depends only on $|X|$, and the notation $\text{Alt}(n)$ or $A_n$ is standard for the alternating group on a set $X$ of size $n$.

Notice that a cycle of even length is a product of an odd number of transpositions. This makes it easy to decide whether a given permutation in cyclic notation is even or odd (but it is also easy to make mistakes!).

**Exercise** If $X$ is finite, what is the order of $G_Y$ as a function of $|Y|$ and $|X|$?

**Definition** A subgroup of $\text{Sym}(X)$ for some set $X$ is called a permutation group or a group of permutations on $X$.

Lemma 2.5  The intersection of any set of subgroups of $G$ is itself a subgroup of $G$.

### 2.2 Cosets and Lagrange’s Theorem

Throughout this subsection, $H$ will be a subgroup of a group $G$.

**Definition** Let $g \in G$. Then the right coset $Hg$ is the subset $\{hg \mid h \in H\}$ of $G$. Similarly, the left coset $gH$ is the subset $\{gh \mid h \in H\}$ of $G$.

**Note.** In the case of additive groups, we denote the coset by $H + g$ rather than by $Hg$.

**Example** Let $G = D_6$ be the dihedral group of order 6 (Subsection 1.10). Then $G$ consists of the 6 permutations (), (1,2,3), (1,3,2), (1,2), (1,3), (2,3), where () represents the identity permutation.

Let us first choose $H = \{((),(1,2,3),(1,3,2))\}$ to be the cyclic subgroup generated by $a = (1,2,3)$. If we put $b = (2,3)$, then we find that $Hb = \{(1,2),(1,3),(2,3)\}$. In fact any right coset of $H$ is equal to either $H$ itself or to $Hb = G \setminus H$. Furthermore, $bH = Hb$, and indeed $Hg = gH$, for all $g \in G$, so the right and left cosets are the same in this example.

Now let us choose $H = \{(),(2,3)\}$ to be the cyclic subgroup generated by $b = (2,3)$. With $a = (1,2,3)$, we have $Ha = \{(1,2,3),(1,3)\}$ and $Ha^2 = \{(1,3,2),(1,2)\}$, but $aH = \{(1,2,3),(1,2)\}$ and $a^2H = \{(1,3,2),(1,3)\}$, so the right and left cosets are not the same in this case.

Note, however, that we always have $g \in Hg$ and $g \in gH$.

**Proposition 2.6**  The following are equivalent for $g,k \in G$:

(i) $k \in Hg$;
(ii) $Hg = Hk$;
(iii) $kg^{-1} \in H$.

12
Proof: Clearly \( Hg = Hk \implies k \in Hg \), so \((ii) \implies (i)\).

For\((i) \implies (ii), suppose that \( k \in Hg \). Then \( k = hg \) for some fixed \( h \in H \). Multiplying this equation on the left by \( h^{-1} \) gives \( g = h^{-1}k \). Let \( f \in Hg \). Then, for some \( h_1 \in H \), we have \( f = h_1g = h_1h^{-1}k \in Hk \), so \( Hg \subseteq Hk \). Similarly, if \( f \in Hk \), then for some \( h_1 \in H \), we have \( f = h_1k = h_1hg \in Hg \), so \( Hk \subseteq Hg \). Thus \( Hg = Hk \), and we have proved that \((i) \implies (ii)\).

If \( k \in Hg \), then, as above, \( k = hg \), and multiplying this on the right by \( g^{-1} \) gives \( kg^{-1} = h \in H \), so \((i) \implies (iii)\).

Finally, if \( kg^{-1} \in H \), then putting \( h = kg^{-1} \), we have \( hg = k \), so \( k \in Hg \), proving \((iii) \implies (i)\).

\[\square\]

**Corollary 2.7** Two right cosets \( Hg_1 \) and \( Hg_2 \) of \( H \) in \( G \) are either equal or disjoint.

Proof: If \( Hg_1 \) and \( Hg_2 \) are not disjoint, then there exists an element \( k \in Hg_1 \cap Hg_2 \), but then \( Hg_1 = Hk = Hg_2 \) by the proposition.

\[\square\]

**Corollary 2.8** The right cosets of \( H \) in \( G \) partition \( G \).

**Proposition 2.9** If \( H \) is finite, then all right cosets have exactly \(|H|\) elements.

Proof: Since \( h_1g = h_2g \implies h_1 = h_2 \) by the cancellation law, it follows that the map \( \phi : H \to Hg \) defined by \( \phi(h) = hg \) is a bijection, and the result follows.

Of course, all of the above results apply with appropriate minor changes to left cosets.

Corollary 2.8 and Proposition 2.9 together imply:

**Theorem 2.10** (Lagrange’s Theorem) Let \( G \) be a finite group and \( H \) a subgroup of \( G \). Then the order of \( H \) divides the order of \( G \).

**Definition** The number of distinct right cosets of \( H \) in \( G \) is called the index of \( H \) in \( G \) and is written as \(|G : H|\).

If \( G \) is finite, then we clearly have \(|G : H| = |G|/|H|\).

**Exercise** Prove that, if \(|G : H| \) is finite, then \(|G : H| \) is also equal to the number of distinct left cosets of \( H \) in \( G \). This is clear if \( G \) is finite, because both numbers are equal to \(|G|/|H|\), but it is not quite so easy if \( G \) is infinite.

**Proposition 2.11** Let \( G \) be a finite group. Then for any \( g \in G \), the order \(|g|\) of \( g \) divides the order \(|G|\) of \( G \).

Proof: Let \(|g| = n\). We saw in Example 4 of Subsection 2.1 that the powers \( \{g^x \mid x \in \mathbb{Z}\} \) of \( g \) form a subgroup \( H \) of \( G \), and we saw in Subsection 1.6 that the distinct powers of \( g \) are \( \{g^x \mid 0 \leq x < n\} \). Hence \(|H| = n\) and the result follows from Lagrange’s Theorem.

As an application, we can now immediately classify all finite groups whose order is prime.

**Proposition 2.12** Let \( G \) be a group having prime order \( p \). Then \( G \) is cyclic; that is, \( G \cong C_p \).

Proof: Let \( g \in G \) with \( 1 \neq g \). Then \(|g| > 1\), but \(|g|\) divides \( p \) by Proposition 2.11, so \(|g| = p\).

But then \( G \) must consist entirely of the powers \( g^k \) \((0 \leq k < p)\) of \( g \), so \( G \) is cyclic.

This result and Proposition 1.11 provide a classification of all groups of order less than 6. We shall soon (Subsection 3.2) be able to deal with groups of order 6 too.
3 Normal Subgroups and Quotient Groups

3.1 Normal Subgroups

Definition A subgroup $H$ of a group $G$ is called normal in $G$ if the left and right cosets $gH$ and $Hg$ are equal for all $g \in G$.

The standard notation for "$H$ is a normal subgroup of $G"$ is $H \triangleleft G$ or $H \unlhd G$. ($H \triangleleft G$ is sometimes but not always used to mean that $H$ is a proper normal subgroup of $G$—i.e. $H \neq G$.)

Examples 1. The two standard subgroups $G$ and $\{1\}$ of any group $G$ are both normal.

2. Any subgroup of an abelian group is normal.

3. In the example $G = D_6$ in Subsection 2.2, we saw that the subgroup $\{(), (1, 2, 3), (1, 3, 2)\} (= \{1, a, a^2\})$ is normal in $G$, but the subgroup $\{(), (2, 3)\} = \{1, b\}$ is not normal in $G$.

In Example 3, the normal subgroup $H = \{1, a, a^2\}$ has index $|G|/|H| = 6/3 = 2$ in $G$. In fact we have the general result:

Proposition 3.1 If $G$ is any group and $H$ is a subgroup with $|G : H| = 2$, then $H$ is a normal subgroup of $G$.

Proof: Assume that $|G : H| = 2$. Then there are only two distinct right cosets of $G$, one of which is $H$, and so by Corollary 2.8, the other one must be $G \setminus H$. The same applies to left cosets. Hence, for $g \in G$, if $g \in H$ then $gH = Hg = H$ and if $g \not\in H$ then $gH = Hg = G \setminus H$.

In either case $gH = Hg$, so $H \triangleleft G$.

So, in Example 6 of Subsection 2.1, Alt($X$) is a normal subgroup of Sym($X$).

The following result often provides a useful method of testing a subgroup for normality:

Proposition 3.2 Let $H$ be a subgroup of the group $G$. Then $H$ is normal in $G$ if and only if $ghg^{-1} \in H$ for all $g \in G$ and $h \in H$.

Proof: Suppose that $H \triangleleft G$, and let $g \in G$, $h \in H$. Then $Hg = gH$, and $gh \in gH$, so $gh \in Hg$, which means that there exists $h' \in H$ with $gh = h'g$. Hence $ghg^{-1} = h' \in H$.

Conversely, assume that $ghg^{-1} \in H$ for all $g \in G$, $h \in H$. Then for $gh \in gH$, we have $ghg^{-1} \in H$, so $gh = h'g$ for some $h' \in H$; i.e. $gh \in Hg$, and we have shown that $gH \subseteq Hg$.

For $hg \in Hg$, we have $g^{-1}hg \in H$ (because $g^{-1}hg = g^{-1}h^{-1}g$ where $g^{-1} = g^{-1}$), so, putting $h' = g^{-1}hg$, we have $hg = gh' \in gH$, and so $Hg \subseteq Hg$. Thus $gH = Hg$, and $H \triangleleft G$.

Example 4. Let $G$ be the dihedral group $D_{12}$. Recall from Subsection 1.10 that $G = \{1, a, a^2, a^3, a^4, a^5, b, ab, a^2b, a^3b, a^4b, a^5b\}$ where $a$ and $b$ are the permutations (1, 2, 3, 4, 5, 6) and (2, 6)(3, 4); the full multiplication table of this group is written out in Subsection 1.10.

Let $H$ be the cyclic subgroup $\{1, a^3\}$ of order 2 of $G$. We claim that $H$ is normal in $G$. To prove this, we have to show that $ghg^{-1} \in H$ for all $g \in G$, $h \in H$. If $h = 1$, then $ghg^{-1} = gg^{-1} = 1 \in H$, so we only need consider $h = a^3$. Then if $g = a^k$ for $0 \leq k \leq 5$, we have $ghg^{-1} = a^k a^3 a^{-k} = a^3 \in H$. The remaining case is $g = a^k b$ for $0 \leq k \leq 5$.

From Subsection 1.10, we have $ba^k = a^{6-3k} = a^{3k}$ and so $ba^kb^{-1} = a^3$, and then, for any $k$, $(a^k b)(a^3 (a^k b)^{-1}) = a^k ba^3 b^{-1} a^{-k} = a^k a^3 a^{-k} = a^3 \in H$, and we have proved that $H \triangleleft G$.

3.2 Classification of Groups of Orders 6 and 8

The following easy lemma was on an Assignment Sheet:
Lemma 3.3 Let $G$ be a group in which $g^2 = 1$ for all $g \in G$. Then $G$ is abelian.

Lemma 3.4 Let $G$ be a group in which $g^2 = 1$ for all $g \in G$, and let $a, b$ be distinct non-identity elements of $G$. Then $\{1, a, b, ab\}$ is a subgroup of $G$ of order 4.

Proof: We cannot have $ab = 1, a$ or $b$ (exercise), so $|\{1, a, b, ab\}| = 4$. To prove it is a subgroup, we check that (i) and (ii) of Proposition 2.2 hold. Since $g^2 = 1$ for all $g \in G$, we have $g = g^{-1}$ for all $g \in G$, and so (ii) is satisfied. (i) is straightforward, given that $G$ is abelian and $a^2 = b^2 = 1$. For example, $b(ab) = bab = bba = a$. □

We can now classify groups of order 6. We have seen two examples so far, the cyclic group $C_6$ and the dihedral group $D_6$. (Notice that the symmetric group $S_3$ is actually equal to $D_6$.) Since $D_6$ is non-abelian (or, alternatively, since it has no element of order 6), these two groups cannot be isomorphic to each other. We shall now prove that they are the only examples.

Proposition 3.5 Let $G$ be a group of order 6. Then $G \cong C_6$ or $G \cong D_6$.

Proof: By Proposition 2.11 the orders of elements $g \in G$ can be 1, 2, 3 or 6. If there is a $g$ with $|g| = 6$, then $G \cong C_6$, so assume not. If all elements had order 1 or 2, then by the last lemma, $G$ would have a subgroup with four elements, which contradicts Lagrange’s Theorem. Hence there is an element $a$ of order 3. Then $N = \{1, a, a^2\}$ has index 2 in $G$, so it is a normal subgroup by Proposition 3.1. Choose $b \in G \setminus N$. Then $G = N \cup Nb = \{1, a, a^2, b, ab, a^2b\}$.

What can $b^2$ be? $b^2 = b, ab$ or $a^2b$ all lead to contradictions by the cancellation law. If $b^2 = a$ or $a^2$ on the other hand, then $b$ has order 6, contrary to assumption. So $b^2 = 1$.

By Proposition 3.2, we have $bab^{-1} \in N$. If $bab^{-1} = 1$, then $ba = b$ which is false. If $bab^{-1} = a$, then $ba = ab$ and then we find that $ab$ has order 6, contrary to assumption. The remaining possibility is $bab^{-1} = a^2$, and then $ba = a^2b = a^{-1}b$ and $G \cong D_6$ by Proposition 1.12. □

The classification of groups of order 8 takes a little more effort, because there are five isomorphism classes, but it can be done using only the results proved so far.

Proposition 3.6 Let $G$ be a group of order 8. Then $G$ is isomorphic to one of $C_8$, $C_4 \times C_2$, $C_2 \times C_2 \times C_2$, $D_8$ and $Q_8$.

Proof: If $G$ has an element of order 8, then it is cyclic ($C_8$), so assume not. If all non-identity elements have order 2, then $G$ is abelian by Lemma 3.3, and by Lemma 3.4 there is a subgroup $H = \{1, a, b, ab\}$ isomorphic to a Klein Four Group $C_2 \times C_2$. Choose $c \in G \setminus H$. Then $G = H \cup Hc$, and it can be checked that $G \cong H \times \langle c \rangle \cong C_2 \times C_2 \times C_2$.

Otherwise, there is an element $a \in G$ of order 4, and a normal subgroup $N = \{1, a, a^2, a^3\}$. Let $b \in G \setminus N$, so $G = N \cup Nb$. As in Proposition 3.5, we cannot have $b^2 \in Nb$, so $b^2 \in N$. If $b^2 = a$ or $a^2$, then $|b| = 8$, contrary to assumption. Hence $b^2 = 1$ or $a^2$. Since $N$ is normal, $bab^{-1} \in N$. As in Proposition 3.5, we cannot have $bab^{-1} = 1$. If $bab^{-1} = a^2$, then $ba^{-2}b^{-1} = bab^{-1}bab^{-1} = a^2a^{-2} = 1$, and then $a^2 = b^{-1}b = 1$, contradiction, so $bab^{-1} \neq a^2$, and hence $bab^{-1} = a$ or $a^3$; that is, $ba = ab$ or $ba = a^3b = a^{-1}b$. We now have four possibilities to analyse:

(i) $b^2 = 1, ba = ab$. Then $G \cong C_4 \times C_2$ by Proposition 1.13.
(ii) $b^2 = a^2, ba = ab$. In this case, we have $(ab)^2 = a^2b^2 = a^4 = 1$, so if we replace $b$ by $ab$, then we are back in Case (i), and $G \cong C_4 \times C_2$ again.
(iii) $b^2 = 1, ba = a^{-1}b$. $G \cong D_8$ by Proposition 1.12.
(iv) $b^2 = a^2, ba = a^{-1}b$. $G \cong Q_8$ by Proposition 1.14. □
3.3 Quotient Groups

The definition of quotient groups depends on the following crucial technical result.

**Lemma 3.7** Let \( N \) be a normal subgroup of a group \( G \), and let \( g, h \in G \). Then the product of any element in the coset \( Ng \) with any element in the coset \( Nh \) is equal to an element in the coset \( N gh \).

**Proof:** Let \( n_1 g \in Ng \) and \( n_2 h \in Nh \). Then, by normality of \( N \), we have \( gN = Ng \), and so \( g n_2 \) is equal to some element \( n_3 g \in Ng \). Hence \((n_1 g)(n_2 h) = n_1 (g n_2) h = n_1 (n_3 g) h = (n_1 n_3) g h \in N gh\), which proves the lemma. \( \square \)

**Definition** If \( A \) and \( B \) are subsets of a group \( G \), then we define their product \( AB = \{ab \mid a \in A, b \in B\} \).

**Lemma 3.8** If \( N \) is a normal subgroup of \( G \) and \( Ng, Nh \) are cosets of \( N \) in \( G \), then \((Ng)(Nh) = Ngh\).

**Proof:** By Lemma 3.7, we have \((Ng)(Nh) \subseteq Ngh\), but \( ngh = (ng)(1h) \in (Ng)(Nh) \), so \( Ngh \subseteq (Ng)(Nh) \), and we have equality. \( \square \)

**Theorem 3.9** Let \( N \) be a normal subgroup of a group \( G \). Then the set \( G/N \) of right cosets \( Ng \) of \( N \) in \( G \) forms a group under multiplication of sets.

**Proof:** We have just seen that \((Ng)(Nh) = Ngh\), so we have closure, and associativity follows easily from associativity of \( G \). Since \((N1)(Ng) = N1g = Ng\) for all \( g \in G \), \( N1 \) is an identity element, and since \((Ng^{-1})(Ng) = Ng^{-1}g = N1, Ng^{-1}\) is an inverse to \( Ng \) for all cosets \( Ng \). Thus the four group axioms are satisfied and \( G/N \) is a group. \( \square \)

**Definition** The group \( G/N \) is called the quotient group (or the factor group) of \( G \) by \( N \).

Notice that if \( G \) is finite, then \(|G/N| = |G : N| = |G|/|N|\).

So, although the quotient group seems a rather complicated object at first sight, it is actually a smaller group than \( G \). If we can find a normal subgroup \( N \) of a group \( G \), then we have partially succeeded in splitting up the study of \( G \) into the study of the two smaller groups \( N \) and \( G/N \). Groups that cannot be split up in this sense are called simple. More precisely:

**Definition** A group \( G \) with \(|G| > 1\) is called simple if its only normal subgroups are \( G \) and \( \{1\} \).

**Example** Cyclic groups of order \( p \) are simple, because by Lagrange’s Theorem, the only possible orders of their subgroups are 1 and \( p \).

**Proposition 3.10** A simple abelian group is cyclic of prime order.

**Proof:** All subgroups of an abelian group \( G \) are normal, so we have only to find a proper non-trivial subgroup to establish non-simplicity. Let \( G \) be a simple abelian group and choose \( 1 \neq g \in G \). If \(|g|\) is infinite, then the subgroup generated by \( g^2 \) is non-trivial and proper, so \( G \) is not simple. If \(|g|\) is finite but not prime, say \(|g| = l m\), then the subgroup generated by \( g^l \) is proper and non-trivial, so \( G \) is not simple. Hence \(|g| = p \) is prime, and we must have \( \langle g \rangle = G \), or \( \langle g \rangle \) would be a proper non-trivial subgroup. Hence \( G \) is cyclic of prime order. \( \square \)

3.4 Examples of Quotient Groups

1. Let \( G \) be the infinite cyclic group \((\mathbb{Z}, +)\), and let \( N = n\mathbb{Z} \) be its subgroup generated by a fixed positive integer \( n \) – let’s take \( n = 5 \) just to be specific. Now, by using Lemma 2.6 (after
changing from multiplicative to additive notation!), we see that the cosets $5\mathbb{Z}+k$ and $5\mathbb{Z}+j$ of $N$ in $G$ are equal if and only if $k \equiv j \pmod{5}$. So there are only 5 distinct cosets, namely $N = N+0$, $N+1$, $N+2$, $N+3$, $N+4$. It is now clear that $G/N$ is isomorphic to the group $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$ (see Subsection 1.7) under the correspondence $N+i \mapsto i$ for $0 \leq i < 5$.

2. Now let $G = \langle g \rangle$ be finite cyclic, and suppose that $|g| = lm$ is composite. Let $N$ be the normal subgroup $\langle g^m \rangle$. We saw in Example 4 of Subsection 2.1 that $N$ has order $l$ and consists of the elements $\{g^k | 0 \leq k < l\}$. Since all cosets have the form $Ng^k$ for some $k \in \mathbb{Z}$, it is clear that $G/N$ is cyclic and is generated by $Ng$. We can calculate its order as $|G/N| = m$. To see this directly, note that (using Lemma 2.6 again) $Ng^k = Ng^j \Leftrightarrow g^{k-j} \in N \Leftrightarrow m|(k-j)$, and so the distinct cosets are $Ng^k$ for $0 \leq k < m$. In particular, $(Ng)^m = Ng^m = N1$ is the identity element of $G/N$, and $|Ng| = m$.

3. For a more complicated example, we take Example 4 of Subsection 3.1, namely $G = D_{12}$ and $N = \{1, a^3\}$. Then $|G/N| = |G|/|N| = 6$. For $g \in G$, let us denote $Ng$ by $\overline{g}$. (This is a commonly used notation, but you must always keep in mind that $\overline{g} = \overline{h}$ does not necessarily imply that $g = h$!) Then, since $a^3 \in N$, we have

$$\overline{a}^3 = (Na)^3 = Na^3 = N1 = \overline{1}$$

is the identity of $G/N$. We also have $\overline{b}^2 = 1$ and $\overline{ba} = \overline{a}^{-1}\overline{b}$, because these relations are inherited from the corresponding relations of $G$. Thus $G/N$ is a group of order 6 satisfying the three relations $\overline{a}^3 = 1$, $\overline{b}^2 = 1$, $\overline{aba} = \overline{a}^{-1}\overline{b}$, and so by Proposition 1.12 $G/N \cong D_6$.

It might be helpful in understanding this example to see the full multiplication table of $G$ again (we saw it already in Subsection 1.10), but this time with the elements arranged according to their cosets. Notice that all elements in each $2 \times 2$ block of this table lie in the same coset of $N$ in $G$. We can then see the multiplication table of $G/N$ by regarding these $2 \times 2$ blocks as single elements (i.e. cosets) in the quotient group.

<table>
<thead>
<tr>
<th></th>
<th>$N$</th>
<th>$Na$</th>
<th>$Na^2$</th>
<th>$Nb$</th>
<th>$Nab$</th>
<th>$Na^2b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N1$</td>
<td>$a^3$</td>
<td>$a$</td>
<td>$a^2$</td>
<td>$b$</td>
<td>$ab$</td>
<td>$a^2b$</td>
</tr>
<tr>
<td>$N$</td>
<td>$a$</td>
<td>$a^3$</td>
<td>$a^2$</td>
<td>$a^5$</td>
<td>$a^3b$</td>
<td>$a^2b$</td>
</tr>
<tr>
<td>$a^4$</td>
<td>$a^5$</td>
<td>$a^2$</td>
<td>$1$</td>
<td>$a^3$</td>
<td>$a^4b$</td>
<td>$a^5b$</td>
</tr>
<tr>
<td>$a^5$</td>
<td>$a^2$</td>
<td>$a^5$</td>
<td>$1$</td>
<td>$a^4$</td>
<td>$a^6b$</td>
<td>$a^5b$</td>
</tr>
<tr>
<td>$Nb$</td>
<td>$b$</td>
<td>$a^3b$</td>
<td>$a^5b$</td>
<td>$a^3b$</td>
<td>$a^4b$</td>
<td>$a^5b$</td>
</tr>
<tr>
<td>$a^4b$</td>
<td>$a^5b$</td>
<td>$a^3b$</td>
<td>$a^5b$</td>
<td>$a^4b$</td>
<td>$a^5b$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a^5b$</td>
<td>$a^3b$</td>
<td>$a^5b$</td>
<td>$a^4b$</td>
<td>$a^5b$</td>
<td></td>
</tr>
<tr>
<td>$Nab$</td>
<td>$ab$</td>
<td>$a^4b$</td>
<td>$a^6b$</td>
<td>$a^3b$</td>
<td>$a^4b$</td>
<td>$a^5b$</td>
</tr>
<tr>
<td>$a^4b$</td>
<td>$a^6b$</td>
<td>$a^3b$</td>
<td>$a^5b$</td>
<td>$a^4b$</td>
<td>$a^5b$</td>
<td></td>
</tr>
</tbody>
</table>

### 4 Homomorphisms and the Isomorphism Theorems

#### 4.1 Homomorphisms

**Definition** Let $G$ and $H$ be groups. A **homomorphism** $\phi$ from $G$ to $H$ is a map $\phi : G \rightarrow H$ such that $\phi(g_1g_2) = \phi(g_1)\phi(g_2)$ for all $g_1, g_2 \in G$.

A homomorphism $\phi$ is called a **monomorphism** if it is an injection; that is, if $\phi(g_1) = \phi(g_2) \implies g_1 = g_2$.
A homomorphism $\phi$ is called an *epimorphism* if it is a surjection; that is, if $\text{im}(\phi) = H$.

A homomorphism $\phi$ is called an *isomorphism* if it is both a monomorphism and an epimorphism.

**Lemma 4.1** Let $\phi : G \to H$ be a homomorphism. Then $\phi(1_G) = 1_H$ and $\phi(g^{-1}) = \phi(g)^{-1}$ for all $g \in G$.

**Proof:** (Recall that $1_G$ and $1_H$ are the identity elements of $G$ and $H$.) Let $\phi(1_G) = h$. Then $1_H h = h = \phi(1_G) = \phi(1_G 1_G) = \phi(1_G) \phi(1_G) = hh$, so $h = 1_H$ by the cancellation law. Similarly, if $g \in G$ and $\phi(g) = h$, then $\phi(g^{-1}) \phi(g) = \phi(g^{-1} g) = \phi(1_G) = 1_H = h^{-1} h = \phi(g)^{-1} \phi(g)$ so $\phi(g^{-1}) = \phi(g)^{-1}$ by the cancellation law. \( \square \)

**Examples 1.** If $H$ is a subgroup of $G$, then the map $\phi : H \to G$ defined by $\phi(h) = h$ for all $h \in H$ is a monomorphism. It is an isomorphism if $H = G$.

2. If $G$ is an abelian group and $r \in \mathbb{Z}$, then $(gh)^r = g^r h^r$ for all $g, h \in G$, so the map $\phi : G \to G$ defined by $\phi(g) = g^r$ is a homomorphism.

**Warning.** This only works when $G$ is abelian.

3. Let $k$ be a fixed element of a group $G$. Then, for $g, h \in G$, we have $kghk^{-1} = kgk^{-1}khk^{-1}$, so the map $\phi : G \to G$ defined by $\phi(g) = kgk^{-1}$ is a homomorphism. In fact it is an isomorphism, because $kgk^{-1} = khk^{-1} \implies g = h$ by the cancellation laws, and each $h \in G$ is equal to $\phi(k^{-1}hk)$. Notice that if $G$ is abelian, then whatever $k$ we choose, we always get $\phi(g) = g$ for all $g$, so these examples are only interesting for non-abelian groups.

We can deduce the following result from Proposition 1.10.

**Proposition 4.2** If $g, k$ are elements of a group $G$, then $g$ and $kgk^{-1}$ have the same order.

The elements $g$ and $kgk^{-1}$ are called *conjugate* elements. We shall be studying this relationship further in Subsection 5.3.

4. Let $G = \{1, a, b, c\}$ be a Klein Four Group. Define $\phi : G \to G$ by $\phi(1) = 1$, $\phi(a) = b$, $\phi(b) = c$, $\phi(c) = a$. Then it is straightforward to check that $\phi$ is an isomorphism.

**Exercise** Find lots more homomorphisms from $G$ to $G$ in this case, including some that are not isomorphisms.

5. Recall that $GL(n, K)$ is the group of invertible $n \times n$ matrices over the field $K$, and that $K^*$ is the multiplicative group of non-zero elements of $K$. By the product of determinants rule, $\det(AB) = \det(A) \det(B)$, it follows that the map $\phi : GL(n, K) \to K^*$ defined by $\phi(g) = \det(g)$ is a homomorphism.

6. Here is an example of a homomorphism from an additive group to a multiplicative group. Let $G = \mathbb{R}$ and $H = \mathbb{R}^*$, and define $\phi : G \to H$ by $\phi(g) = \exp(g)$. Then $\phi(g_1 + g_2) = \phi(g_1) \phi(g_2)$, which says that $\phi$ is a homomorphism. In fact $\phi$ is a monomorphism but not an epimorphism, because $\text{im}(\phi)$ is the subgroup of positive real numbers.

### 4.2 Kernels and Images

**Definition** Let $\phi : G \to H$ be a homomorphism. Then the *kernel* $\text{ker}(\phi)$ of $\phi$ is defined to be the set of elements of $G$ that map onto $1_H$; that is,

$$\text{ker}(\phi) = \{ g \in G \mid \phi(g) = 1_H \}.$$
Note that by Lemma 4.1 above, \( \ker(\phi) \) always contains \( 1_G \).

**Proposition 4.3** Let \( \phi : G \to H \) be a homomorphism. Then \( \phi \) is a monomorphism if and only if \( \ker(\phi) = \{1_G\} \).

**Proof:** Since \( 1_G \in \ker(\phi) \), if \( \phi \) is a monomorphism, then we must have \( \ker(\phi) = \{1_G\} \). Conversely, suppose that \( \ker(\phi) = \{1_G\} \), and let \( g_1, g_2 \in G \) with \( \phi(g_1) = \phi(g_2) \). Then \( 1_H = \phi(g_1^{-1}g_2) = \phi(g_1^{-1}g_2) \) (by Lemma 4.1), so \( g_1^{-1}g_2 \in \ker(\phi) \) and hence \( g_1^{-1}g_2 = 1_G \) and \( g_1 = g_2 \). So \( \phi \) is a monomorphism.

The following theorem says that the set of normal subgroups of \( G \) is equal to the set of kernels of homomorphisms with domain \( G \).

**Theorem 4.4** (i) Let \( \phi : G \to H \) be a homomorphism. Then \( \ker(\phi) \) is a normal subgroup of \( G \). (ii) Let \( N \) be a normal subgroup of a group \( G \). Then the map \( \phi : G \to G/N \) defined by \( \phi(g) = Ng \) is a homomorphism (in fact an epimorphism) with kernel \( N \).

**Proof:** (i) Checking that \( \ker(\phi) \) is a subgroup of \( G \) is straightforward, using Proposition 2.2. If \( g \in G \) and \( k \in K \), then

\[
\phi(gkg^{-1}) = \phi(g)\phi(k)\phi(g^{-1}) = \phi(g)1_K\phi(g^{-1}) = 1_K,
\]

so \( gkg^{-1} \in \ker(\phi) \) and \( K \leq G \) by Proposition 3.2.

(ii) It is straightforward to check that \( \phi \) is an epimorphism, and \( \phi(g) = 1_H \iff Ng = N1_G \iff g \in N \), so \( \ker(\phi) = N \). The epimorphism defined in (ii) of the above theorem is called the natural or canonical homomorphism/epimorphism from \( G \) to \( G/N \).

The image \( \im(\phi) \) of a homomorphism is just its image as a map, and the following proposition is straightforward to prove.

**Proposition 4.5** Let \( \phi : G \to H \) be a homomorphism. Then \( \im(\phi) \) is a subgroup of \( H \) (but not necessarily a normal subgroup).

**Example 7.** Let \( G = H = D_{12} \), the dihedral group of order 12. We saw in Subsection 1.10 that \( G = \{ a^k \mid 0 \leq k < 6 \} \cup \{ a^6b \mid 0 \leq k < 6 \} \). We define \( \phi : G \to H \) by \( \phi(a^k) = a^{2k} \) and \( \phi(a^k b) = a^{2k}b \) for \( 0 \leq k < 6 \). We claim that \( \phi \) is a homomorphism. It seems at first sight as though we need to check that \( \phi(gh) = \phi(g)\phi(h) \) for all 144 ordered pairs \( g, h \in G \), but we can group these tests into the four distinct types listed in Subsection 1.10. We will make free use of the fact that \( a^m = 1 \) when \( 6 | m \).

(i) \( \phi(a^ka^l) = \phi(a^{k+l}) \) or \( \phi(a^{k+l-6}) = a^{2(k+l)} \) or \( a^{2(k+l-6)} = a^{2k}a^{2l} = \phi(a^k)\phi(a^l) \);
(ii) \( \phi(a^ka^l) = \phi(a^k)\phi(a^l) - \) this is similar to (i);
(iii) \( \phi((a^k b)a^l) = \phi(a^{k-l}) \) or \( \phi(a^{k-l+6}) = a^{2(k-l)}b \) or \( a^{2(k-l+6)}b = a^{2k}a^{-2}b = a^{2k}ba^{2l} = \phi(a^k b)\phi(a^l) \);
(iv) \( \phi((a^k b)(a^l b)) = \phi(a^k b)\phi(a^l b) - \) this is similar to (iii).

So \( \phi \) really is a homomorphism. We can check that the only elements of \( G \) with \( \phi(g) = 1 \) are \( g = 1 \) and \( g = a^3 \), so \( \ker(\phi) = \{1, a^3\} \), which is the normal subgroup that we considered in Example 3 of Subsection 3.4. \( \im(\phi) \) consists of the 6 elements \( 1, a^2, a^4, b, a^2b, a^4b \) of \( G \).

In general, if \( \phi : G \to H \) is a homomorphism and \( J \) is a subset of \( H \), then we define the complete inverse image of \( J \) under \( \phi \) to be the set \( \phi^{-1}(J) = \{ g \in G \mid \phi(g) \in J \} \). It is easy to check, using Proposition 2.2, that if \( J \) is a subgroup of \( H \), then \( \phi^{-1}(J) \) is a subgroup of \( G \).
In the case where \( \phi \) is the canonical epimorphism \( G \to G/N \) of Theorem 4.4(ii), \( \phi^{-1}(J) \) is a union of cosets of \( N \), and if \( J \) is a subgroup, then \( \phi^{-1}(J) \) is a subgroup of \( G \) containing \( N \). Conversely, if \( I \) is a subgroup of \( G \) containing \( N \), then \( J = I/N \) is a subgroup of \( G/N \) with \( \phi^{-1}(J) = I \), and so we get the following proposition, which is useful when working with quotient groups. We shall use it in the proof of Sylow's Theorem (5.14) later on.

**Proposition 4.6** If \( N \leq G \), then the subgroups of \( G/N \) are precisely the quotient groups \( I/N \), for subgroups \( I \) of \( G \) that contain \( N \).

### 4.3 The Isomorphism Theorems

**Theorem 4.7** (First Isomorphism Theorem) Let \( \phi : G \to H \) be a homomorphism with kernel \( K \). Then \( G/K \cong \text{im}(\phi) \). More precisely, there is an isomorphism \( \overline{\phi} : G/K \to \text{im}(\phi) \) defined by \( \overline{\phi}(Kg) = \phi(g) \) for all \( g \in G \).

**Proof:** The trickiest point to understand in this proof is that we have to show that \( \overline{\phi}(Kg) = \phi(g) \) really does define a map from \( G/K \) to \( \text{im}(\phi) \). The reason that this is not obvious is that we can have \( Kg = Kh \) with \( g \neq h \), and when that happens we need to be sure that \( \phi(g) = \phi(h) \). This is called checking that the map \( \overline{\phi} \) is well-defined. In fact, once you have understood what needs to be checked, then doing it is quite easy, because \( Kg = Kh \Rightarrow g = kh \) for some \( k \in K = \ker(\phi) \), and then \( \phi(g) = \phi(k)\phi(h) = \phi(h) \).

Clearly \( \text{im}(\overline{\phi}) = \text{im}(\phi) \), and it is straightforward to check that \( \overline{\phi} \) is a homomorphism. Finally, \( \overline{\phi}(Kg) = 1_H \iff \phi(g) = 1_H \iff g \in K \iff Kg = K1 = 1_{G/K} \), and so \( \overline{\phi} \) is a monomorphism by Proposition 4.3. Thus \( \overline{\phi} : G/K \to \text{im}(\phi) \) is an isomorphism, which completes the proof.

Let us illustrate this theorem using Example 6 from the last subsection. Note that the elements of \( G = D_{12} \) are listed in two separate columns in the diagram, in different orders, once for the domain and once for the codomain of \( \phi \). The elements of \( \text{im}\phi \) are printed slightly to the left of those not in \( \text{im}(\phi) \) in the codomain column.

\[
\begin{align*}
1 & \quad \{ a^3 \} = N \\
\phi & \quad \{ a \} \\
\phi^{-1}(a) & \quad \{ a^2 \} \\
\phi^{-1}(a^2) & \quad \{ a^3 \} \\
\phi^{-1}(a^3) & \quad \{ a^4 \} \\
\phi^{-1}(a^4) & \quad \{ a^5 \} \\
\phi^{-1}(a^5) & \quad \{ a^6 \} \\
\phi^{-1}(a^6) & \quad \{ b \} \\
\phi^{-1}(b) & \quad \{ ab \} \\
\phi^{-1}(ab) & \quad \{ a^2b \} \\
\phi^{-1}(a^2b) & \quad \{ a^3b \} \\
\phi^{-1}(a^3b) & \quad \{ a^4b \} \\
\phi^{-1}(a^4b) & \quad \{ a^5b \} \\
\phi^{-1}(a^5b) & \quad \{ a^6b \}
\end{align*}
\]

The other two isomorphism theorems are less important, and are used mainly in more advanced courses on group theory. Before we can state the Second Isomorphism Theorem, we
need a lemma. Recall from Subsection 3.7 that the product of two subsets \( A \) and \( B \) of \( G \) is
defined as \( AB = \{ ab \mid a \in A, b \in B \} \). This is not usually a subgroup of \( G \), even when \( A \) and \( B \)
are subgroups (can you find an example to demonstrate this?). However, we have:

**Lemma 4.8** If \( H \) is any subgroup and \( K \) is a normal subgroup of a group \( G \), then \( HK = KH \)
is a subgroup of \( G \).

**Proof:** Let \( hk \in HK \). Then, by normality of \( K \), \( hk \in hK = Kh \subseteq KH \) so \( HK \subseteq KH \)
and similarly \( KH \subseteq HK \), and we have equality. The product of two elements of \( HK \) lies
in \( HKHK = HHKK = HK \), and the inverse \( k^{-1}h^{-1} \) of an element \( hk \in HK \) lies in
\( KH = HK \), so \( HK \) is a subgroup of \( G \) by Proposition 2.2. \( \square \)

**Theorem 4.9** (Second Isomorphism Theorem) Let \( H \) be any subgroup and let \( K \) be a normal
subgroup of a group \( G \). Then \( H \cap K \) is a normal subgroup of \( H \) and \( H/(H \cap K) \cong HK/K \).

**Proof:** Use Proposition 3.2 to show that \( H \cap K \subseteq H \). Let \( \phi : G \to G/K \) be the natural map
(see Theorem 4.4(ii)). Then \( \phi(H) \) is the set of cosets \( Kh \) for \( h \in H \), which together form
the subgroup \( KH/K = HK/K \) of \( G/K \); in other words \( \text{im}(\phi_H) = HK/K \). Also \( \ker(\phi_H) =
H \cap \ker(\phi) = H \cap K \). Now, by applying Theorem 4.7 to \( \phi_H \), we get \( H/(H \cap K) \cong HK/K \). \( \square \)

**Example** Let \( K \) and \( H \) be respectively the subgroups \( \{1, a^2, a^4\} \) and \( \{1, a^3, b, a^3b\} \) of \( G = D_{12} \). Then \( K \cong C_3 \) and \( H \) is a Klein Four Group. \( K \) is a normal subgroup: to check this,
use Proposition 3.2; for example,
\[
(a^k b)a^2(a^k b)^{-1} = a^k ba^2b^{-1}a^{-k} = a^k a^4a^{-k} = a^4
\]
for all \( k \in \mathbb{Z} \). Clearly \( H \cap K = \{1\} \) is trivial, and so \( H/(H \cap K) \cong H \) is also a Klein Four
Group. By direct calculation, we find \( HK = G \), so \( HK/K = G/K \) is a Klein Four Group.

**Theorem 4.10** (Third Isomorphism Theorem) Let \( K \subseteq H \subseteq G \), where \( H \) and \( K \) are both
normal subgroups of \( G \). Then \( (G/K)/(H/K) \cong G/H \).

**Proof:** Define \( \phi : G/K \to G/H \) by \( \phi(Kg) = Hg \) for all \( g \in G \). As in the proof of
Theorem 4.7, we need to check that this is well-defined; that is, that \( Kg_1 = Kg_2 \implies \phi(g_1) = \phi(g_2) \).
This is easy, because \( K \subseteq H \), so \( Kg_1 = Kg_2 \implies Hg_1 = Hg_2 \).
Since \( \text{im}(\phi) = G/H \) and \( \ker(\phi) = H/K \), the result follows by applying Theorem 4.7 to \( \phi \). \( \square \)

5 Groups Acting On Sets

5.1 Definition and Examples

**Definition** Let \( G \) be a group and \( X \) a set. An action of \( G \) on \( X \) is a map \( \cdot : G \times X \to X \),
which satisfies the properties:

(i) \( (gh) \cdot x = g \cdot (h \cdot x) \) for all \( g, h \in G \), \( x \in X \);

(ii) \( 1_G \cdot x = x \) for all \( x \in X \).

**Note 1.** In this definition, the image of \( (g, x) \) under the map \( \cdot \) is denoted by \( g \cdot x \).

2. We have actually defined a left action of \( G \) on \( X \). A right action can be defined analogously
as a map \( X \times G \to X \).

**Proposition 5.1** Let \( \cdot \) be an action of the group \( G \) on the set \( X \). For \( g \in G \), define the
map \( \phi(g) : X \to X \) by \( \phi(g)(x) = g \cdot x \). Then \( \phi(g) \in \text{Sym}(X) \), and \( \phi : G \to \text{Sym}(X) \) is a
homomorphism.
There was one example with two orbits. Let
In most of the examples that we have seen so far, there is just one orbit, the whole of
only if there exists a \( g \) (Cayley’s Theorem)
Theorem 5.3
by the cancellation law, so the action is faithful. From Proposition 5.2, we can now deduce:
\[ \text{as an exercise. The equivalence classes of} \]
Definition 5.2 Orbits and Stabilisers
Sym(\( X \))
That is, to a subgroup of
The action is said to be faithful if \( K = \{1\} \). In this case, Theorem 4.7 says that \( G \cong G/K \cong \text{im}(\phi) \), which we state as a proposition:

Proposition 5.2 If \( \cdot \) is a faithful action of \( G \) on \( X \), then \( G \) is isomorphic to a subgroup of \( \text{Sym}(X) \).

Examples 1. If \( G \) is a subgroup of \( \text{Sym}(X) \), then we can define an action of \( G \) on \( X \) simply by putting \( g \cdot x = g(x) \) for \( x \in X \). This action is faithful.

2. Let \( P \) be a regular hexagon, and let \( G = \{1, a, a^2, a^3, a^4, a^5, b, ab, a^2b, a^3b, a^4b, a^5b\} = D_{12} \)
be the group of isometries of \( P \). In Subsection 1.10, we defined \( a \) and \( b \) to be the permutations
(1, 2, 3, 4, 5, 6) and (2, 6)(3, 5) of the set \( \{1, 2, 3, 4, 5, 6\} \) of vertices of \( P \), and so this immediately
gives us an action of \( G \) on the vertex set.

There are some other related actions however. We could instead take \( X \) to be the set \( E = \{e_1, e_2, e_3, e_4, e_5, e_6\} \) of edges of \( P \), where \( e_1 \) is the edge joining 1 and 2, \( e_2 \) joins 2 and 3, etc. The map \( \phi \) of the action of \( G \) on \( E \) is then given by \( \phi(a) = (e_1, e_2, e_3, e_4, e_5, e_6) \),
\( \phi(b) = (e_1, e_6)(e_2, e_5)(e_3, e_4) \). (Notice that any homomorphism is fully specified by the images
of a set of group generators, because the images of all other elements in the group can be
calculated from these.) This action is still faithful.

As a third possibility, let \( D = \{d_1, d_2, d_3\} \) be the set of diagonals of \( P \), where \( d_1 \) joins vertices
1 and 4, \( d_2 \) joins 2 and 5, and \( d_3 \) joins 3 and 6. Then map \( \phi \) of the action of \( G \) on \( D \) is defined by \( \phi(a) = (d_1, d_2, d_3) \), and \( \phi(b) = (d_2, d_3) \). This action is not faithful, and its kernel is the
normal subgroup \( \{1, a^3\} \) of \( G \) that we have already studied. The image is isomorphic to \( D_6 \).

3. There is a faithful action called the left regular action, which we can define for any group
\( G \). Here we put \( X \) to be the underlying set of \( G \) and simply define \( g \cdot x \) to be \( gx \) for all \( g \in G, x \in X \). Conditions (i) and (ii) of the definition obviously hold, so we have defined an
action. If \( g \) is in the kernel \( K \) of the action, then \( gx = x \) for all \( x \in X \), which implies \( g = 1 \)
by the cancellation law, so the action is faithful. From Proposition 5.2, we can now deduce:

Theorem 5.3 (Cayley’s Theorem) Every group \( G \) is isomorphic to a permutation group.
(That is, to a subgroup of \( \text{Sym}(X) \) for some set \( X \).)

5.2 Orbits and Stabilisers

Definition Let \( \cdot \) be an action of \( G \) act on \( X \). We define a relation \( \sim \) on \( X \) by \( x \sim y \) if and
only if there exists a \( g \in G \) with \( y = g \cdot x \). Then \( \sim \) is an equivalence relation – the proof is left
as an exercise. The equivalence classes of \( \sim \) are called the orbits of \( G \) on \( X \). In particular,
the orbit of a specific element \( x \in X \), which is denoted by \( G \cdot x \) or by \( \text{Orb}_G(x) \) is
\[ \{ y \in X \mid \exists g \in G \text{ with } g \cdot x = y \} \].

In most of the examples that we have seen so far, there is just one orbit, the whole of \( X \).
There was one example with two orbits. Let \( X \) be a set, and \( Y \) a proper non-empty subset
of $X$. In Example 7 of Subsection 2.1, we defined the subgroup $\text{Sym}(X)^Y$ of $\text{Sym}(X)$ to be \{ $g \mid g \in \text{Sym}(X)$, $g(y) \in Y$ and $g^{-1}(y) \in Y \ \forall y \in Y$ \}. Denote this subgroup by $G$, and give it the obvious action on $X$ with $g \cdot x = g(x)$ for $g \in G$, $x \in X$. Then there are two orbits, $Y$ and $X \setminus Y$.

In a similar way, for any partition of $X$, we can can define a subgroup of $\text{Sym}(X)$ having the sets in this partition as the orbits.

**Definition** Let $G$ act on $X$ and let $x \in X$. Then the stabiliser of $x$ in $G$, which is denoted by $G_x$ or by $\text{Stab}_G(x)$ is \{ $g \in G \mid g \cdot x = x$ \}.

The proof of the following proposition is left as an exercise.

**Proposition 5.4** Let $G$ act on $X$ and $x \in X$. Then

(i) $\text{Stab}_G(x)$ is a subgroup of $G$;

(ii) $\cap_{x \in X} \text{Stab}_G(x)$ is the kernel of the action of $G$ on $X$.

The next theorem is a very fundamental result in group theory.

**Theorem 5.5** (The Orbit-Stabiliser Theorem) Let the finite group $G$ act on $X$, and let $x \in X$. Then $|G| = |\text{Orb}_G(x)||\text{Stab}_G(x)|$.

**Proof:** Let $y \in \text{Orb}_G(x)$. Then there exists a $g \in G$ with $g \cdot x = y$. Let $H = \text{Stab}_G(x)$. For an element $g' \in G$, we have

$$g' \cdot x = y \iff g' \cdot x = g \cdot x \iff g^{-1}g' \cdot x = x \iff g^{-1}g' \in H \iff g' \in gH.$$ 

So the elements $g'$ with $g' \cdot x = y$ are precisely the elements of the coset $gH$. But, by Proposition 2.9 (applied to left rather than to right cosets!), we have $|gH| = |H|$. In other words, for each $y \in \text{Orb}_G(x)$, there are precisely $|H|$ elements $g'$ of $G$ with $g' \cdot x = y$. Hence the total number of such $y \in \text{Orb}_G(x)$ must be $|G|/|H|$, which proves the result. $\square$

### 5.3 Conjugacy Classes

In Example 3 of Subsection 5.1, a group $G$ was made to act on the set of its own elements by multiplication on the left; that is, $g \cdot x = gx$ for $g, x \in G$.

There is another important action of $G$ on $X$ = $G$, which is defined by $g \cdot x = g x g^{-1}$ for $g, x \in G$.

It is easy to check that conditions (i) and (ii) of the definition hold, so this does indeed define an action. This action is called *conjugation*. The orbits of the action are called the *conjugacy classes* of $G$, and elements in the same conjugacy class are said to be *conjugate* in $G$. So $g, h \in G$ are conjugate if and only if there exists $f \in G$ with $h = f g f^{-1}$. We will write $\text{Cl}_G(g)$ for the orbit of $g$; that is the conjugacy class containing $g$. We have seen already in Proposition 4.2 that conjugate elements have the same order.

What is $\text{Stab}_G(g)$ for this action? By definition it consists of the elements $f \in G$ for which $f \cdot g = g$; that is, $f g f^{-1} = g$, or equivalently $f g = g f$. In other words, it consists of those $f$ that commute with $g$. It is called the *centraliser of $g$ in $G$* and is written as $C_G(g)$.

By applying the Orbit-Stabiliser Theorem (5.5), we get:

**Proposition 5.6** Let $g \in G$. Then $|\text{Cl}_G(g)| = |G|/|C_G(g)|$. 

23
The kernel $K$ of the action consists of those $f \in G$ that fix and hence commute with all $g \in G$. This is called the centre of $G$ and is denoted by $Z(G)$. So we have

$$Z(G) = \{ f \in G \mid fg = gf \quad \forall g \in G \}.$$ 

Note that $g \in Z(G)$ if and only if $\text{Cl}_G(g) = \{g\}$.

It is high time that we worked out some examples!

**Examples 1.** Let $G$ be an abelian group. Then $Z(G) = G$, $C_G(g) = G$ and $\text{Cl}_G(g) = \{g\}$ for all $g \in G$.

2. Let $G = \{a^k \mid 0 \leq k < n\} \cup \{a^k b \mid 0 \leq k < n\}$ be the dihedral group $D_{2n}$. Then of course, as always, $\text{Cl}_G(1) = \{1\}$. To compute the remaining classes, we calculate $fgf^{-1}$ in the four cases, when $f$ and $g$ have the form $a^k$ ($1 \leq k < n$) and $a^k b$ ($0 \leq k < n$). Since $b^2 = 1$, we have $b = b^{-1}$, so we shall always substitute $b$ for $b^{-1}$ in our calculations.

(i) $f = a^l$, $g = a^k$: $fgf^{-1} = g$.
(ii) $f = a^l b$, $g = a^k$: $fgf^{-1} = ba^k b = a^{-k} = g^{-1}$.
(iii) $f = a^l$, $g = a^k b$: $fgf^{-1} = a^{k+l} b a^{-l} = a^{k+l} a^{-l} b = a^{k+2l} b$.
(iv) $f = a^l b$, $g = a^k b$: $fgf^{-1} = a^l b a^k b a^{-l} = a^l b a^k a^{-l} = a^l a^{-k} b = a^{2l-k} b$.

The cases when $n$ is odd and even are different. Suppose first that $n$ is odd. Then, by (i) and (ii), $a$ and $a^{-k} = a^{n-k}$ are conjugate for all $k$, and we have the distinct conjugacy classes $\{a^k, a^{n-k}\}$ for $1 \leq k \leq (n-1)/2$, all of which contain just two elements. By (iii), we see that $b$ is conjugate to $a^{2l} b$ for $0 \leq l < n$, and when $n$ is odd, this actually includes all elements $a^l b$ for $0 \leq l < n$. (For example, $ab = a^{2l} b$ with $l = (n+1)/2$.) So the set $\{a^k b \mid 0 \leq k < n\}$ forms a single conjugacy class. Geometrically, this is not surprising, because these $n$ elements are all reflections that pass through one vertex and the centre of the polygon $P$ of which $G$ is the group of isometries.

Now suppose that $n$ is even. Then, when $k = n/2$, we have $a^k = a^{-k}$, and so $\{a^{n/2}\}$ is a conjugacy class of size 1 (and hence $a^{n/2} \in Z(G)$). We also have the classes $\{a^k, a^{n-k}\}$ of size 2 for $1 \leq k \leq (n-2)/2$. In this case, the reflections $a^k b$ split up into two conjugacy classes of size $n/2$, namely $\{a^{2k} b \mid 0 \leq k < n/2\}$ and $\{a^{2k+1} b \mid 0 \leq k < n/2\}$. Geometrically these correspond to the two different types of reflections: those about lines that pass through two vertices of $P$ and those about lines that bisect two edges of $P$.

3. Let $G = \text{Sym}(X)$ and let $f, g \in G$. Let us write $g$ in cyclic notation, and suppose that one of the cycles of $g$ is $(x_1, x_2, \ldots, x_r)$. Then $g(x_1) = x_2$, and so $fg(x_1) = f(x_2)$ and hence $fgf^{-1}(f(x_1)) = f(x_2)$. Similarly, we have $fgf^{-1}(f(x_i)) = f(x_{i+1})$ for $1 \leq i < r$ and $fgf^{-1}(f(x_r)) = f(x_1)$. Hence $fgf^{-1}$ has a cycle $(f(x_1), f(x_2), \ldots, f(x_r))$, and we have:

**Proposition 5.7** Given a permutation $g$ in cyclic notation, we obtain the conjugate $fgf^{-1}$ of $g$ by replacing each element $x \in X$ in the cycles of $g$ by $f(x)$.

For example, if $X = \{1, 2, 3, 4, 5, 6, 7\}$, $g = (1, 5)(2, 4, 7, 6)$ and $f = (1, 3, 5, 7, 2, 4, 6)$, then $fgf^{-1} = (3, 7)(4, 6, 2, 1)$.

In general, we say that a permutation has cycle-type $2^{p_2} 3^{p_3} \cdots$, if it has exactly $r_i$ cycles of length $i$, for $i \geq 2$. So, for example, $(1, 15)(2, 4, 6, 8, 7)(5, 9)(3, 11, 12, 13, 10)(14, 15, 16)$ has cycle-type $2^{3} 3^{5} 5^{2}$. By Proposition 5.7, conjugate permutations have the same cycle-type, and conversely, it is easy to see that if $g$ and $h$ have the same cycle-type, then there is an $f \in \text{Sym}(X)$ with $fgf^{-1} = h$. For example, if $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $g = (1, 5, 9)(2, 4, 6, 8)(7, 10)$ and $h = (1, 5)(2, 10, 9)(3, 6, 8, 7)$, then we can choose $f$ to map $1, 5, 9, 2, 4, 6, 8, 7, 10$ to $2, 10, 9, 3, 6, 8, 7, 1, 5, 4$, respectively, so $f = (1, 2, 3, 4, 6, 8, 7)(5, 10)$. $f$ is not unique; can you find some other possibilities? Hence we have:
Proposition 5.8 Two permutations of Sym(X) are conjugate in Sym(X) if and only if they have the same cycle-type.

For example, $S_3$ has three conjugacy classes, corresponding to cycle-types $1$, $2^1$, $3^1$, and $S_4$ has five conjugacy classes, corresponding to cycle-types $1$, $2^1$, $2^2$, $3^1$, $4^1$.

5.4 The Simplicity of $A_5$

In Subsection 3.3, we defined a group $G$ to be simple if its only normal subgroups are $\{1\}$ and $G$, and we saw that the only abelian simple groups are the cyclic groups of prime order. There are also infinitely many finite non-abelian simple groups. These were eventually completely classified into a number of infinite families, together with 26 examples known as sporadic groups, which do no belong to an infinite family. The work on this proof went on for decades, and was finally completed in about 1981.

One of the infinite families of finite non-abelian simple groups consists of the alternating groups, which do not belong to an infinite family. The work on this proof went on for decades, and was finally completed in about 1981.

For example, consider the subgroup $H$ of $A_5$, and we saw that the only abelian simple groups are the cyclic groups of prime order. There is one permutation of cycle-type $1$, 15 of type $2$, $3!$ of cycle-type $3$, and 24 of type $5^1$, making 60 elements in total.

The problem is that these are classes in $S_n$, and two permutations could conceivably be conjugate in $S_n$, but not in $A_n$, in which case the corresponding class would split up into more than one conjugacy class in $A_n$.

In fact, the 15 permutations of cycle-type $2^2$ forms a single class in $A_n$. To see this, let $g = (x_1, x_2)(x_3, x_4)(x_5)$ and $h = (y_1, y_2)(y_3, y_4)(y_5)$ be in this class. Then the permutations $f_i$ which maps $x_i$ to $y_i$ for $1 \leq i \leq 5$ and $f_2$ which maps $x_1 \rightarrow y_2$, $x_2 \rightarrow y_1$ and $x_i \rightarrow y_i$ for $3 \leq i \leq 5$ both satisfy $fgf^{-1} = h$, and since $f_2 = f_1(x_1, x_2)$, one of $f_1$ and $f_2$ is an even permutation and lies in $A_n$. Hence $g$ and $h$ are conjugate in $A_n$.

Similarly, the 20 permutations of cycle-type $3^1$ are all conjugate in $A_n$, because if $g = (x_1, x_2, x_3)(x_4)(x_5)$ and $h = (y_1, y_2, y_3)(y_4)(y_5)$, then $f_1$ mapping $x_i \rightarrow y_i$ for $1 \leq i \leq 5$ and $f_2$ mapping $x_1 \rightarrow y_1$ for $1 \leq i \leq 3$ and $x_4 \rightarrow y_5$, $x_5 \rightarrow y_4$ both satisfy $fgf^{-1} = h$.

However, for the cycle-type $5^1$, if we take $g = (1, 2, 3, 4, 5)$, then the five permutations $f_i$ with $f_i g f_i^{-1} = h$ are $(4, 5)g^k$ for $0 \leq k \leq 4$, and these are all odd permutations, so $g$ and $h$ are not conjugate in $A_n$. We can check that the 12 permutations $(1, 2, 3, 4, 5)$ for which the map $i \rightarrow x_i$ $(2 \leq i \leq 5)$ is an even permutation of $(2, 3, 4, 5)$ are all conjugate to $g$ in $A_n$, whereas the remaining 12, for which it is an odd permutation, are conjugate to $h$ in $A_n$. Hence this class in $S_n$ splits into two conjugacy classes, each of size 12, in $A_n$. Summing
Lemma 5.10 $A_5$ has 5 conjugacy classes, of sizes 1, 15, 20, 12, 12.

Theorem 5.11 $A_5$ is a simple group.

Proof: By Lemma 5.9, a normal subgroup $N$ of $A_5$ would be a union of conjugacy classes of $A_5$. But no combination of the numbers 1, 15, 20, 12, 12 that contains 1 adds up to a divisor of 60 other than 1 or 60, and so the result follows by Lagrange’s Theorem (2.10).

5.5 Sylow’s Theorem

By Lagrange’s Theorem (2.10), the order of a subgroup $H$ of a finite group $G$ always divides the order of $G$. An obvious converse question to ask is whether a group $G$ has subgroups of all orders that divide $|G|$. This is true for some groups. For example, we saw in Subsection 2.1 that it is true for finite cyclic groups. However, it is not true in general, and the smallest counterexample is described in the following result:

Proposition 5.12 $A_4$ has no subgroup of order 6.

Proof: $|A_4| = 24/2 = 12$, so a subgroup $H$ of order 6 would have index 2, and would be normal by Proposition 3.1, and hence by Lemma 5.9 a union of conjugacy classes of $A_4$. By Proposition 3.5, a group of order 6 is cyclic or dihedral, and in either case it contains at least one element of order 2. But the only elements of order two in $A_4$ are $(1, 2)(3, 4)$, $(1, 3)(2, 4)$ and $(1, 4)(2, 3)$, and these are all conjugate in $A_4$ (check!), so they must all lie in $H$. But then $H$ has a subgroup of order 4, namely $\{(), (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\}$, contradicting Lagrange’s Theorem.

The Norwegian mathematician Sylow proved a number of theorems about subgroups of groups of prime power order. We shall just prove the first one here, which asserts their existence. There are several proofs of this theorem. We have chosen one which demonstrates the use of conjugacy classes, quotient groups and induction. We start with a lemma.

Lemma 5.13 Let $G$ be an abelian group and let $p$ be a prime dividing $|G|$. Then $G$ has an element of order $p$.

Proof: We use induction on $|G|$. The result is vacuously true when $|G| = 1$, so suppose that $|G| > 1$. Let $g \in G$ with $1 \not= g$. By replacing $g$ by a suitable power of itself if necessary, we may assume that the order $q$ of $g$ is prime. If $q = p$ then we are done. Otherwise, because $G$ is abelian, the subgroup $N$ of $G$ generated by $g$ is normal, and the quotient group $G/N$ has order $|G|/q$, which is still divisible by $p$. Hence, by induction applied to $G/N$, $G/N$ has an element $hN$ of order $p$. Then $(hN)^p = 1_{G/N} = N$, and so $h^p \in N$. Either $h^p = 1$, and $|h| = p$, or $h^p = g^k$ for some $k$ with $1 \leq k < q$, and then $|h| = pq$ and $|h^q| = p$. In either case we have found an element of $G$ of order $p$.

Theorem 5.14 Let $G$ be a finite group and suppose that $p^n$ divides $|G|$, where $p$ is prime and $n > 0$. Then $G$ has a subgroup of order $p^n$.

Proof: We use induction on $|G|$. The result is clear when $|G| = 1$, so suppose that $|G| > 1$. Let the conjugacy classes of $G$ be $C_1, C_2, \ldots, C_r$, and let $g_i$ be an element in $C_i$. We assume that $C_1 = \{1\}$ where $g_1 = 1$. Then clearly

$$|C_1| + |C_2| + \ldots + |C_r| = |G| \quad (\dagger)$$

26
By Proposition 5.6, \(|C_i| = |G|/|C_G(g_i)|\), and so \(|C_i|\) divides the order of \(|G|\). Let us choose the numbering such that \(|C_i|\) is not divisible by \(p\) for \(1 \leq i \leq s\), and \(|C_i|\) is divisible by \(p\) for \(s + 1 \leq i \leq r\). Since \(|G|\) is divisible by \(p\), we must have \(s > 1\), because if \(s = 1\), then the left and right hand sides of Equation (†) would be congruent to 1 and 0 mod \(p\).

Suppose first that there is some \(C_i\) with \(2 \leq i \leq s\) and \(|C_i| > 1\). Then \(|C_G(g_i)| < |G|\) and so \(C_G(g_i)\) is a proper subgroup of \(G\), and since \(p\) does not divide \(|C_i|\), \(|C_G(g_i)|\) must be divisible by \(p^n\). Then we can apply induction to \(C_G(g_i)\) and deduce that \(C_G(g_i)\) has a subgroup \(H\) of order \(p^n\). But then \(H\) is also a subgroup of \(G\), and we are done.

Hence we may assume that \(|C_i| = 1\) for \(1 \leq i \leq s\). By Equation (†) again, we must have \(p\) divides \(|C_i|\). Now we saw above that \(|C_i| = 1\) if and only if \(g_i \in Z(G)\), so the centre \(Z(G)\) of \(G\) has order \(s\) and is equal to \(\{g_1, g_2, \ldots, g_s\}\). Since \(p\) divides \(|C_i|\), it follows from the lemma above that \(Z(G)\) has an element \(g\) of order \(p\). Let \(N\) be the subgroup \(\{1, g, g^2, \ldots, g^{p^n-1}\}\) of \(G\) generated by \(g\). Since \(g \in Z(G)\), we have \(g^i \in Z(G)\) for all \(i\), and so \(h(g^i) - 1 = g^i\) for all \(i\) and all \(h \in G\), and \(N \leq G\).

Now the quotient group \(G/N\) has order \(|G|/|N| = |G|/p\), and by induction applied to \(G/N\) we know that \(G/N\) has a subgroup of order \(p^n-1\). By Proposition 4.6, such a subgroup has the form \(H/N\) where \(H\) is a subgroup of \(G\) of order \(|N|/|H/N| = |N|p^n - 1 = p^n\), and we are done. \(\square\)

**Definition** Suppose that \(n\) is the largest power of the prime \(p\) that divides \(|G|\), so \(|G| = p^nq\) where \(q\) is not divisible by \(p\). A subgroup of \(G\) of order \(p^n\) is called a *Sylow \(p\)-subgroup* of \(G\).

We know from Theorem 5.14 that \(G\) has at least one Sylow \(p\)-subgroup. Here are the remaining theorems of Sylow, which we shall not prove here.

**Theorem 5.15** Let \(G\) be a finite group, \(p\) a prime, and \(|G| = p^nq\), where \(p\) does not divide \(|G|\). Then:

(i) The number \(r\) of Sylow \(p\)-subgroups of \(G\) satisfies \(r \equiv 1 \pmod{p}\);

(ii) Any two Sylow \(p\)-subgroups of \(G\) are conjugate in \(G\);

(iii) Any subgroup of \(G\) of order \(p^m\) for \(1 \leq m \leq n\) is contained in a Sylow \(p\)-subgroup of \(G\).

**Appendix**

**Proof of Proposition 2.4:** This proposition says that no permutation can be both even and odd. We can assume that the set \(X\) being permuted is \(\{1, 2, 3, \ldots, n\}\) for some \(n > 0\). Let \(X^{(2)}\) be the set of all unordered pairs \(\{i, j\}\) of elements of \(X\). For an integer \(x\), \(\text{sgn}(x)\) is defined to be 1, 0, or \(-1\), when \(x > 0\), \(x = 0\) and \(x < 0\), respectively. For a permutation \(\phi\) of \(X\), the sign \(\text{sgn}(\phi)\) of \(\phi\) is defined as:

\[
\text{sgn}(\phi) = \prod_{\{i,j\} \in X^{(2)}} \text{sgn} \left( \frac{\phi(j) - \phi(i)}{j - i} \right).
\]

Let \(\psi = (k, l)\) be a transposition. We claim that \(\text{sgn}(\phi \psi) = -\text{sgn}(\phi)\). To see this, consider the effect of replacing \(\phi\) by \(\phi \psi\) in the terms \(\text{sgn} \left( \frac{\phi(j) - \phi(i)}{j - i} \right)\) in the product above. If \(\{k, l\} \cap \{i, j\}\) is empty, then the term is unchanged. For those terms where \(\{k, l\} \cap \{i, j\}\) have one number in common, say \(i = k\), the product of the two terms \(\text{sgn} \left( \frac{\phi(j) - \phi(k)}{j - k} \right)\) and \(\text{sgn} \left( \frac{\phi(l) - \phi(k)}{l - k} \right)\) will be unchanged. Finally the term \(\text{sgn} \left( \frac{\phi(k) - \phi(l)}{k - l} \right)\) will change sign, so the whole product changes sign, which establishes the claim.

Since the identity permutation has sign 1, it now follows that if \(\phi\) is an even permutation then \(\text{sgn}(\phi) = 1\), and if \(\phi\) is an odd permutation then \(\text{sgn}(\phi) = -1\). The result follows. \(\square\)
Part II

Ring Theory

6  Definition, Examples and Elementary Properties

6.1 Definitions

Definition A ring is a set $R$ together with two binary operations $+ : R \times R \rightarrow R$ and $\cdot : R \times R \rightarrow R$ that satisfy the following properties:

(i) (Group under addition) $(R, +)$ is an abelian group.

(ii) (Associativity) For all $a, b, c \in R$, $(ab)c = a(bc)$.

(iii) (Distributivity) For all $a, b, c \in R$, $(a + b)c = ac + bc$ and $a(b + c) = ab + ac$.

(iv) (Identity) There exists an element $1 = 1_R \in R$ such that, for all $a \in R$, $1a = a1 = a$.

Note that we nearly always write $ab$ in place of $a \cdot b$. Some books omit Axiom (iv).

Lemma 6.1 Let $R$ be a ring. Then $R$ has a unique identity element.

Proof: Let $1$ and $1'$ be two identity elements of $R$. Then, $1 = 11' = 1'$.

As usual, the zero element of $(R, +)$ will be denoted by $0_R$ or usually just $0$.

Lemma 6.2 In any ring $R$, $0a = a0 = 0$ for all $a \in R$.

Proof: $0a = (0 + 0)a = 0a + 0a$, so $0a = 0$ by cancellation in $(R, +)$. Similarly $a0 = 0$.

What happens if $1 = 0$? The proof of the following lemma is left as an exercise.

Lemma 6.3 Let $R$ be a ring such that $0 = 1$. Then $R = \{0\}$.

The ring $\{0\}$ a ring is called the zero ring. All other rings will be called non-zero rings.

Definition A ring $R$ is commutative if it satisfies

(v) (Commutativity) For all $a, b \in R$, $ab = ba$.

Definition If $a$ and $b$ are non-zero elements of a ring $R$ with $ab = 0$, then $a$ and $b$ are called zero divisors. A non-zero commutative ring $R$ is called an integral domain (or just domain) if it has no zero divisors; that is, if $a, b \in R$, $ab = 0$ implies $a = 0$ or $b = 0$.

Definition An element $a$ of a ring $R$ is called a unit if it has a two-sided inverse under multiplication; that is, if there exists $b \in R$ with $ab = ba = 1$.

Exercise Show that the units in $R$ form a group under multiplication.

A non-zero ring $R$ is called a division ring if $R \setminus \{0\}$ is a group under multiplication; that is, if all of its non-zero elements are units. So a field is a commutative division ring.

Exercise Show that a field is a domain.

Proposition 6.4 (Cancellation laws in domains) If $R$ is a domain, $a, b, c \in R$ with $a \neq 0$, and $ab = ac$ or $ba = ca$, then $b = c$.

Proof: $ab = ac \Rightarrow a(b - c) = 0$ and since $R$ is a domain with $a \neq 0$, this implies $b - c = 0$, so $b = c$. The proof for $ba = ca$ is similar.

28
Proposition 6.5  A finite integral domain is a field.

Proof: Let \( R = \{r_0, r_1, \ldots, r_n\} \) be a finite domain with \( 0 = r_0 \). By Proposition 6.4, for fixed \( i > 0 \), the \( n \) products \( r_ir_j \) \((1 \leq j \leq n)\) are all distinct and non-zero. Since there are \( n \) possible values for these \( n \) products, they all occur exactly once. In particular, we have \( r_ir_j = 1 \) for some \( j \), so \( R \) is a field. \( \Box \)

6.1.1 Examples – Numbers

\( \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \) and \( \mathbb{C} \) are commutative rings under the usual addition and multiplication. They are all domains, and \( \mathbb{Q}, \mathbb{R}, \) and \( \mathbb{C} \) are fields.

In Subsection 1.7 we saw that, for a fixed \( n > 0 \), \( \mathbb{Z}_n = \{0, 1, \ldots, n-1\} \) forms a group under addition modulo \( n \). We can also define multiplication modulo \( n \), and this makes \( \mathbb{Z}_n \) into a commutative ring. For instance, in \( \mathbb{Z}_6 \), \( 4 + 5 = 3 \) (residue of 9 modulo 6), \( 4 \cdot 5 = 2 \) (residue of 20 modulo 6), \( 4 \cdot 3 = 0 \) (residue of 12 modulo 6).

Proposition 6.6  \( \mathbb{Z}_n \) is a domain if and only if \( n \) is prime, in which case it is a field.

Proof: If \( n = 1 \) then \( \mathbb{Z}_n \) is the zero ring so is by definition not a domain. If \( n = ab \) with \( 1 < a, b < n \), then \( ab = 0 \) with \( a \neq 0 \neq b \) in \( \mathbb{Z}_n \), so \( \mathbb{Z}_n \) is not a domain. If \( n \) is prime and \( 0 < a, b < n \), then \( n \) does not divide \( ab \) (we shall prove this formally later), so \( ab \neq 0 \) in \( \mathbb{Z}_n \), which is therefore a domain, and also a field by Proposition 6.5. \( \Box \)

Exercise  What are the units in \( \mathbb{Z}_n \).

6.1.2 Examples – Matrices

If \( R \) is a ring then the set \( M_n(R) \) of \( n \times n \)-matrices with coefficients in \( R \) is another ring. The multiplication and addition is the same as you have learnt in linear algebra: \((a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})\) and \((a_{ij}) \cdot (b_{ij}) = (\sum_k a_{ik}b_{kj})\).

For the zero ring \( R \), the ring \( M_n(R) \) is also zero. For a non-zero ring \( M_n(R) \) is commutative if and only if \( R \) is commutative and \( n = 1 \). The units in \( M_n(R) \) are of course the invertible matrices.

6.1.3 Examples – Polynomials

Let \( R \) be a ring, and let \( x_1, \ldots, x_n \) be independent commuting variables. The polynomials in \( x_i \)-s with coefficients in \( R \) form the polynomial ring \( R[x_1, \ldots, x_n] \) under the usual addition and multiplication of polynomials.

A monomial is an expression \( x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \), where \( \alpha_1, \ldots, \alpha_n \) are non-negative integers. A polynomial is a linear combination of monomials with coefficients in \( R \). For example, \( 3x_1^2x_2/2 + 11x_1 - 2x_1x_2^3/3 \in \mathbb{Q}[x_1, x_2] \).

Observe that \( R[x_1, \ldots, x_n] \) is commutative if and only if \( R \) is commutative.

The most important example is the ring \( R[x] \) of polynomials in a single variable \( x \). So each \( f \in R[x] \) has the form \( f = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \), with \( a_i \in R \). If \( a_n \neq 0 \), then we define \( n \) to be the degree \( \deg(f) \) of \( f \), and \( a_n \) is called the leading coefficient of \( f \). A polynomial is called monic, if its leading coefficient is 1. Note that non-zero elements of \( R \) have degree 0, and we leave the degree of the zero polynomial undefined, although some authors define this to be \(-1\) or \(-\infty\).
6.2 Subrings

**Definition** A subset $S$ of a ring $R$ is called a subring of $R$ if it forms a ring under the same operation as that of $R$ with the same identity element.

**Proposition 6.7** Let $R$ be a ring and let $S$ be a subgroup of $(R, +)$. Then $S$ is a subring of $R$ if and only if

(i) $a_1, a_2 \in S \implies a_1a_2 \in S$; and

(ii) $1_R \in S$.

**Proof:** As a subgroup of $(R, +)$, $(S, +)$ is an abelian group, closed under the multiplication and containing the identity. Thus, $S$ has two operations and identity for multiplication. All ring axioms of $S$ easily follow from the corresponding axioms of $R$. 

**Lemma 6.8** The intersection of any set of subrings of $R$ is itself a subring.

6.3 Isomorphisms and direct products

**Definition** An isomorphism $\phi : R \rightarrow S$ between two rings $R$ and $S$ is a bijection from $R$ to $S$ such that $\phi(r_1r_2) = \phi(r_1)\phi(r_2)$ and $\phi(r_1 + r_2) = \phi(r_1) + \phi(r_2)$ for all $r_1, r_2 \in R$. Two rings $R$ and $S$ are called isomorphic if there is an isomorphism between them. In this case we write $R \cong S$.

If $\phi : R \rightarrow S$ is an isomorphism, then we know from Lemma 4.1 for groups that $\phi(0_R) = 0_S$. Since $\phi(1_R)$ is an identity element of $S$, it follows from Lemma 6.1 that $\phi(1_R) = 1_S$.

**Definition** Let $R$ and $S$ be two rings. We define the direct product $R \times S$ of $R$ and $S$ to be the set $\{(r, s) \mid r \in R, s \in S\}$ of ordered pairs of elements from $R$ and $S$, with the obvious component-wise addition and multiplication $(r_1, s_1) + (r_2, s_2) = (r_1 + r_2, s_1 + s_2)$, $(r_1, s_1)(r_2, s_2) = (r_1r_2, s_1s_2)$ for $r_1, r_2 \in R$ and $s_1, s_2 \in S$.

It is straightforward to check that $R \times S$ is a ring under these operations. Note that the identity element is $(1_R, 1_S)$.

But $R$ and $S$ are not isomorphic to subrings of $R \times S$ in general. Indeed, $R$ can be thought of as the elements of the form $(r, 0_S)$, and these elements define a ring isomorphic to $R$. But its identity element is $(1_R, 0_S)$, which is not the same as the identity element $(1_R, 1_S)$ of $R \times S$.

Note also that the direct product of two non-zero rings is never a domain, since $(1_R, 0_S)(0_R, 1_S) = (0_R, 0_S)$, the zero element of $R \times S$.

**Exercise** $(r, s)$ is a unit in $R \times S$ if and only if $r$ is a unit in $R$ and $s$ is a unit in $S$.

In the proof of the next proposition, we shall use the following fact, which was proved last year in Foundations. We shall prove it again later on in the course. Integers $m$ and $n$ are coprime (have no common prime divisors) if and only if there exist integers $a$ and $b$ such that $am + bn = 1$.

**Proposition 6.9 (The Chinese Remainder Theorem)** The rings $\mathbb{Z}_m \times \mathbb{Z}_n$ and $\mathbb{Z}_{mn}$ are isomorphic if and only if $m$ and $n$ are coprime.

**Proof:** The “only if” bit follows from Proposition 1.10. If $m$ and $n$ are not relatively prime then their least common multiple $l = \text{lcm}(m, n) < mn$. The abelian additive groups are not isomorphic because the order of 1 is $mn$ in $(\mathbb{Z}_{mn}, +)$ but for any element $(a, b) \in \mathbb{Z}_m \times \mathbb{Z}_n$, we have $l(a, b) = (la, lb) = 0$, so no element of $\mathbb{Z}_m \times \mathbb{Z}_n$ has order $mn$. 

30
If $m$ and $n$ are coprime then there exist integers $a, b$ such that $am + bn = 1$. Let $(x)_n$ denote the residue of $x$ mod $n$. We define $\phi : \mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n$ by $\phi(x) = ((x)_m, (x)_n)$. It is clear that $\phi(x + y) = \phi(x) + \phi(y)$ and $\phi(xy) = \phi(x)\phi(y)$. It remains to show that $\phi$ is a bijection. We do it by writing the inverse map explicitly: $\phi^{-1}(y, z) = (amz + bny)_{mn}$. Since both sets have the same cardinality $mn$, it suffices to observe that

$$\phi(\phi^{-1}(y, z)) = \phi((amz + bny)_{mn}) = (((amz + bny)_{mn})_m, ((amz + bny)_{mn})_n) =

((amz + bny)_m, (amz + bny)_n) = ((bny)_m, (amz)_n) = (y, z)$$

with the last equality ensured by $(bn)_m = 1 = (am)_n$ since $am + bn = 1$. □

Using induction on $k$, one derives the following corollary.

**Corollary 6.10** If $n = p_1^{a_1} \cdots p_k^{a_k}$ is a decomposition of $n$ into a product of distinct primes then $\mathbb{Z}_{n} \cong \mathbb{Z}_{p_1^{a_1}} \times \cdots \times \mathbb{Z}_{p_k^{a_k}}$ as rings.

### 7 Homomorphisms, Ideals, Quotient Rings and the Isomorphism Theorems

The theory of ring homomorphisms is very similar to that of group homomorphisms.

#### 7.1 Homomorphisms

**Definition** Let $R$ and $S$ be rings. A ring homomorphism $\phi$ from $R$ to $S$ is a function $\phi : R \to S$ such that $\phi(1_R) = 1_S$, and for all $r_1, r_2 \in R$ we have and $\phi(r_1 + r_2) = \phi(r_1) + \phi(r_2)$ $\phi(r_1r_2) = \phi(r_1)\phi(r_2)$.

From Lemma 4.1 for groups we have $\phi(0_R) = 0_S$. Unlike in the case of a ring isomorphism, we cannot deduce that $\phi(1_R) = 1_S$ from the other conditions, so we choose to make it part of the definition (although in some books it is not!).

As for group homomorphisms, we define a ring monomorphism to be an injective homomorphism and a ring epimorphism to be a surjective homomorphism. So it is an isomorphism if and only if it is both a monomorphism and an epimorphism.

The image $\text{im}(\phi)$ of a ring homomorphism is just its image as a function, and it is straightforward to check using Proposition 6.7 that it is a subring of $S$.

The kernel $\ker(\phi)$ of a ring homomorphism is defined to be its kernel as a homomorphism of additive groups. That is,

$$\ker(\phi) = \{ r \in R \mid \phi(r) = 0_S \}.$$

**Examples**

1. For a fixed integer $n > 0$, there is a ring homomorphism $\phi : \mathbb{Z} \to \mathbb{Z}_n$ in which, for $k \in \mathbb{Z}$, $\phi(k) = k \pmod n$. That is, $\phi(k)$ is the residue of $k$ modulo $n$. Then $\ker(\phi) = n\mathbb{Z}$.

2. If $\phi : R \to S$ is a ring homomorphism, then there is an induced homomorphism $\psi : R[x] \to S[x]$, defined by

$$\psi(a_nx^n + \cdots + a_1x + a_0) = \phi(a_n)x^n + \cdots + \phi(a_1)x + \phi(a_0).$$

Then $\ker(\psi)$ consists of those polynomial in which all coefficients lie in $\ker(\phi)$.

Similarly, there is an induced homomorphism of matrix rings $\psi : M_n(R) \to M_n(S)$.
3. Here is another example related to polynomials. Let $R$ be a subring of $S$ and let $\alpha$ be a fixed element of $S$. Then the map $\phi : R[x] \to S$ defined by $\phi(f) = f(\alpha)$, which evaluates the polynomial $f$ at the point $\alpha$, is a ring homomorphism in which $\ker(\phi)$ consists of those polynomials with $f(\alpha) = 0$.

4. Let $R$ be a commutative ring of characteristic $p$, that is, $px = 0$ for any $x \in R$, where $p$ is a prime number. The ring $R$ admits a homomorphism, $\phi : R \to R$ defined by $\phi(x) = x^p$. Clearly $\phi(xy) = \phi(x)\phi(y)$. The identity $(x + y)^p = x^p + y^p$ is sometimes called the Freshman’s dream binomial formula. It holds because the commutativity of $x$ and $y$ implies that

$$(x + y)^p = x^p + y^p + \sum_{k=1}^{p-1} \frac{p!}{k!(p-k)!} x^k y^{p-k}$$

and all binomial coefficients in the sum are divisible by $p$.

7.2 Ideals

Definition A subset $I$ of a ring $R$ is called an ideal in $R$ if

(i) $I$ is a subgroup of $(R, +)$;
(ii) For all $x \in R$ and $y \in I$, we have $xy \in I$ and $yx \in I$.

Notice that ideals are different from subrings. They are both subgroups of $(R, +)$, but subrings satisfy Condition (ii) for ideals only when $x$ and $y$ are both in $I$. On the other hand, subrings must by definition contain $1_R$, whereas it is an easy exercise to show that an ideal contains $1_R$ only when $I = R$.

When $R$ is a commutative ring, it is routine to check that the subset $(a) = \{ ra \mid r \in R \}$, consisting of all multiples of $a$ in $R$, is an ideal of $R$, and it is known as the principal ideal generated by $a$. This is also often written as $aR$ or $Ra$.

For an arbitrary ring, the principal ideal $(a)$ is equal to the set of finite sums $\{ \sum_{i=1}^{k} r_i a s_i \mid r_i, s_i \in R \}$, but this will not be needed for this course.

Exercise If $R$ is commutative, then $(a) = R$ if and only if $a$ is a unit of $R$.

In $\mathbb{Z}$, the principal ideals are the sets $(n) = n\mathbb{Z}$ of multiples of $n$, for some fixed $n \geq 0$. We shall see later that these are the only ideals of $\mathbb{Z}$.

In the theory of homomorphisms, ideals of rings correspond to normal subgroups of groups.

Proposition 7.1 Let $\phi : R \to S$ be a ring homomorphism. Then $\ker(\phi)$ is an ideal in $R$.

Proof: By Theorem 4.4, $K = \ker(\phi)$ is an additive subgroup of $R$. If $r \in K$, $x \in R$ then $\phi(xr) = \phi(x)\phi(r) = 0_S\phi(r) = 0_S$. Hence $xr \in K$. Similarly, $rx \in K$ and $K$ is an ideal. \qed

Exercise. Let $\phi : R \to S$ be a ring homomorphism. Show that, if $J$ is an ideal of $S$, then $\phi^{-1}(J)$ is an ideal of $R$. Give an example where $I$ is an ideal of $R$ but $\phi(I)$ is not an ideal of $S$.

7.3 Quotient Rings

Since an ideal $I$ of a ring $R$ is a subgroup of $(R, +)$, we can consider its cosets $I + a$ for $a \in R$. We know from Theorem 3.9 that they form a group under addition, with the operation $(I + a_1) + (I + a_2) = I + (a_1 + a_2)$. The following proposition defines the quotient ring $R/I$.

Proposition 7.2 The cosets of an ideal form a ring under addition in the quotient group and the multiplication $(I + a)(I + b) = I + ab$. 

32
Proof: We know already from group theory that \( R/I \) is a group under addition, so Axiom (i) of the definition of a ring is satisfied. We have to verify that the multiplication in \( R/I \) is well-defined. To do this, suppose that \( I + a = I + x \) and \( I + b = I + y \). So \( a - x, b - y \in I \), and \( ab = ab - ay + ay - xy + xy = a(b - y) + (a - x)y + xy \). But by the definition of an ideal, we have \( a(b - y) \in I \) and \((a - x)y \in I\), so \( ab - xy \in I \) and \( I + ab = I + xy \), and multiplication is well-defined. Axioms (ii), (iii) and (iv) follow immediately from the same properties in \( R \), noting that \( 1_{R/I} = I + 1 \).

Examples 1. The quotient ring \( \mathbb{Z}/(n) \) is isomorphic to the ring \( \mathbb{Z}_n \) of the residues modulo \( n \). The isomorphism \( \mathbb{Z}_n \to \mathbb{Z}/(n) \) is just \( m \to m + (n) \). This is an instance of the first isomorphism for rings, which we shall prove shortly.

2. Let \( F \) be a field and \( f \in F[x] \) with \( \deg(f) > 0 \). Let \( I = (f) \), the principal ideal of \( F[x] \) generated by \( f \). Since \( f \) is not changed by multiplying \( f \) by a constant, we may assume that \( f = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \) has leading coefficient 1.

Let us consider the quotient ring \( Q = F[x]/(f) \). The elements of \( Q \) are the cosets \((f) + g \) with \( g \in F[x] \). Since \( f = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \), we can replace \( x^n \) by \(-a_{n-1}x^{n-1} - \cdots - a_1x - a_0 \) in \( g \) without changing the coset \((f) + g \). By doing this repeatedly, we can assume that \( \deg(g) < n \). For example, if \( f = x^2 + 2x - 3 \), then
\[
(f) + x^3 = (f) - x(2x - 3) = (f) + 2(2x - 3) + 3x = (f) + 7x - 6.
\]

On the other hand, if \( g, g' \in F[x] \) with \( g \neq g' \) and \( \deg(g), \deg(g') < \deg(f) \), then \( g - g' \notin (f) \), so \((f) + g \) and \((f) + g' \) are in distinct cosets of \((f) \). Hence there is a bijection between \( Q \) and the polynomials of degree less than \( \deg(f) \), where multiplication is done modulo \( f \). In particular, if \( \deg(f) = 1 \) then \( Q \cong F \).

Proposition 7.3 Let \( I \) be an ideal of a ring \( R \). Then the map \( \phi : R \to R/I \) defined by \( \phi(a) = I + a \) is a surjective ring homomorphism with kernel \( I \).

Proof: It is straightforward to check that \( \phi \) is a surjective ring homomorphism. Since \( 0_{R/I} = I \), we have \( \phi(a) = 0_{R/I} \iff I + a = I \iff a \in I \), so \( \ker(\phi) = I \).

Theorem 7.4 (First Isomorphism Theorem for Rings) Let \( \phi : R \to S \) be a ring homomorphism with kernel \( I \). Then \( R/I \cong \text{im}(\phi) \). More precisely, there is an isomorphism \( \overline{\phi} : R/I \to \text{im}(\phi) \) defined by \( \overline{\phi}(I + a) = \phi(a) \) for all \( a \in R \).

Proof: By Theorem 4.7, \( \overline{\phi} \) is a well-defined isomorphism of abelian groups under addition. It is straightforward to check that it is a ring homomorphism.

7.4 Maximal ideals

Definition An ideal \( I \) of a ring \( R \) is called maximal, if \( I \neq R \), but if \( J \) is any ideal of \( R \) with \( I \subseteq J \subseteq R \), then \( I = J \) or \( J = R \).

Theorem 7.5 An ideal \( I \) in a commutative ring \( R \) is maximal if and only if \( R/I \) is a field.

Proof: First suppose that \( I \) is maximal. Since \( I \neq R \), \( 1 \notin I \), so \( I + 0 \neq I + 1 \). We have to check that, for each \( x \in R \setminus I \), \((I + x)\) has a multiplicative inverse in \( R/I \). Since \( I \) is maximal and \( x \notin I \), the ideal \((I + x)\) is equal to \( R \), so \( 1 \in I + (x) \), and hence there exists \( y \in R \) with \( 1 \in I + xy \). Then \( I + 1 = I + xy = (I + x)(I + y) \).

Conversely, suppose that \( R/I \) is a field. Then \( I + 1 \neq I + 0 \), so \( I \neq R \). Let \( J \) be an ideal of \( R \) with \( I \subseteq J \subseteq R \). If \( I \neq J \) and \( x \in J \setminus I \), then \( I + x \neq I + 0 \), so \( I + x \) has a multiplicative inverse in \( R/I \), and there exists \( y \in R \) with \((I + x)(I + y) = I + 1 \). Hence \( I + 1 = I + xy \), so \( 1 \in I + (x) \subseteq J \) and thus \( J = R \).
8 Euclidean Domains, Principal Ideal Domains, and Unique Factorisation Domains

The ring $R$ will be an integral domain (and hence commutative) throughout this section.

8.1 Euclidean and Principal Ideal Domains

As you saw in Foundations, in the ring $\mathbb{Z}$ of integers, we have a Euclidean/division algorithm:

**Proposition 8.1** For any $a, b \in \mathbb{Z}$ with $0 \neq b$, there exist $q, r \in \mathbb{Z}$ with $a = qb + r$, where $0 \leq r < |b|$.

This can be proved by considering the least element $r$ of the set \{a – qb | q ∈ Z, a – qb ≥ 0\} and showing that $0 ≤ r < |b|$.

There is a similar result for polynomial rings $F[x]$ in one variable when $F$ is a field.

**Proposition 8.2** For any $f, g \in F[x]$ with $0 \neq g$, there exist $q, r \in F[x]$ with $f = qg + r$, where either $r = 0$ or $\deg(r) < \deg(g)$.

**Proof:** If $f = 0$ we take $q = r = 0$, and otherwise we use induction on $d = \deg(f)$. If $d = 0$, then $f \in F$ and then take $q = 0, r = f$ if $\deg(g) > 0$, and $q = fg^{-1}, r = 0$ if $\deg(g) = 0$.

So suppose that $d > 0$ and $f = a_dx^d + a_{d-1}x^{d-1} + \cdots + a_0$. If $\deg(g) > d$, then take $q = 0, r = f$.

Otherwise $g = b_ex^e + \cdots + b_0$ with $e = \deg(g) \leq d$. Now $\deg(f - a_db_e^{-1}x^{d-e}g) < d$ and the result follows by applying induction to $f - a_db_e^{-1}x^{d-e}g$. 

These two examples motivate the following definition.

**Definition** A Euclidean domain (abbreviated ED) is a domain $R$ that admits a norm function $\nu : R \setminus \{0\} \to \mathbb{N} \cup \{0\}$ such that

(i) $\nu(ab) ≥ \nu(b)$ for all $a, b \in R \setminus \{0\}$;

(ii) $\forall a, b \in R$ with $b \neq 0, \exists q, r \in R$ such that $a = qb + r$ and either $r = 0$ or $\nu(r) < \nu(b)$.

**Examples 1.** $\mathbb{Z}$ and $F[x]$ are Euclidean domains with $\nu(a) = |a|$ and $\nu(a) = \deg(a)$, respectively.

2. Here is another example. Let $i = \sqrt{-1}$ and let $\mathbb{Z}[i] = \{x + iy | x, y \in \mathbb{Z}\}$ be the ring of Gaussian integers. It is straightforward to check using Proposition 6.7 that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$, so it is certainly a domain. For $z \in \mathbb{Z}[i]$ define $\nu(z) = |z|^2$; that is, $\nu(x + iy) = x^2 + y^2$.

Since $\nu(z_1z_2) = \nu(z_1)\nu(z_2)$, Condition (i) holds. Of course, we can define $\nu(z)$ in the same way for all $z \in \mathbb{C}$, and we still have $\nu(z_1z_2) = \nu(z_1)\nu(z_2)$.

Let $a, b \in \mathbb{Z}[i]$ with $b \neq 0$. Then we have $a/b = x + iy$ with $x, y \in \mathbb{Q}$. Choose $x_0, y_0 \in \mathbb{Z}$ with $|x - x_0| ≤ 1/2$ and $|y - y_0| ≤ 1/2$. Then

$$a = b(x + iy) = b(x_0 + iy_0) + b((x - x_0) + i(y - y_0)) = qb + r,$$

where $q = (x_0 + iy_0) \in \mathbb{Z}[i]$ and $r = b((x - x_0) + i(y - y_0))$. Since $r = a - qb$, we have $r \in \mathbb{Z}[i]$, and $\nu(r) = \nu(b)\nu((x-x_0) + i(y-y_0)) ≤ \nu(b)(1/4 + 1/4) < \nu(b)$, so Condition (ii) holds, and $\mathbb{Z}[i]$ is an ED.

We can prove various nice properties of EDs, but most of these properties hold in a more general class of rings, which we shall now introduce. Recall that, in a commutative ring $R$, the principal ideals are those of the form $(a) = aR$ for some fixed $a \in R$.

**Definition** A domain $R$ is called a principal ideal domain (abbreviated PID) if every ideal of $R$ is principal.
Theorem 8.3 Any Euclidean domain is a principal ideal domain. (ED $\implies$ PID)

Proof: Let $I$ be an ideal in a Euclidean domain $R$. If $I = \{0\}$ then $I = (0)$ is principal. Otherwise, choose $b \in I \setminus \{0\}$ such that $\nu(b)$ is as small as possible. By definition of ideal, $(b) \subseteq I$. Let us now prove the opposite inclusion. For an arbitrary $a \in I$ we can write $a = bq + r$ with either $r = 0$ or $\nu(r) < \nu(b)$. If $r \neq 0$ then $r = a - bq \in I$ with $\nu(r) < \nu(b)$, contradicting the choice of $b$. So $a = bq \in (b)$ and $I = (b)$ is principal. $\Box$

Examples It turns out that the domain $\mathbb{Z}[\alpha]$ with $\alpha = (1 + \sqrt{-19})/2$ is a PID but not a ED, but we shall not prove that here. We shall however prove later that $\mathbb{Z}[\alpha]$ with $\alpha = \sqrt{-5}$ is not a PID.

Various familiar properties of divisibility that hold in $\mathbb{Z}$ hold in the more general context of PIDs.

Definition Let $x, y \in R$ we say that $x$ divides $y$ and write $x|y$ if $y = xr$ for some $r \in R$.

The following lemma is straightforward.

Lemma 8.4 The following statements are equivalent for all $x, y \in R$.

(i) $x|y$;
(ii) $y \in (x)$;
(iii) $(x) \supseteq (y)$.

Definition Let $x, y \in R$. We say that $x$ and $y$ are associate (and write $x \sim y$) if both $x|y$ and $y|x$.

Lemma 8.5 The following statements are equivalent in a domain $R$:

(i) $x \sim y$;
(ii) $(y) = (x)$;
(iii) There exists a unit $q \in R$ with $x = qy$.

Proof: (i) $\iff$ (ii) follows from the previous lemma.

(i) $\implies$ (iii): For $x = 0$, we have $x \sim y \iff y = 0$, so assume that $x \neq 0 \neq y$. There exist $q, r \in R$ such that $x = qy$ and $y = rx$. Then $x = qy = q(rx)$ and $(1 - qr)x = 0$. Because $R$ is a domain, $1 - qr = 0$ and so $q$ is a unit. (iii) $\implies$ (i) is easy. $\Box$

Examples In $\mathbb{Z}$, the only units are $\pm 1$, so $x \sim y \iff |x| = |y|$.

In $F[x]$, with $F$ a field, the units are the non-zero constants, so $x \sim y \iff x = ay$ for some $a \in F \setminus \{0\}$. Note that every polynomial is associate to a unique monic polynomial (leading coefficient 1).

In $\mathbb{Z}[i]$, the multiplicative inverse of $a + ib$ in $\mathbb{C}$ is $(a - ib)/(a^2 + b^2)$, which lies in $\mathbb{Z}[i]$ if and only if $a^2 + b^2 = 1$. So there are four units, $\pm 1$ and $\pm i$, and $x \sim y \iff x = \pm y$ or $x = \pm iy$.

Definition Let $x, y \in R$. A greatest common divisor $\gcd(x, y)$ (also called highest common factor) is an element $d \in R$ such that

(i) $d|\gcd(x, y);
(ii) if \ z \in R \ with \ z|x \ and \ z|y, \ then \ z|d$.

A least common multiple $\text{lcm}(x, y)$ is an element $l \in R$ such that

(i) $x|l \ and \ y|l$;
(ii) if $z \in R \ with \ x|z \ and \ y|z, \ then \ l|z$.
This definition can be generalised in the obvious way to the gcd or lcm of any set of elements of $R$.

Any two greatest common divisors $\gcd(x, y)$ of $x$ and $y$ must divide each other, and so are associates. Similarly for lcms. So we have uniqueness up to associate elements. The same holds for the gcd or lcm of a set of elements of $R$. We often write things like $\gcd(4, 6) = 2$ to mean that 2 is a gcd of 4 and 6. This is a mild abuse of notation, since it is also true that $\gcd(4, 6) = -2$, but of course $2 \neq -2$. Note that $\gcd(0, x) = x$ and $\lcm(0, x) = 0$ for any $x \in R$.

Existence of greatest common divisors is a bit trickier. They do not necessarily exist in an arbitrary domain, but they do in a PID.

**Proposition 8.6** If $R$ is PID then $\lcm(x, y)$ and $\gcd(x, y)$ exist for any pair of elements $x, y \in R$. Furthermore, there exist $r, s \in R$ with $\gcd(x, y) = rx + sy$.

(We have already used this last property in $\mathbb{Z}$ in the proof of Proposition 6.9.)

**Proof:** Let $I = \{rx + sy \mid r, s \in R\}$. It is routine to check that $I$ is an ideal of $R$ (it is the sum $(x) + (y)$ of the ideals $(x)$ and $(y)$), so it is principal, and hence $I = (d)$ for some $d \in R$. Similarly, the intersection $(x) \cap (y)$ is an ideal, so is equal to $(l)$ for some $l \in R$. We claim that $d$ is a greatest common divisor and $l$ is a least common multiple of $x$ and $y$. Indeed, $(x) \subseteq (d) \supseteq (y)$ and whenever $x \subseteq (z) \supseteq (y)$ it follows that $(z) \supseteq (x) + (y) = (d)$. Similarly, $(x) \supseteq (l) \subseteq (y)$ and whenever $x \supseteq (z) \subseteq (y)$ it follows that $(z) \subseteq (x) \cap (y) = (l)$. \qed

In an ED, we can compute $r, s \in R$ with $\gcd(x, y) = rx + sy$ by repeated application of the Euclidean algorithm. If $\nu(x) > \nu(y)$ then we divide $y$ into $x$ and get a remainder $x_1$, then divide $x_1$ into $x$ to get a remainder $x_2$ and carry on until the remainder is 0. At this stage, the last $x_i$ found is equal to $\gcd(x, y)$ (it is left as an exercise to prove this), and $r$ and $s$ can be computed by substituting back into the equations found.

For example, in $\mathbb{Z}$ let $x = 26, y = 6$. Then

$$
\begin{align*}
26 & = 2 \times 10 + 6 \\
10 & = 1 \times 6 + 4 \\
6 & = 1 \times 4 + 2 \\
4 & = 2 \times 2 + 0
\end{align*}
$$

so $\gcd(26, 6) = 2$, and substituting for 4 and then 6 gives $2 = 6 - 4 = 6 - (10 - 6) = 2 \times 6 - 10 = 2 \times (26 - 2 \times 10) - 10 = 2 \times 26 - 5 \times 10$.

### 8.2 Prime and Irreducible Elements

There are two different ways to say what a prime number is. We are going to see that these lead to two different notions in an arbitrary domain.

**Definition** Let $r \in R \setminus \{0\}$. We say that $r$ is **irreducible** if $r$ is not a unit, but $r = ab$ with $a, b \in R$ implies that either $a$ or $b$ is a unit.

We say that $r \in R \setminus \{0\}$ is **prime** if $r$ is not a unit, and $r \mid xy$ implies that $r \mid x$ or $r \mid y$.

**Proposition 8.7** If $R$ is a domain, then any prime element of $R$ is irreducible.

**Proof:** Let $r$ be prime and $r = ab$. Then, since $r \mid r = ab$, we have $r \mid a$ or $r \mid b$. Without loss of generality, $r \mid a$. Since $a \mid r$, we have $r \sim a$, so $r = aq$ with $q$ a unit. Then $q = b$ by Proposition 6.4, so $b$ is a unit, and $r$ is irreducible. \qed

**Proposition 8.8** If $R$ is a PID, then any irreducible element of $R$ is prime.
PROOF: This one is a little tricky! Let \( r \) be irreducible and suppose that \( r|ab \) with \( a, b \in R \). An element \( c = \gcd(r, a) \) exists by Proposition 8.6. Then \( r = ct \) for some \( t \in R \). Since \( r \) is irreducible, either \( c \) or \( t \) is a unit. We consider both cases.

If \( t \) is a unit, then \( r \sim c \) and \( c|a \), so \( r|a \).

So suppose that \( c \) is a unit. By Proposition 8.6, we have \( c = xa + yr \) for some \( x, y \in R \), and multiplying by \( b \) gives \( cb = xab + yrb \). Now we know that \( r|ab \) and clearly \( r|yb \), so \( r|cb \); that is, \( ru = cb \) for some \( u \in R \). But \( c \) a unit means that we can multiply by \( c^{-1} \) and conclude that \( r|b \).

\[\square\]

Example Let \( R = \mathbb{Z}[\sqrt{-5}] = \{ a + b\sqrt{-5} \mid a, b \in \mathbb{Z} \} \). Then 
\[6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})\]
in \( R \). We claim that \( 2 \) is irreducible but not prime.

\( 2 \) does not divide \( 1 \pm \sqrt{-5} \) because \( 2x = 1 \pm \sqrt{-5} \) implies that \( x = 1/2 \pm \sqrt{-5}/2 \), which is not an element of \( R \). Hence, \( 2 \) is not prime.

Let us show that \( 2 \) is irreducible. If \( 2 = ab \) with \( a = x + y\sqrt{-5}, \ b = s + t\sqrt{-5} \in R \) then 
\[4 = |a|^2|b|^2 = (x^2 + 5y^2)(s^2 + 5t^2)\]. Clearly, \(|a|^2, |b|^2 \in \mathbb{N}\). If \(|a|^2 = 1 \) then \( a^{-1} = x - y\sqrt{-5} \in R \) and \( a \) is a unit in \( R \). Similarly, if \(|a|^2 = 4 \) then \(|b|^2 = 1 \) and \( b \) is a unit in \( R \). Finally, it is clear that there are no integers \( x, y \) with \( x^2 + 5y^2 = 2 \), so we cannot have \(|a|^2 = 2 \). Hence \( 2 \) is irreducible as claimed. It can shown similarly that \( 3, (1 + \sqrt{-5}) \) and \( (1 - \sqrt{-5}) \) are all irreducible in \( R \).

So by Proposition 8.8, we have

\[\text{Corollary 8.9 } \mathbb{Z}[\sqrt{-5}] \text{ is not a PID and hence also not a ED.}\]

8.3 Number Fields

Another interesting and easy property of PIDs is the following.

Proposition 8.10 For \( \alpha \neq 0 \), the ideal \((\alpha)\) in a PID \( R \) is maximal if and only if \( \alpha \) is irreducible.

PROOF: If \((\alpha)\) is maximal then \((\alpha) \neq R\), so \( \alpha \) cannot be a unit. If \( \alpha = bc \), then \((\alpha) \subseteq (b) \subseteq R\), and so either \((\alpha) = (b)\), in which case \( c \) is a unit, or \((b) = R\), in which case \( \alpha \) is irreducible.

Conversely, if \( \alpha \) is irreducible, then \( \alpha \) is not a unit, so \((\alpha) \neq R\). Suppose \((\alpha) \subseteq (b) \subseteq R\). Then \( \alpha = bc \), so \( a = bc \) for some \( c \in R \) and then either \( c \) is a unit, in which case \((\alpha) = (b)\), or \( b \) is a unit, in which case \((b) = R\).

\[\square\]

In Example 2 in Subsection 7.3, we saw that, if \( F \) is a field and \( f \in F[x] \) with \( \deg(f) > 0 \), then the elements of the quotient ring \( F[x]/(f) \) correspond to polynomials in \( F[x] \) of degree less than \( f \), where multiplication is done modulo \( f \); that is, \( f \) becomes equal to 0 in the quotient ring.

When \( f \) is irreducible, Proposition 8.10 and Theorem 7.5 together tell us that \( F[x]/(f) \) is a field. The case \( F = \mathbb{Q} \) is particularly important, and as we are about to see, \( \mathbb{Q}[x]/(f) \) is isomorphic to a subfield of \( \mathbb{C} \) in this case.

Definition An element \( \alpha \in \mathbb{C} \) is said to be algebraic (over \( \mathbb{Q} \)) if it satisfies a polynomial equation \( f(\alpha) = 0 \) for some \( f \in \mathbb{Q}[x] \) with \( \deg(f) > 0 \). Otherwise \( \alpha \) is called transcendental.

So, for example, \( \sqrt{2} \) is algebraic, and it has been proved that the well-known constants \( \pi \) and \( e \) are transcendental.

37
We saw in Example 3 of Subsection 7.1 that, for any \( \alpha \in \mathbb{C} \), \( f(x) \mapsto f(\alpha) \) defines a ring homomorphism \( \phi : \mathbb{Q}[x] \rightarrow \mathbb{C} \). The property of \( \alpha \) being transcendental is equivalent to \( \ker(\phi) = \{0\} \), in which case the First Isomorphism Theorem for Rings (Theorem 7.4) tells us that \( \text{im}(\phi) \cong \mathbb{Q}[x] \).

When \( \alpha \) is algebraic, there exist non-zero \( f \in \ker(\phi) \) and, since \( \ker(\phi) \) is an ideal of the PID \( F[x] \), we have \( \ker(\phi) = (f) \) for some \( f \in F[x] \), where in fact we see from the proof of Theorem 8.3 that \( f \) is a polynomial of smallest possible non-zero degree in \( \ker(\phi) \).

Although \( f \) is not unique, any two possible \( f \) would divide each other and thus be associates. By multiplying by a constant, we can assume that \( f \) has leading coefficient equal to 1, and then it is unique, and is known as the minimal polynomial of \( \alpha \) (over \( \mathbb{Q} \)).

If \( f = gh \) with \( \deg(g), \deg(h) < \deg(f) \) then \( f(\alpha) = 0 \) implies \( g(\alpha) = 0 \) or \( h(\alpha) = 0 \), contradicting the fact that \( f \) has smallest possible degree in \( \ker(\phi) \). So \( f \) is irreducible.

Summing up, we have proved:

**Proposition 8.11** If \( \alpha \) is an algebraic element of \( \mathbb{C} \), then there is a unique non-zero polynomial \( f \in \mathbb{Q}[x] \) with leading coefficient 1 such that \( f(\alpha) = 0 \), and \( f \) is irreducible.

By the First Isomorphism Theorem for Rings, we have \( \text{im}(\phi) \cong \mathbb{Q}[x]/(f) \) which, as we saw above, is a field, so \( \text{im}(\phi) \) is a subfield of \( \mathbb{C} \), denoted by \( \mathbb{Q}[\alpha] \). Fields of this type are called number fields and are the object of study of Algebraic Number Theory.

As a simple example, consider \( \mathbb{Q}[\sqrt{2}] = \{ a + b\sqrt{2} \mid a, b \in \mathbb{Q} \} \), corresponding to \( f = x^2 - 2 \).

It is not completely obvious that this is a field, but this becomes clear from the identity

\[
\frac{a + b\sqrt{2}}{c + d\sqrt{2}} = \frac{ac - 2bd + (bc - ad)\sqrt{2}}{c^2 - 2d^2},
\]

which is derived by multiplying the top and bottom of the left hand side by \( c - d\sqrt{2} \).

### 8.4 Unique Factorisation Domains

**Definition** A domain \( R \) is a FD (factorisation domain) if each non-unit \( x \in R \setminus \{0\} \) admits a factorisation \( x = r_1 r_2 \cdots r_n \) where \( r_i \) are irreducible elements.

An FD \( R \) is a UFD (unique factorisation domain) if for any two factorisations of an element \( x = r_1 r_2 \cdots r_n = s_1 s_2 \cdots s_m \), where all \( r_i \) and \( s_i \) are irreducible, we have \( m = n \) and there exists \( \sigma \in S_n \) such that \( r_i \sim s_{\sigma(i)} \) for all \( i \).

**Proposition 8.12** Let \( R \) be a FD. Then \( R \) is a UFD if and only if every irreducible element is prime.

**Proof:** For the only if part we consider an irreducible element \( x \) such that \( x|ab \). Factorising \( a = r_1 \cdots r_k \) and \( b = r_{k+1} \cdots r_n \), we get a factorisation \( ab = r_1 \cdots r_n \). On the other hand, \( ab = xy \) for some \( y \in R \). Factorising \( y = s_1 \cdots s_t \), we get another factorisation \( ab = xs_1 \cdots s_t \).

By the UFD property, \( x \) is associate to \( r_i \) for some \( i \). If \( i \leq k \) then \( x|a \). If \( i > k \) then \( x|b \).

For the if part, suppose that every irreducible element is prime, and consider two factorisations \( x = r_1 r_2 \cdots r_n = s_1 s_2 \cdots s_m \) of a non-unit element \( x \) into irreducibles. We use induction on \( n + m \). If \( n = m = 1 \), then \( x = r_1 = s_1 \) and the result is clear. Otherwise, assuming without loss that \( n > 1 \), since \( r_n|x \) and \( r_n \) is prime, we have \( r_n|s_i \) for some \( i \) with \( 1 \leq i \leq m \).

Hence, \( r_n q = s_i \), and irreducibility of \( s_i \) implies that \( q \) is a unit. Now the result follows by applying the inductive assumption on \( r_1 r_2 \cdots r_{n-1} = (qs_1) \cdots s_{i-1} s_{i+1} \cdots s_m \). \( \Box \)
Theorem 8.13  A PID is a UFD. (So we have ED \implies PID \implies UFD.)

Proof: Using Propositions 8.8 and 8.12, it suffices to show that R is a FD.
We have to factorise an arbitrary non-unit \( x \in R \setminus \{0\} \) into irreducibles. Suppose there is such an \( x \) that cannot be so factorised. So \( x \) is not irreducible, and we can write \( x = x_{1,1}x_{1,2} \) where \( x_{1,i} \) are not units. At least one of \( x_{1,1} \) and \( x_{1,2} \) is not irreducible, so we can factorise it further.

We are going to repeat these steps over and over again. Step \( n + 1 \) starts with \( x = x_{n,1}x_{n,2} \cdots x_{n,k} \) where none of \( x_{n,i} \) are units. They cannot all be irreducible, or we would have successfully factorised \( x \), so pick all of \( x_{n,i} \) which are not irreducible, and write them as a product of two non-units resulting in \( x = x_{n+1,1}x_{n+1,2} \cdots x_{n+1,t} \). This process goes on forever.

It is now more or less clear that we can find an infinite sequence of elements \( x = x_0, x_1, x_2, \ldots \), where each \( x_n = x_{n,i} \) for some \( i \), \( x_{i+1}|x_i \) but \( x_i \nmid x_{i+1} \).

For those who prefer a more formal argument, we define a binary tree. The root of the tree is the element \( x \). The nodes at level \( n \) are elements \( x_{n,i} \) for all \( i \). If \( x_{n,i} \) is irreducible, it does not have any upward edges. If \( x_{n,i} = x_{n+1,j}x_{n+1,j+1} \), it has two upward edges going to \( x_{n+1,j} \) and \( x_{n+1,j+1} \). Since the process has not terminated, the tree is infinite. This means there is an infinite path in this tree starting from the root and going upward. Let \( x_n = x_{n,i} \) be the element of this infinite path at level \( n \). In particular, \( x_0 = x \).

We have done all this hard work to obtain the ascending chain of ideals

\[
\cdots \supset (x_{n+1}) \supset (x_n) \supset \cdots \supset (x_1) \supset (x_0)
\]

with all of the inclusions proper. Let \( I = \bigcup_{n=1}^{\infty} (x_n) \). Then \( I \) is an ideal. To see this, let \( r, s \in I \). Then, for some \( m, n \geq 0 \), we have \( r \in (x_m), r \in (x_m) \) and assuming without loss that \( m \geq n \), we have \( r, s \in (x_m) \), so \( r+s \in (x_m) \in I \). The other conditions for being an ideal can be similarly checked.

Since \( R \) is a PID, \( I = (d) \) for some \( d \in R \). Then \( d \in (y_n) \) for some \( n \). This implies that \( I = (d) \subseteq (y_n) \). But this contradicts the fact that \( (y_n) \) is properly contained in \( (y_{n+1}) \subseteq I \), and this contradiction completes the proof.

Examples 1. All of our EDs \( \mathbb{Z}, \mathbb{Z}[i], F[X] \) (\( F \) a field) are UFDs.
2. \( \mathbb{Z}[\sqrt{-5}] \) is a FD but not a UFD. We have seen that it is not a UFD since

\[
6 = 2 \cdot 3 = (1 + i\sqrt{5})(1 - i\sqrt{5})
\]

are two distinct factorisations into irreducibles. To prove it is an FD, use induction on \( |a| \). If \( |a| = 1 \), then \( a \) is a unit (in fact the only units are \( \pm 1 \)). If \( 0 \neq a = bc \) with \( b \) and \( c \) non-units, \( 0 < |b| < |a| \) and \( 0 < |c| < |a| \), so by inductive hypothesis \( b \) and \( c \) factorise into irreducibles, and hence so does \( a \).

3. \( \mathbb{Z}[x] \) is a UFD, which will be proved later, but not a PID. The ideal

\[
(2) + (x) = \{ 2a + xb \mid a, b \in \mathbb{Z}[x] \}
\]

is not principal.

Exercises Prove the assertion in Example 3 that \( (2) + (x) \) is not principal. Prove also that, for a field \( F \) and \( n > 1 \), the polynomial ring \( F[x_1, \ldots x_n] \) is not a PID.

We saw in Proposition 8.6 that any two elements in a PID have a gcd and an lcm. We now show that the same holds in a UFD.

39
Proposition 8.14 Any finite collection of elements in a UFD has a gcd and an lcm.

Proof: Let $a_1, a_2, \ldots, a_k$ be elements in a UFD $R$. Then each $a_i$ is divisible by only finitely many irreducible elements, where we regard associate elements as being the same. Let $p_1, p_2, \ldots, p_n$ be a complete list of the irreducible elements (or, more precisely, representatives of their equivalence classes under associativity) dividing any of the $a_i$. Then we can write each $a_i$ in the form $u_i p_1^{e_{i1}} p_2^{e_{i2}} \cdots p_n^{e_{in}}$, where each $e_{ij} \geq 0$ and $u_i$ is a unit. It is now not hard to see that

\[
\begin{align*}
gcd(a_1, a_2, \ldots, a_k) &= p_1^{m_1} p_2^{m_2} \cdots p_n^{m_n}, \quad \text{and} \\
\lcm(a_1, a_2, \ldots, a_k) &= p_1^{M_1} p_2^{M_2} \cdots p_n^{M_n},
\end{align*}
\]

where, for each $j$ with $1 \leq j \leq n$, $m_j = \min(e_{1j}, e_{2j}, \ldots, e_{kj})$ and $M_j = \max(e_{1j}, e_{2j}, \ldots, e_{kj})$.

\[\square\]

8.5 Primes in $F[x]$

Throughout this section $F[x]$ will be the polynomial ring over a field $F$. Since this is an ED, primes and irreducibles are the same. We start with the following easy observation.

Proposition 8.15 (Remainder Theorem) Let $f = f(x) \in F[x]$. Then for $a \in F$, $f(a) = 0$ if and only if $x - a$ divides $f$.

Proof: Since $F[x]$ is an ED, we can divide $f(x)$ by $x - a$ with a remainder:

\[f(x) = g(x)(x - a) + r.\]

Since $\deg(x - a) = 1$, we have $r = 0$ or $\deg(r) < 1$, so $r \in F$ (a constant polynomial). Substituting $x = a$, we arrive at $f(a) = r$, and the result follows. \[\square\]

Corollary 8.16 If $0 \neq f \in F[x]$, then $f(a) = 0$ for at most $\deg(f)$ distinct values of $a \in F$. (In other words, a polynomial of degree $d$ has at most $d$ roots.)

Proof: Induction on $\deg(f)$. If $\deg(f) = 0$ then, since $f \neq 0$, it is a non-zero constant, so $f(a)$ is never 0. If $\deg(f) > 0$ and $f(a) = 0$, then by Proposition 8.15, $f = (x - a)g$ with $\deg(g) = \deg(f) - 1$. Now, if $f(b) = 0$, then either $b = a$ or $g(b) = 0$, and by inductive hypothesis there are at most $\deg(f) - 1$ possible $b$ with $f(b) = 0$, so the result follows. \[\square\]

Theorem 8.17 Let $F$ be a field. Then any finite subgroup of the multiplicative group $F \setminus \{0\}$ is a cyclic group.

Proof: Suppose $G \leq F \setminus \{0\}$ is not cyclic, and let $|G| = N$. By the classification of finite abelian groups (from Algebra 1), $G$ is isomorphic to $C_{n_1} \times \cdots \times C_{n_m}$, a direct product of cyclic groups of orders $n_1 | n_2 | \ldots | n_m$ and $N = n_1 n_2 \cdots n_m$. Since $G \leq F \setminus \{0\}$ is not cyclic, $m > 1$. Now, putting $n = n_m$, we have $(x_1, \ldots, x_m)^n = (x_1^n, \ldots, x_m^n) = (1, \ldots, 1)$ for all $x_j \in C_{n_j}$, and so $g^n = 1$ for any $g \in G$. But then, for $f(x) = x^n - 1$, there are $N > n$ elements $a \in F$ with $f(a) = 0$, contradicting Corollary 8.16. \[\square\]

Using Proposition 6.6, we have the following corollary:

Corollary 8.18 If $p$ is prime, then the multiplicative group $\mathbb{Z}_p \setminus \{0\}$ is a cyclic group of order $p - 1$.

Definition A field $F$ is algebraically closed if for any $f(x) \in F[x]$ of degree at least 1 there exists $a \in F$ such that $f(a) = 0$. 

40
Example $\mathbb{C}$ is the best known example of an algebraically closed field. Another example is the subfield $\mathbb{A}$ of $\mathbb{C}$ consisting of the algebraic numbers, which are the numbers $a \in \mathbb{C}$ with $f(a) = 0$ for some $f \in \mathbb{Q}[x]$. (It is not obvious that $\mathbb{A}$ is a field, but it is!)

**Proposition 8.19** If $F$ is an algebraically closed field then the irreducibles in $F[x]$ are the polynomials of degree 1. So each irreducible is an associate of $x - a$ for a unique $a \in F$.

**Proof:** The element $x - a$ is irreducible because any of its divisors must have degree 1 or 0. If it is 0, the divisor is a unit. If it is 1, the divisor is associate to $x - a$.

To show that $x - a$ and $x - b$ are not associate when $a \neq b$, notice that since the units in $F[x]$ are the non-zero constant polynomials, $x - a$ is associate only to $cx - ca$ for all $c \in F^*$.

Finally, if we have an irreducible $f \in F[x]$, then $f$ has degree at least 1. Since $F$ is algebraically closed, there is $a \in F$ such that $f(a) = 0$. By Proposition 8.15, $x - a$ divides $f$. Hence $f$ is associate to $x - a$ and $\deg(f) = 1$.

You should prove the next proposition yourself. It is an excellent exercise but we won’t use it in this course.

**Proposition 8.20** The monic primes in $\mathbb{R}[x]$ are $x - a$ and $x^2 + bx + c$ for all possible $a, b, c \in \mathbb{R}$ with $b^2 - 4c < 0$.

It is much more difficult to test whether a polynomial in $\mathbb{Q}[x]$ is prime, and there is no easily stated necessary and sufficient condition for this. We shall see later in Theorem 9.5 that this problem is essentially equivalent to testing polynomials in $\mathbb{Z}[x]$ for irreducibility in $\mathbb{Z}[x]$. Factorising polynomials in $\mathbb{Z}[x]$ into irreducibles is one of the major research problems in computer algebra.

### 8.6 Gaussian primes

In this section, we investigate which elements of $\mathbb{Z}[i]$ are prime. We will call them Gaussian primes. Again $\mathbb{Z}[i]$ is an ED, so primes and irreducibles are the same.

Let us recall that $\nu(x) = |x|^2 = xx^*$, where $x^*$ is the complex conjugate of $x$. It is useful to remember that if $xy \in \mathbb{Z}[i]$ then $\nu(x)\nu(y)$ in $\mathbb{Z}$.

**Proposition 8.21** If $x \in \mathbb{Z}[i]$ and $\nu(x)$ is prime in $\mathbb{Z}$ then $x$ is Gaussian prime.

**Proof:** Suppose $y|x$. Hence $\nu(y)|\nu(x) = p$ in $\mathbb{Z}$, which forces $\nu(y)$ to be $p$ or 1. If $\nu(y) = p$ then $y$ is associate to $x$. If $\nu(y) = 1$ then $y$ is a unit. So $x$ is irreducible and hence prime. \(\square\)

**Proposition 8.22** Let $p \in \mathbb{Z}$ be a prime in $\mathbb{Z}$. Then either $p$ is a Gaussian prime or $p = xx^*$ are Gaussian primes.

**Proof:** If $p$ is not a Gaussian prime, then there exists a Gaussian prime $x$ such that $p = xy$ and neither $x$ nor $y$ is a unit. Hence, $\nu(x)\nu(y) = \nu(p) = p^2$. This forces $\nu(x) = \nu(y) = p$, which makes $x$ and $y$ prime by Proposition 8.21. Finally, $xx^* = \nu(x) = p = xy$, so $x^* = y$. \(\square\)

**Proposition 8.23** Let $q \in \mathbb{Z}[i]$ be a Gaussian prime. Then either $\nu(q)$ is a prime or a square of a prime.

**Proof:** Let $n = \nu(q) = qq^*$. Take the decomposition of $n$ into primes in $\mathbb{Z}$, say $n = p_1 \cdots p_t$. Then $q|p_j$ in $\mathbb{Z}[i]$ for some $j$ and hence $\nu(q)$ divides $\nu(p_j) = p_j^2$. \(\square\)
Theorem 8.24 The prime elements in \( \mathbb{Z}[i] \) are obtained from the prime elements \( \mathbb{Z} \). Each prime \( p \in \mathbb{Z} \) congruent to 3 modulo 4 is a Gaussian prime. The prime \( p = 2 \) gives rise to a Gaussian prime \( q \) such that \( 2 = qq^* \sim q^2 \). Each prime \( p \in \mathbb{Z} \) congruent 1 modulo 4 gives rise to two conjugate Gaussian primes \( q \) and \( q^* \) such that \( p = qq^* \).

**Proof:** Let \( q \) be a Gaussian prime. Then, by Proposition 8.23, \( \nu(q) = p \) or \( p^2 \) for some prime \( p \in \mathbb{Z} \). In the first case \( \nu(q) = p \), Proposition 8.22 tells us that \( p \) gives rise to two Gaussian primes \( q, q^* \) such that \( p = qq^* = \nu(q) \).

In the second case \( \nu(q) = p^2 \), since \( q \) is prime in \( \mathbb{Z}[i] \) and \( q \mid p^2 \), we have \( q \mid p \) in \( \mathbb{Z}[i] \), and hence \( p = qs \) for some \( s \in \mathbb{Z}[i] \). Then \( |s| = |p|/|q| = \nu(p)/\nu(q) = p^2/p^2 = 1 \), so \( s \) is a unit \( (s^{-1} = s^*/|s|^2) \). Hence \( q \) is associate to \( p \). We saw earlier that the units in \( \mathbb{Z}[i] \) are \( \pm 1 \) and \( \pm i \), so \( q = \pm p \) or \( \pm ip \) in this case.

It remains to see that everything is controlled by the value of \( p \) modulo 4. If \( p = qq^* \) and \( q = x + yi \) then \( p = qq^* = x^2 + y^2 \). Since \( 1^2 = 1, 2^2 = 0, 3^2 = 1 \) in \( \mathbb{Z}_4 \), we cannot have \( x^2 + y^2 = 3 \) in \( \mathbb{Z}_4 \), and so a prime in \( \mathbb{Z} \) congruent to 3 modulo 4 cannot be represented as \( qq^* \). Thus, it must be prime in \( \mathbb{Z}[i] \) as well.

For \( p = 2 \), we have \( 2 = (1 - i)(1 + i) \) where, by Proposition 8.21, \( 1 \pm i \) are Gaussian primes. Since \( (1 - i) = -i(1 + i) \) and \( -i \) is a unit, \( (1 - i) \) and \( (1 + i) \) are associates in \( \mathbb{Z}[i] \), so \( 2 \sim q^2 \) with \( q = 1 + i \).

Finally, we must prove that a prime \( p \) congruent to 1 modulo 4 is not a Gaussian prime. We saw in Proposition 8.18 that the multiplicative group \( \mathbb{Z}_p \setminus \{0\} \) is cyclic of order \( p - 1 \), and since \( 4|p - 1 \), this group has an element \( a \) of order 4. Now the polynomial \( x^2 - 1 \) of degree 2 has at most 2 roots in \( \mathbb{Z}_p \), and since 1 and \( -1 (= p - 1) \) are roots, \( -1 \) is the only element of order 2 in \( \mathbb{Z}_p \setminus \{0\} \). Since \( a^2 \) has order 2, we have \( a^2 = -1 \) in \( \mathbb{Z}_p \). So in \( \mathbb{Z} \), we have that \( a^2 + 1 \equiv 0 \mod{p} \), and there exists \( b \in \mathbb{Z} \) with \( a^2 + 1 = kp \), so \( (a + i)(a - i) = kp \) in \( \mathbb{Z}[i] \).

If \( p \) were a Gaussian prime, then we would have \( p|a + i \) or \( p|a - i \) in \( \mathbb{Z}[i] \), which is clearly false, since multiples of \( p \) in \( \mathbb{Z}[i] \) have the form \( px + ipy \) for \( x, y \in \mathbb{Z} \). So \( p \) is not a Gaussian prime when \( p \) is congruent to 1 mod 4, which completes the proof.

Corollary 8.25 (Fermat) The prime 2 and every prime congruent to 1 modulo 4 is a sum of positive integer squares in a unique way.

**Proof:** If \( p = 2 \) or \( p \) is congruent to 1 mod 4 then by Theorem 8.24 we have \( p = qq^* = a^2 + b^2 \), where \( q = a + ib \) is a Gaussian prime. If \( p = a^2 + b^2 = c^2 + d^2 \) with \( a, b, c, d > 0 \), then \( p = (a + ib)(a - ib) = (c + id)(c - id) \) are two prime decompositions in \( \mathbb{Z}[i] \). Since \( \mathbb{Z}[i] \) is a UFD and the only units in \( \mathbb{Z}[i] \) are \( \pm 1 \) and \( \pm i \), we have \( a + ib = \pm(c + id) \) or \( a + ib = \pm i(c + id) \) and hence \( a = c, b = d \) or \( a = d, b = c \).

As an application, we will now determine which integers can be expressed as the sum of two integer squares. So let \( S = \{a^2 + b^2 \mid a, b \in \mathbb{Z}\} \). Clearly \( c \geq 0 \) for all \( n \in S \).

Theorem 8.26 We have \( S = T \), where \( T = \{n^2p_1p_2 \cdots p_k \mid n, k \in \mathbb{Z}, k \geq 0, \) and the \( p_i \) are distinct positive primes in \( \mathbb{Z} \), each equal to 2 or congruent to 1 modulo 4 \).

**Proof:** First note that, for \( a^2 + b^2 \) and \( c^2 + d^2 \) in \( S \), we have
\[
(a^2 + b^2)(c^2 + d^2) = \nu(a + ib)^{\nu(c + id)} = \nu((a + ib)(c + id)) = 
\nu((ac - bd) + i(ad + bc)) = (ac - bd)^2 + (ad + bc)^2,
\]
so \( S \) is closed under multiplication.
Clearly \( n^2 \in S \) for any \( n \in \mathbb{Z} \), and if \( p > 0 \) is prime with \( p_i = 2 \) or \( p_i \) congruent to 1 mod 4 then by Theorem 8.24 we have \( p = qq^* = a^2 + b^2 \), where \( q = a + ib \) is a Gaussian prime, so \( p \in S \). So \( T \subseteq S \).

Assume that \( S \neq T \) and let \( c = a^2 + b^2 \) be the smallest element of \( S \setminus T \). We can write any positive integer as \( n^2 p_1 p_2 \cdots p_k \) for distinct primes \( p_i \), so \( c \notin T \) implies that some \( p_i \) is congruent to 3 mod 4. Put \( p = p_i \). Then, by Theorem 8.24, \( p \) is a Gaussian prime dividing \( a^2 + b^2 = (a + ib)(a - ib) \), so \( p \) divides \( a + ib \) or \( a - ib \). In either case, this implies \( p \mid a \) and \( p \mid b \), so in fact \( p^2 \mid c \). Since the primes \( p_i \) are distinct, this forces \( p \mid n \) and hence \( p \mid n^2 \). But then \( (a/p)^2 + (b/p)^2 = c/p^2 = (n/p)^2 p_1 p_2 \cdots p_k \) is a smaller element of \( S \setminus T \). This contradiction shows that \( S = T \). \( \square \)

Remark. Not all positive integers (for example 7) are the sum of three integer squares, but there is a famous theorem of Lagrange that says that all positive integers can be expressed as the sum of four integer squares.

You should be able to prove for yourself that there are infinitely many primes in \( \mathbb{N} \) that are congruent to 3 modulo 4. Just adapt the proof that there are infinitely many primes. It is harder to prove that there are infinitely many primes congruent to 1 modulo 4, but we can do that as another application.

(There is a much more advanced result in Analytic Number Theory, due to Dirichlet, which says that any arithmetic progression \( \{ a + nd \mid n \in \mathbb{N} \} \) in which \( a \) and \( d \) are coprime integers contains infinitely many primes.)

Corollary 8.27 There are infinitely many primes in \( \mathbb{N} \) that are congruent to 1 modulo 4.

Proof: Suppose there are only finitely many, say \( p_1, \ldots, p_n \). Let \( p_0 = 2 \), \( q_0 = 1 + i \), \( p_j = q_j q_j^* \) be a prime decomposition of \( p_j \). Let us consider a prime decomposition of \( x = 2p_1 p_2 \cdots p_n + i \in \mathbb{Z}[i] \). No prime \( p \in \mathbb{Z} \) congruent to 3 modulo 4 divides \( x \), because \( x/p \) has 1/p as the coefficient at i, so it is not in \( \mathbb{Z}[i] \). Hence, one of the Gaussian primes \( q_j \) (if it is \( q_j^* \) swap the notation between \( q_j \) and \( q_j^* \)) divides \( x \). So \( p_j = \nu(q_j) \nu(x) = 4p_1^2 p_2^2 \cdots p_n^2 + 1 \), which is a contradiction since \( \nu(x) \) is congruent to 1 modulo all \( p_j \). \( \square \)

9 Polynomial Rings over UFDs

9.1 Fields of fractions

In Foundations, the construction of the field \( \mathbb{Q} \) of rationals from the integers \( \mathbb{Z} \) was described in outline. We can perform this construction with any integral in place of \( \mathbb{Z} \). The details are straightforward but unfortunately a little tedious!

Let \( R \) be a domain. We consider the set \( W = R \times (R \setminus \{0\}) = \{ (x, y) \in R \times R \mid y \neq 0 \} \). We define an equivalence relation on \( W \) by \( (a, b) \sim (c, d) \) whenever \( ad = bc \). It left as an exercise to show that this is, indeed, an equivalence relation. An equivalence class of \( (a, b) \) is called a fraction and denoted \( a/b \). Let \( Q = Q(R) \) be the set of all of the equivalence classes on \( W \).

Proposition 9.1 If \( R \) is a domain then \( Q(R) \) is a field under the operations

\[
\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd},
\]

and \( \pi : R \to Q(R) \), with \( \pi(r) = r/1 \) is an injective ring homomorphism

Proof: We have to show that these operations are well-defined. Then we have to establish all the axioms of a field. Finally we have to show that \( \pi \) is an injective ring homomorphism.
To show that the operations are well defined, we need to observe first that the denominators of the results are non-zero because \( R \) is a domain. It remains to prove that the result is independent of the representative of the equivalence class. Given \( a/b = x/y \) and \( c/d = u/w \), we need to show that \( ac/bd = xu/yw \) and \( (ad + bc)/bd = (xw + yu)/yw \). The first equality requires \( acyw = bdxu \), which follows directly from \( ay = bx \) and \( cw = du \). The second equality requires \( adyw + bcyw = bdxw + bdyu \). Rewriting it, we get \( adyw - bdxw = bdyu - bcyw \), which holds because

\[
adyw - bdxw = dw(ay - bx) = 0 \text{ and } bdyu - bcyw = by(du - cw) = 0.
\]

The list of axioms of a field is long and we have to go and check them all. But we are in a good shape because we know that the operations are well-defined, so we can use our usual intuition about fractions. The associativity of addition is probably the hardest axiom to check

\[
\left( \frac{a}{b} + \frac{c}{d} \right) + \frac{e}{f} = \frac{ad + bc}{bd} + \frac{e}{f} = \frac{adf + (bcf + bde)}{bdf} = \frac{a}{b} + \frac{cf + de}{df} = \frac{a}{b} + \left( \frac{c}{d} + \frac{e}{f} \right).
\]

The commutativity of addition is easier:

\[
\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} = \frac{c + a}{d + b}.
\]

The zero and the additive inverse are usual: \( 0 = 0/1 \) and \( -(a/b) = (-a)/b \) with all the checks routine. The associativity of multiplication is straightforward

\[
\left( \frac{a}{b} \cdot \frac{c}{d} \right) \cdot \frac{e}{f} = \frac{ac}{bd} \cdot \frac{e}{f} = \frac{ace}{bdf} = \frac{a}{b} \cdot \frac{ce}{df} = \frac{a}{b} \cdot \frac{c}{d} \cdot \frac{e}{f} = \frac{a}{b} \cdot \frac{c}{d} \cdot \frac{e}{f}
\]

as well as the commutativity:

\[
\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} = \frac{c}{d} \cdot \frac{a}{b}.
\]

The unity and the multiplicative inverse are usual: \( 1 = 1/1 \) and \( (a/b)^{-1} = b/a \) with all the checks routine. It is worth noticing though why \( a \neq 0 \). Indeed, \( a = 0 \) if and only if \( a \cdot 1 = b \cdot 0 \) if and only \( a/b = 0/1 = 0 \). Finally, we have to check distributivity but it suffices to do it on one side only because the multiplication is commutative:

\[
\left( \frac{a}{b} + \frac{c}{d} \right) \cdot \frac{e}{f} = \frac{ad + bc}{bd} \cdot \frac{e}{f} = \frac{ade + bce}{bdf} = \frac{ad}{bdf} + \frac{bce}{bdf} = \frac{a}{b} \cdot \frac{e}{f} + \frac{c}{d} \cdot \frac{e}{f}.
\]

The map \( \pi \) is a ring homomorphism because \( \pi(1_R) = 1/1 = 1_Q \),

\[
\pi(xy) = \frac{xy}{1} = \frac{x}{1} \cdot \frac{y}{1} = \pi(x) \cdot \pi(y), \quad \pi(x + y) = \frac{x + y}{1} = \frac{x}{1} + \frac{y}{1} = \pi(x) + \pi(y).
\]

Finally, \( x \) is in the kernel if and only if \( x/1 = 0/1 \) if and only if \( x \cdot 1 = 0 \cdot 1 \) if and only if \( x = 0 \).

\[\square\]

**Definition** \( Q = Q(R) \) is called the field of fractions of a domain \( R \).

**Examples 1.** \( Q(\mathbb{Z}) = \mathbb{Q} \).

2. \( Q(F[x]) \) is equal to the field of rational functions \( p/q \) (\( p, q \in F[x], q \neq 0 \)) in one variable \( x \). This is commonly denoted by \( F(x) \) (so \( F[x] \) and \( F(x) \) are not the same!).

3. \( Q(\mathbb{Z}[i]) = \mathbb{Q}[i] \).
9.2 Polynomial rings over UFDs are UFDs

Our goal in this subsection is to prove that if \( R \) is a UFD then so is \( R[x] \).

If \( R \) is an integral domain and \( f, g \in R[x] \), then \( \deg(fg) = \deg(f) + \deg(g) \), and hence:

**Proposition 9.2** If \( R \) is an integral domain then so is \( R[x] \).

Another easy result is:

**Proposition 9.3** If \( R \) is an integral domain then an irreducible element of \( R \) remains irreducible in \( R[x] \), and the units in \( R \) and in \( R[x] \) are the same.

Note that some of these properties can fail and give surprising results when \( R \) is not a UFD. For example, if \( R = \mathbb{Z}_4 \) (the ring of integers modulo 4) and \( f = 2x + 1 \), we get \( f^2 = 1 \), so \( f \) is a unit in \( R[x] \setminus R \). But here we shall confine our attention to domains \( R \), for which things behave more predictably.

For the remainder of this section, we shall assume that \( R \) is a UFD. We saw in Proposition 8.14 that any finite set of elements in \( R \) has a gcd.

**Definition** An element \( 0 \neq f = a_0 + a_1x + \cdots + a_nx^n \in R[x] \) is called primitive if \( \gcd(a_0, a_1, \ldots, a_n) = 1 \).

So any non-zero \( f \in R[x] \) can be written as \( af_0 \), where \( a \in R \) if the gcd of the coefficients of \( f \), and \( f_0 \) is primitive.

**Proposition 9.4** The product of two primitive polynomials is primitive.

**Proof:** Let \( f = a_0 + a_1x + \cdots + a_nx^n \) and \( g = b_0 + b_1x + \cdots + b_nx^n \) be primitive, and suppose that some irreducible \( p \) divides \( c_i \) for \( 0 \leq i \leq m + n \), where \( fg = c_0 + c_1x + \cdots + c_{m+n}x^{m+n} \).

If \( f \) is primitive, \( p \) cannot divide every \( a_i \), so suppose that \( p \mid a_i \) for \( 0 \leq i < k \), but \( p \mid b_k \), where \( k \geq 0 \). Similarly, choose \( l \) such that \( p \mid b_l \) for \( 0 \leq l < k \), but \( p \mid a_l \), where \( l \geq 0 \). We have \( c_{k+l} = \sum_{i=0}^{k+l} a_ib_{k+l-i} \), where we take any undefined coefficients to be 0. Since \( p \) is prime that does not divide \( a_k \) or \( b_l \), it does not divide \( a_kb_l \), but \( p \) does divide every other term in this sum. So \( p \nmid c_{k+l} \), which is a contradiction. Hence \( f \) must be primitive.

**Theorem 9.5** Let \( R \) be a UFD with field of fractions \( Q = Q(R) \). Then a primitive polynomial in \( R[x] \) is irreducible if and only if it is irreducible in \( Q[x] \).

**Proof:** Let \( f \in R[x] \) be primitive. If \( \deg(f) = 0 \) then \( f \) is a unit in \( R \) and so is irreducible in neither \( R[x] \) nor \( Q[x] \). So assume that \( \deg(f) > 0 \). Suppose that \( f = gh \) with \( g, h \in R[x] \) and \( g, h \) non-units. By primitivity, we cannot have \( \deg(g) = 0 \) or \( \deg(h) = 0 \), so if \( f \) is reducible in \( R[x] \) then it is also reducible in \( Q[x] \).

Conversely, assume that \( f \) is primitive and reducible in \( Q[x] \), so \( f = gh \) with \( g, h \in Q[x] \) and \( g, h \) non-units – so \( \deg(g), \deg(h) > 0 \). Let \( a_1 \) be the least common multiple of all the denominators of the coefficients of \( g(x) \), so \( a_1g \in R[x] \). Now let \( a_2 \) be the greatest common divisor of all the coefficients of \( a_1g(x) \), Put \( a = a_1/a_2 \). Then \( a \in Q \) and \( ag \in R[x] \) with \( ag \) primitive. Similarly, we define \( b = b_1/b_2 \in Q \) with \( bh \in R[x] \) and \( bh \) primitive. So, by Proposition 9.4, \( f' := abgh = abf \) is primitive, and hence \( a_2b_2f' = a_1b_1f \) with \( f \) and \( f' \) both primitive. So \( a_1b_1 \) and \( a_2b_2 \) are both equal to the gcd of the coefficients of \( a_1b_1f \) and, by uniqueness of the gcd, \( a_1b_1 \) and \( a_2b_2 \) are associates in \( R \). But this means that \( ab = (a_1b_1)/(a_2b_2) \) is a unit \( u \in R \), and then \( f = (ag)(u^{-1}bh) \) is a factorisation of \( f \) in \( R[x] \), so \( f \) is reducible in \( R[x] \).
In particular, we have Gauss’ Lemma, which says that a primitive irreducible polynomial in \( \mathbb{Z}[x] \) remains irreducible in \( \mathbb{Q}[x] \).

**Corollary 9.6** If \( R \) is a UFD then there are two kinds of irreducibles in \( R[X] \): irreducible elements in \( R \) and primitive elements in \( R[X] \) that are irreducible in \( \mathbb{Q}[X] \).

**Proof:** This follows from Proposition 9.3, Theorem 9.5, and the fact that a polynomial of non-zero degree in \( R[x] \) that is not primitive is reducible. \( \square \)

**Theorem 9.7** If \( R \) is a UFD then so is \( R[x] \).

**Proof:** It is clear that we can factorise any \( f \in R[x] \) into a product of irreducible elements of \( R \) and irreducible primitive polynomials in \( R[x] \). For uniqueness, suppose that

\[
p_1p_2\cdots p_kf_1f_2\cdots f_n = q_1q_2\cdots q_lg_1g_2\cdots g_m
\]

be two factorisations of the same polynomial in \( R[x] \), where the \( p_i \) and \( q_i \) are irreducibles in \( R \), and the \( f_i \) and \( g_i \) are irreducible primitive polynomials in \( R[x] \).

By Theorem 9.5, the \( f_i \) and \( g_i \) are also irreducible elements of \( Q[x] \) (with \( Q \) the field of fractions of \( R \)), whereas the \( p_i \) and \( q_i \) are units in \( Q[x] \). We saw earlier that \( Q[x] \) is an ED, so it is a UFD. Hence \( n = m \) and, after permuting the \( g_i \) if necessary, \( f_i \) and \( g_i \) are associates in \( Q[x] \) for \( 1 \leq i \leq n \). This means that, for each \( i \), we have \( g_i = a_if_i \) for some \( a_i \in Q \). But then \( a_i = b_i/c_i \) with \( b_i, c_i \in R \), and \( c_if_i = b_ig_i \). Since \( f_i \) and \( g_i \) are primitive, \( c_i \) and \( b_i \) are both geds of the coefficients of \( c_if_i \), so \( b_i \) and \( c_i \) are associates in \( R \), and hence the \( a_i \) are all units in \( R \).

We can now cancel the \( f_i \) and conclude that \( p_1p_2\cdots p_k = q_1q_2\cdots q_l \), where \( a = a_1a_2\cdots a_n \) is a unit in \( R \). It now follows from the fact that \( R \) is a UFD that \( k = l \) and, after permuting the \( q_i \) if necessary, \( p_i \) and \( q_i \) are associates for all \( i \). So we have uniqueness. \( \square \)

**Examples 1.** \( Z[x] \) is a UFD. We saw earlier that \( Z[x] \) is not a PID.

2. By repeated application of Theorem 9.7, we find that \( R[x_1, x_2, \ldots, x_n] \) is a UFD for any \( n > 0 \) and any UFD \( R \).

As we remarked earlier, it is not easy to check polynomials in \( Z[x] \) for irreducibility. There is however a sufficient (but not necessary) condition, known as Eisenstein’s criterion that is often useful. We shall state it for \( Z[x] \), but the same proof words for \( R[x] \) for any UFD \( R \).

**Proposition 9.8** Let \( f = a_0 + a_1x + \cdots + a_nx^n \) be a primitive polynomial in \( Z[x] \), and suppose there is prime \( p \in Z \) such that \( p \not| a_n \), \( p | a_i \) for \( 0 \leq i < n \) and \( p^2 \not| a_0 \). Then \( f \) is irreducible in \( Z[x] \) (hence also in \( Q[x] \)).

**Proof:** Since \( f \) is primitive, if it is reducible in \( Z[x] \) then \( f = (b_mx_m + \cdots + b_0)(c_{x^l} + \cdots + c_0) \) with each \( b_i, c_i \in Z \) and \( 0 < l, m < n \). Then \( a_0 = b_0c_0 \) and, since \( p | a_0 \) but \( n^2 \not| a_0 \), \( p \) must divide one, but not both, of \( b_0 \) and \( c_0 \). Suppose without loss that \( p | b_0 \), \( p \not| c_0 \). Now \( a_1 = b_0c_1 + b_1c_0 \) so \( p | a_1 \) implies \( p \mid b_1c_0 \), and since \( p \not| c_0 \), we get \( p | b_1 \). Similarly, if \( n \geq 3 \), then \( p | a_2 = b_0c_2 + b_1c_1 + b_2c_0 \) implies \( p | b_2 \) and, since \( m < n \), we can prove by induction that \( p | b_m \). But then \( p | a_n = b_mc_n \), contrary to assumption. This contradiction proves that \( f \) is irreducible. \( \square \)

**Examples 1.** \( 2 + 12x + 10x^2 + 3x^3 \) is irreducible using the prime 2.

2. We could replace the condition “\( p \not| a_n \), \( p | a_i \) for \( 0 \leq i < n \) and \( p^2 \not| a_0 \)” by “\( p \not| a_0 \), \( p | a_i \) for \( 0 < i < n \) and \( p^2 \not| a_n \)” in Proposition 9.8, and the result would still be true, using a similar proof. So, for example \( 4 - 9x + 36x^2 - 3x^3 \) is irreducible, using the prime 3.

46
9.3 The cyclotomic polynomials

For a positive integer $n$, let $\zeta_n = e^{2\pi i/n}$. Then $\zeta_n$ is a complex $n$-th root of unity. The $n$ powers $1 = \zeta_n^0, \zeta_n^1, \ldots, \zeta_n^{n-1}$ form all of the $n$-th roots of unity. These are distinct roots of the polynomial $x^n - 1$ and so in $\mathbb{C}[x]$ we have the factorisation

$$x^n - 1 = \prod_{i=0}^{n-1} (x - \zeta_n^i)$$

(9.9)

Note that the set of $n$-th roots of unity forms a cyclic group under multiplication with generator $\zeta_n$. An $n$-th root of unity is called primitive if it is not an $m$-th root of $1$ for any $m \in \mathbb{Z}$ with $1 \leq m < n$. So the primitive roots are those of order $n$ in the group; that is, the generators of the group.

Recall that gcd($0,n$) = $n$ for any $n > 0$.

**Lemma 9.10** Let $n > 0$, and for $0 \leq i < n$ define $d_i = \gcd(n, i)$. Then $\zeta_n^i$ has order $n/d_i$, and so is a primitive $(n/d_i)$-th root of unity. In particular, $\zeta_n^i$ is a primitive $n$-th root of $1$ if and only if $d_i = 1$.

**Proof:** Since $d_i|n$, $(\zeta_n^i)^{n/d_i} = (\zeta_n^m)^{i/d_i} = 1$, so the order of $\zeta_n^i$ divides $n/d_i$. On the other hand, if $(\zeta_n^i)^m = 1$ for some $m > 0$, then $n$ divides $m$, so $n/d_i$ divides $m(i/d_i)$. But by the definition of gcd, $n/d_i$ and $i/d_i$ are coprime, so $n/d_i$ divides $m$. So the order of $\zeta_n^i$ is exactly $n/d_i$ as claimed. □

For each integer $n > 0$, we define the $n$-th cyclotomic polynomial to be

$$\Phi_n(x) = \prod_{0 \leq i < n\atop \gcd(i,n) = 1} (x - \zeta_n^i)$$

So $\deg(\Phi_n(x)) = \phi(n)$, where $\phi$ is the Euler phi-function, defined as

$$\phi(n) = \{ i \mid 0 \leq i < n, \gcd(i,n) = 1 \}$$

for integers $n > 0$.

From the lemma above, we see that each power $\zeta_n^i$ is a $d$-th root of unity for some divisor $d$ of $n$. Conversely, for any $d|n$, the powers $\zeta_n^i$ of $\zeta_n$ with $0 \leq i < n$ for which $(n/d)|i$ are precisely the $d$-th roots of unity, so the sequence $\{ \zeta_n^i \mid 0 \leq i < n \}$ contains each $d$-th root of unity and hence each primitive $d$-th root of unity exactly once. So, from Equation 9.9, we have:

$$x^n - 1 = \prod_{d|n} \Phi_d(x)$$

(9.11)

This equation enables us to compute $\Phi_d(x)$ recursively. Here are the first few:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\Phi_d(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$x - 1$</td>
</tr>
<tr>
<td>2</td>
<td>$x + 1$</td>
</tr>
<tr>
<td>3</td>
<td>$x^2 + x + 1$</td>
</tr>
<tr>
<td>4</td>
<td>$x^2 + 1$</td>
</tr>
<tr>
<td>5</td>
<td>$x^4 + x^3 + x^2 + x + 1$</td>
</tr>
<tr>
<td>6</td>
<td>$x^2 - x + 1$</td>
</tr>
<tr>
<td>7</td>
<td>$x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$</td>
</tr>
<tr>
<td>8</td>
<td>$x^4 + 1$</td>
</tr>
<tr>
<td>9</td>
<td>$x^6 + x^3 + 1$</td>
</tr>
<tr>
<td>10</td>
<td>$x^4 - x^3 + x^2 - x + 1$</td>
</tr>
</tbody>
</table>

Observe, in particular that, for $p$ prime,

$$\Phi_p(x) = (x^p - 1)/(x - 1) = x^{p-1} + x^{p-2} + \cdots + x + 1.$$
Proposition 9.12 For each $n \geq 0$, $\Phi_n(x)$ is monic and has coefficients in $\mathbb{Z}$.

Proof: Induction on $n$. Since $\Phi_1(x) = x - 1$, the result is true for $d = 1$, so suppose $n > 1$. Then by Equation 9.11, we have $x^n - 1 = f(x)\Phi_n(x)$, where $f(x)$ is the product of the $\Phi_d(x)$ over all divisors of $n$ other than $n$ itself. By inductive hypothesis, $f(x)$ is monic and has coefficients in $\mathbb{Z}$. So $\Phi_n(x)$ is the result of dividing $x^n - 1$ by $f(x)$. By considering the usual process of dividing one polynomial by another and the fact that $f(x)$ is monic, we see that $\Phi_n(x)$ must also be monic have integers as coefficients.

It was at one time conjectured that the only coefficients that can occur in $\Phi_n(x)$ are 0, 1, and −1, but this turned out to be false. For example, $\Phi_{105}(x)$:

\[
x^{48} + x^{47} + x^{46} - x^{42} - 2x^{41} - x^{39} + x^{36} + x^{35} + x^{33} + x^{32} + x^{31} - x^{28} - x^{26} - x^{24} - x^{22} + x^{20} + x^{17} + x^{16} + x^{15} + x^{14} + x^{13} + x^{12} - x^9 - x^8 - 2x^7 - x^6 - x^5 + x^2 + x + 1
\]

Our final objective is to prove:

Theorem 9.13 $\Phi_n$ is irreducible in $\mathbb{Z}[x]$ for all $n > 0$.

For $n$ prime, although Eisenstein’s irreducibility criterion does not apply directly to $\Phi_n(x)$, it can be applied to show that $\Phi_n(x + 1)$ and hence also $\Phi_n(x)$ is irreducible - this will be an exercise on an assignment sheet. The proof for general $n$ is more difficult and will involve some new ideas.

Although derivatives are only normally defined for functions over $\mathbb{R}$ or $\mathbb{C}$, for an arbitrary field $F$ and $f = a_n x^n + \cdots + a_1 x + a_0 \in F[x]$, we can define the formal derivative $Df$ of $f$ to be

\[na_n x^{n-1} + (n - 1)a_{n-1} x^{n-2} + \cdots + 2a_2 x + a_1.\]

It can be checked that formal derivatives have many of the familiar properties of derivatives, including:

(i) $D(f+g) = Df + Dg$
(ii) $D(af) = aD(f)$
(iii) $D(fg) = fDg + gDf$

for all $f, g \in F[x]$ and $a \in F$.

((i) and (ii) are clear from the definition. To prove (iii), do it first for $f = x^n$, $g = x^m$ and then use (i), (ii) and $f(g + h) = fg + fh$ for $f, g, h \in F[x]$.)

In particular, note that if $f = e^2 g$ with $e, f, g \in F[x]$, then $Df = 2geDe + e^2Dg$ is divisible by $e$.

Let $p$ be a prime not dividing $n$. For $a \in \mathbb{Z}$, we denote its residue modulo $p$ in $\mathbb{Z}_p$ by $\bar{a}$. Then the map $\mathbb{Z}[x] \to \mathbb{Z}_p[x]$ that maps $a_n x^n + \cdots + a_1 x + a_0$ to $\bar{a}_n x^n + \cdots + \bar{a}_1 x + \bar{a}_0$ is a ring homomorphism (Example 2 in Subsection 7.1), and we denote the image of a polynomial $f \in \mathbb{Z}[x]$ under this map by $\bar{f}$.

Lemma 9.14 If $p$ is a prime not dividing $n$, and $\bar{\cdot}$ is as just defined, then $\gcd(x^n - 1, D(x^n - 1)) = 1$.

Proof: We have $D(x^n - 1) = \pi x^{n-1}$ and so $xD(x^n - 1) - \pi x^{n-1} = \pi$. But $p \not| n$ means that $\pi$ is a unit in $\mathbb{Z}_p$, so we can multiply by $\pi^{-1}$, and the result follows from Proposition 8.6.

Corollary 9.15 We cannot have $e^2(x^n - 1)$ with $e \in \mathbb{Z}_p[x]$ and $\deg(e) > 0$. 

48
Lemma 9.16 (i) \( a^p = a \) for all \( a \in \mathbb{Z}_p \); (ii) \( f(x^p) = f(x)^p \) for all \( f \in \mathbb{Z}[x] \).

Proof: (i) Since the multiplicative group \( \mathbb{Z}_p \setminus \{1\} \) has order \( p-1 \), we have \( a^{p-1} = 1 \) for all \( a \in \mathbb{Z}_p \setminus \{1\} \), and hence \( a^p = a \). But this is also true for \( a = 0 \), so it is true for all \( a \in \mathbb{Z}_p \).

(ii) Let \( f = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{Z}_p[x] \). Recall from Example 4 of Subsection 7.1 that, for a ring of characteristic \( p \) (that is, a ring in which \( px = 0 \) for all \( x \in R \)), the map \( x \mapsto x^p \) is a ring homorphism of \( R \). This applies to \( \mathbb{Z}_p[x] \), so \( f(x^p) = a_n^p x^{np} + \cdots + a_1^p x^p + a_0^p \), which by (i) is equal to \( f(x^p) \).

We shall also need the following easy lemma.

Lemma 9.17 Let \( F \) and \( K \) be fields with \( F \subseteq K \), and let \( p, q \) be coprime polynomials in \( F[x] \). Then \( p \) and \( q \) are also coprime as elements of \( K[x] \).

Proof: This follows from Proposition 8.6, since \( p, q \) coprime in \( F[x] \) implies that there exist \( r, s \in F[x] \) with \( pr + qs = 1 \). But then, since \( r, s \in K[x] \), \( p, q \) are coprime in \( K[x] \).

We turn now to the proof of Theorem 9.13. Suppose for a contradiction that \( \Phi(x) \) is reducible for some \( n > 0 \), so we can write \( \Phi(x) = f(x) g(x) \) with \( f, g \in \mathbb{Z}[x] \) and \( \deg(f), \deg(g) > 0 \), where we may assume that \( f \) is irreducible and \( f(\zeta_n) = 0 \). Since \( \Phi \) is monic, we may assume that \( f, g \) are monic.

Proposition 9.18 Let \( p \) be a prime not dividing \( n \). Then \( f(\zeta_n^p) = 0 \).

Proof: Suppose not. Since \( p \) does not divide \( n \), \( \zeta_n^p \) is a primitive \( n \)-th root of 1 and hence \( \Phi_n(\zeta_n^p) = 0 \). So \( f(\zeta_n^p) \neq 0 \) implies \( g(\zeta_n^p) = 0 \), and hence \( \zeta_n \) is a root of the polynomial \( h(x) := g(x^p) \). Since \( f \) and \( h \) have a common root \( \zeta_n \), they cannot be coprime in \( \mathbb{C}[x] \), and so, by the lemma above, they are not coprime in \( \mathbb{Q}[x] \). But since \( f \) is irreducible in \( \mathbb{Q}[x] \), we must have \( \gcd(f, h) = f \), and hence \( f | h \). Also, since \( f, h \) are monic, we have \( f | h \) in \( \mathbb{Z}[x] \).

So with \( n \) defined as above, we have, in \( \mathbb{Z}_p[x] \), \( \Phi_n = f \overline{g} \), and \( f \overline{h} \), where \( \overline{h}(x) = g(x^p) = g(x)^p \) by Lemma 9.16. So we cannot have \( \gcd(\overline{f}, \overline{g}) = 1 \), since this would imply that \( \gcd(\overline{f}, \overline{h}) = 1 \), contradicting \( \overline{f} | \overline{h} \). So \( \Phi_n = \overline{f} \overline{g} \) is divisible by \( \overline{f}^2 \), where \( \overline{e} = \gcd(\overline{f}, \overline{g}) = 1 \), with \( \deg(e) > 1 \). But then \( \overline{f}^2 | x^n - 1 \), contradicting Corollary 9.15.

Corollary 9.19 Let \( 0 \leq i < m \) with \( \gcd(i, n) = 1 \). Then \( f(\zeta_n^i) = 0 \).

Proof: We can write \( i = p_1 p_2 \ldots p_r \), where the \( p_j \) are not necessarily distinct primes not dividing \( n \). By Proposition 9.18, \( f(\zeta_n^{p_1 p_2}) = 0 \). But we can apply it again with \( \zeta_n^{p_1} \) in place of \( \zeta_n \) and conclude that \( f(\zeta_n^{p_1 p_2}) = 0 \). So an inductive proof shows that \( f(\zeta_n^i) = 0 \).

We can now immediately complete the proof of Theorem 9.13. By this last corollary, \( f \) has at least \( \phi(n) \) distinct roots in \( \mathbb{C}[x] \), and so \( \deg(f) \geq \phi(n) = \deg(\Phi_n) \), which contradicts our assumption that \( \Phi_n = gh \) with \( \deg(g), \deg(h) > 0 \).