We write $\mathbb{N}$ for the set of positive integers, $\mathbb{Z}$ for the set of integers, $\mathbb{Q}$ for the set of rational numbers, $\mathbb{R}$ for the set of real numbers and $\mathbb{C}$ for the set of complex numbers.

1.1 The Algebra of Matrices

1.1.1. Definition. By a field $F$, we mean a non-empty set of elements with two laws of combination, which we call an addition and a multiplication satisfying:

\begin{enumerate}
\item[(+1)] To every pair of elements $a, b \in F$ there is associated a unique element, called their sum, which we denote by $a + b$.
\item[(+2)] Addition is associative; $(a + b) + c = a + (b + c)$.
\item[(+3)] Addition is commutative; $a + b = b + a$.
\item[(+4)] There exists an element, which we denote by 0, such that $a + 0 = a$ for all $a \in F$.
\item[(+5)] For each $a \in F$ there exists an element, which we denote by $-a$ such that $a + (-a) = 0$.
\item[(\times 1)] To every pair of elements $a, b \in F$ there is associated a unique element, called their product, which we denote by $ab$, or $a \cdot b$.
\item[(\times 2)] Multiplication is associative; $(ab)c = a(bc)$.
\item[(\times 3)] Multiplication is commutative; $ab = ba$.
\item[(\times 4)] There exists an element different from 0, which we denote by 1, such that $a \cdot 1 = a$ for all $a \in F$.
\item[(\times 5)] For each $a \in F$, $a \neq 0$, there exists an element which we denote by $a^{-1}$, such that $a \cdot a^{-1} = 1$.
\item[(\times +)] Multiplication is distributive with respect to addition: $(a + b)c = ac + bc$.
\end{enumerate}

1.1.2. Example. The set of numbers $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ under ordinary addition and multiplication are fields. However, the set of integers $\mathbb{Z}$ satisfies every item in the definition except $(\times 5)$, so it is not a field.

1.1.3. Definition. Let $F$ be a field. An $m \times n$ (m by n) matrix $A$ with $m$ rows and $n$ columns with entries over $F$ is a rectangular array of the form

\[
A = \begin{bmatrix}
a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\
\vdots & \ddots & \vdots & & \vdots \\
a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\
\vdots & \ddots & \vdots & & \vdots \\
a_{m1} & \cdots & a_{mj} & \cdots & a_{mn}
\end{bmatrix},
\]

where $a_{ij} \in F$ for all $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$. We write $M_{m,n}(F)$ for the set of $m \times n$ matrices with entries in $F$ and we write $M_n(F)$ for $M_{n,n}(F)$ the set of square matrices of order $n$. We call a $1 \times n$ matrix, a row vector and we call an $m \times 1$ matrix, a column vector.
1.1.4. Remark. As a shortcut, we often use the notation $A = [a_{ij}]$ to denote the matrix $A$ with entries $a_{ij}$. Notice that when we refer to the matrix we put parentheses—as in “$[a_{ij}]$”, and when we refer to a specific entry we do not use the surrounding parentheses—as in “$a_{ij}$”.

1.1.5. Definition. The $m \times n$ zero matrix $0_{m \times n} \in M_{m,n}(F)$ is the matrix with 0’s everywhere,

$$0_{m \times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$ 

When $m = n$, we write $0_n$ as an abbreviation for $0_{n \times n}$.

1.1.6. Definition. Two $m \times n$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are equal if $a_{ij} = b_{ij}$ for all $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$.

1.1.7. Definition. The $n \times n$ identity matrix $I_n \in M_n(F)$ is the matrix with 1’s on the main diagonal and 0’s everywhere else,

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$ 

1.1.8. Definition. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ matrices and a scalar $r \in F$. The matrix $A + rB$ is the matrix $C \in M_{m,n}(F)$ with entries $C = [c_{ij}]$ where

$$c_{ij} = a_{ij} + rb_{ij}.$$ 

1.1.9. Proposition. Let $A$, $B$ and $C$ be matrices of the same size, and let $r$ and $s$ be scalars in $F$. Then

1. $A + B = B + A$ 
2. $(A + B) + C = A + (B + C)$ 
3. $A + 0 = A$ 
4. $r(A + B) = rA + rB$ 
5. $1A = A$ 
6. $(r + s)A = rA + sA$ 
7. $r(sA) = (rs)A$

1.1.10. Definition. Let $A$ be an $m \times n$ matrix with columns $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$ and $\vec{x}$ is a column vector in $F^n$. The product of $A$ and $\vec{x}$ denoted by $A\vec{x}$ is the linear combination of the columns of $A$ using the corresponding entries in $\vec{x}$ as weights. That is,

$$A\vec{x} = \begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} := x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n.$$ 

If $B$ is an $n \times p$ matrix with columns $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_p$, then the product of $A$ and $B$, denoted by $AB$, is the $m \times p$ matrix with columns $A\vec{b}_1, A\vec{b}_2, \ldots, A\vec{b}_p$. In other words,

$$AB = A \begin{bmatrix} \vec{b}_1 \\ \vec{b}_2 \\ \vdots \\ \vec{b}_p \end{bmatrix} := \begin{bmatrix} A\vec{b}_1 \\ A\vec{b}_2 \\ \vdots \\ A\vec{b}_p \end{bmatrix}.$$
The above definition of $AB$ is a good for theoretical work. When $A$ and $B$ have small sizes, the following method is more efficient when working by hand. Let $A = [a_{ij}] \in M_{m,n}(F)$ and $B = [b_{ij}] \in M_{n,p}(F)$. Then the matrix product $AB$ is defined as the matrix $C = [c_{ij}] \in M_{m,p}(F)$ with entries

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj},$$

that is,

$$
\begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}
\begin{bmatrix}
  b_{11} & b_{12} & \cdots & b_{1p} \\
  \vdots & \vdots & \ddots & \vdots \\
  b_{n1} & b_{n2} & \cdots & b_{np}
\end{bmatrix}
= 
\begin{bmatrix}
  c_{11} & \cdots & c_{1p} \\
  \vdots & \ddots & \vdots \\
  c_{m1} & \cdots & c_{mp}
\end{bmatrix}.
$$

1.1.11. Proposition. Let $A$ be $m \times n$ and let $B$ and $C$ have sizes for which the indicated sums and products are defined.

1. $r(AB) = (rA)B = A(rB)$ for any scalar $r$
2. $A(B + C) = AB + AC$ and $(B + C)A = BA + CA$
3. $A(BC) = (AB)C$
4. $A0_{n \times k} = 0_{m \times k}$ and $0_{k \times m} A = 0_{k \times n}$
5. $I_mA = A = AI_n$

1.1.12. Definition. If $A$ is a square matrix of order $n$, then we write $A^k$ for $A \cdot \cdots \cdot A$ ($k$ copies).

1.1.13. Remarks. Properties above are analogous to properties of real numbers. But NOT ALL real number properties correspond to matrix properties.

1. It is not the case that $AB$ always equal to $BA$.
2. Even if $AB = AC$, then $B$ may not equal to $C$.
3. It is possible for $AB = 0$ even if $A \neq 0$ and $B \neq 0$.

1.1.14. Definition. Let $A$ be an $n \times n$ matrix. We say that $A$ is invertible or nonsingular and has the $n \times n$ matrix $B$ as inverse if $AB = BA = I_n$.

If $B$ and $C$ are $n \times n$ matrices with $AB = I_n$ and $CA = I_n$, then the associativity of multiplication implies that

$$B = I_n B = (CA)B = C(AB) = CI_n = C.$$

Hence, an inverse for $A$ is unique if it exists and we write $A^{-1}$ for this inverse.

1.1.15. Theorem. Suppose $A$ and $B$ are invertible matrices of the same size. Then the following results hold.

1. $A^{-1}$ is invertible and $(A^{-1})^{-1} = A$, i.e., $A$ is the inverse of $A^{-1}$.
2. $AB$ is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

1.1.16. Corollary. Let $GL_n(F)$ be the set of all invertible $n \times n$ matrices over $F$. Then:

1. $I_n \in GL_n(F)$.
2. $A, B \in GL_n(F) \Rightarrow AB \in GL_n(F)$.
3. $A \in GL_n(F) \Rightarrow A^{-1} \in GL_n(F)$. 
1.1. Exercises.  

1. Let $A$ be $m \times n$, and let column vectors $\vec{z}$, $\vec{y}$ and $\vec{z}$ and matrices $B$ and $C$ have sizes for which the indicated sums and products are defined. Prove that:
   
   (a) $A(\vec{z} + \vec{y}) = A\vec{z} + A\vec{y}$ and $(B + C)\vec{z} = B\vec{z} + C\vec{z}$
   
   (b) $A(B + C) = AB + AC$ and $(B + C)A = BA + CA$
   
   (c) $A(B\vec{z}) = (AB)\vec{z}$
   
   (d) $A(BC) = (AB)C$

2. For any two subsets of $U$, we define

   $$A + B = (A \setminus B) \cup (B \setminus A) \quad \text{and} \quad A \cdot B = A \cap B.$$  

   Note that $A + B = (A \cup B) \setminus (A \cap B) = (A \cup B) \cap (A \cap B)^c = (A \cup B) \cap (A^c \cup B^c)$.

   Prove that for subsets $A$, $B$ and $C$ of $U$:
   
   (a) $A + B = B + A$
   
   (b) $A + \emptyset = A = \emptyset + A$
   
   (c) $A + A = \emptyset$ and $A \cdot A = A$
   
   (d) $(A + B) + C = (A \cup B \cup C) \cap (A \cup B^c \cup C^c) \cap (A^c \cup B \cup C^c) \cap (A^c \cup B^c \cup C)$
   
   (e) $A + (B + C) = (A + B) + C$
   
   (f) $A + B = A + C \Rightarrow B = C$
   
   (g) $A \cdot (B + C) = A \cdot B + A \cdot C$

3. Let $\mathbb{C} = \{a + ib : a, b \in \mathbb{R} \text{ and } i^2 = -1\}$ be the set of complex numbers with ordinary addition and multiplication. Note that $\mathbb{R} = \{a + i0 : a \in \mathbb{R}\} \subset \mathbb{C}$. Recall that the conjugate of $z = a + ib$ is given by $\overline{z} = a - ib$ and $|z| = \sqrt{a^2 + b^2}$. It is easy to see that

   $$\overline{\overline{z}} = z,$$  

   $$\overline{z + w} = \overline{z} + \overline{w},$$  

   and $$\overline{zw} = \overline{z}\overline{w}.$$  

   Moreover, if $z = a + ib$, then

   $$z\overline{z} = (a + ib)(a - ib) = a^2 + b^2 = |z|^2,$$

   so

   $$\frac{1}{z} = \frac{\overline{z}}{|z|^2} = \frac{a}{a^2 + b^2} - i \left( \frac{b}{a^2 + b^2} \right) \iff z \neq 0.$$  

   Define $\mathbb{Q}[i] = \{a + ib : a, b \in \mathbb{Q}\}$ and $\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\}$.

   (a) Prove that $|zw| = |z||w|$.

   (b) Prove that if $z \in \mathbb{Q}[i]$, then $z \neq 0 \iff \frac{1}{z} \in \mathbb{Q}[i]$.

   (c) Prove that $\mathbb{Q}[i]$ under ordinary addition and multiplication satisfies all items in 1.1.1.

   (d) Prove that $\mathbb{Z}[i]$ under ordinary addition and multiplication satisfies every item in 1.1.1 except $(\times b)$. Elements in $\mathbb{Z}[i]$ are called Gaussian integers.

   (e) Find all $z \in \mathbb{Z}[i]$ such that $\frac{1}{z} \in \mathbb{Z}[i]$.

4. Let $H$ be a nonempty subset of $\mathbb{C}$ such that $a - b \in H$ for all $a, b \in H$.

   Prove that $0 \in H$ and $(a \in H \Rightarrow -a \in H)$.

1.2 Integers

1.2.1. Theorem. [Division Algorithm] Given integers $a$ and $b$, with $b \neq 0$, there exist unique integers $q$ and $r$ satisfying

   $$a = qb + r,$$  

   where $0 \leq r < |b|$.

The integers $q$ and $r$ are called, respectively, the quotient and remainder in the division of $a$ by $b$.

Proof. To prove this theorem, we must use the well-ordering principle, namely, "every nonempty set $S$ of nonnegative integers contains a least element; that is, there is some integer $a$ in $S$ such that $a \leq b$ for all $b \in S$".
1.2. Integers

Existence: First we shall assume that \( b > 0 \). Let \( S = \{ a - xb : x \in \mathbb{Z} \text{ and } a - xb \geq 0 \} \subseteq \mathbb{N} \cup \{0\} \). We shall show that \( S \neq \emptyset \). Since \( b \geq 1 \), we have \( |a|b \geq |a| \), so

\[
a - (-|a|)b = a + |a|b \geq a + |a| \geq 0,
\]

Then \( a - (\neg|a|)b \in S \), so \( S \neq \emptyset \). By the well-ordering principle, \( S \) contains a least element, call it \( r \). Then \( a - qb = r \) for some \( q \in \mathbb{Z} \). Since \( r \in S, r \geq 0 \) and \( a = qb + r \). It remains to show that \( r < b \). Suppose that \( r \geq b \). Thus,

\[
0 \leq r - b = a - qb - b = a - (q + 1)b,
\]

so \( r - b \leq r \) and \( r - b \in S \). This contradicts the minimality of \( r \). Hence, \( r < b \).

Next, we consider the case in which \( b < 0 \). Then \( |b| > 0 \) and the previous argument gives \( q', r \in \mathbb{Z} \) such that

\[
a = q'|b| + r, \quad \text{where} \quad 0 \leq r < |b|.
\]

Since \( |b| = -b \), we may take \( q = -q' \) to arrive at

\[
a = qb + r, \quad \text{where} \quad 0 \leq r < |b|
\]

as desired.

Uniqueness: Let \( q, q', r, r' \in \mathbb{Z} \) be such that

\[
a = qb + r \quad \text{and} \quad a = q'b + r',
\]

where \( 0 \leq r, r' < |b| \). Then

\[
(q - q')b = r' - r.
\]

Since \( 0 \leq r, r' < |b| \), we have \( |r' - r| < |b| \), so \( |b||q - q'| = |r' - r| < |b| \). This implies that \( 0 \leq |q - q'| < 1 \), hence \( q = q' \) which also forces \( r = r' \).

\[
1.2.2. \text{Definition.} \quad \text{An integer } b \text{ is said to be divisible by an integer } a \neq 0, \text{ in symbols } a \mid b, \text{ if there exists some integer } c \text{ such that } b = ac. \text{ We write } a \nmid b \text{ to indicate that } b \text{ is not divisible by } a.
\]

Note that \( a \mid b \Leftrightarrow -a \mid b \), so we may consider only positive divisors. The next theorem contains elementary properties of divisibility.

\[
1.2.3. \text{Theorem.} \quad \text{For integers } a, b \text{ and } c, \text{ the following statements hold:}
\]

1. \( a \mid 0, 1 \mid a, a \mid a \).
2. \( a \mid 1 \) if and only if \( a = \pm 1 \).
3. If \( a \mid b \), then \( a \mid (b), (a) \mid b \) and \( (a) \mid (b) \).
4. If \( a \mid b \) and \( c \mid d \), then \( ac \mid bd \).
5. If \( a \mid b \) and \( b \mid c \), then \( a \mid c \).
6. \( (a \mid b \) and \( b \mid a) \) if and only if \( a = \pm b \).
7. If \( a \mid b \) and \( b \neq 0 \), then \( |a| \leq |b| \).
8. If \( a \mid b \) and \( a \mid c \), then \( a \mid (bx + cy) \) for arbitrary integers \( x \) and \( y \).

\[
1.2.4. \text{Definition.} \quad \text{An integer } p > 1 \text{ is called a prime number, or simply a prime, if its only}
\]

positive divisors are 1 and \( p \). An integer greater than 1 which is not a prime is termed composite.

\[
1.2.5. \text{Example.} \quad 2, 3, 5, 11, 2011 \text{ are primes. } 6, 8, 12, 2558 \text{ are composite numbers.}
\]
1.2.6. **Definition.** Let \(a\) and \(b\) be given integers, with at least one of them different from zero. The **greatest common divisor (gcd)** of \(a\) and \(b\), denoted by \(\gcd(a, b)\), is the positive integer \(d\) satisfying:

1. \(d \mid a\) and \(d \mid b\),
2. for all \(c \in \mathbb{Z}\), if \(c \mid a\) and \(c \mid b\), then \(c \leq d\).

Basic properties of gcd are collected in the next theorem.

1.2.7. **Theorem.** Let \(a\) and \(n\) be integers not both zero.

1. If \(d = \min\{ax + ny : x, y \in \mathbb{Z}\}\), then \(d = \gcd(a, n)\).
2. If \(\gcd(a, n) = d\), then \(\exists x, y \in \mathbb{Z}, ax + ny = d\).
3. \(\gcd(a, n) = 1\) if and only if \(\exists x, y \in \mathbb{Z}, ax + ny = 1\).

**Proof.** (1) The given set contains \(a^2 + n^2\), so it is not empty and \(d\) exists by the well-ordering principle. Then \(d = ax + ny > 0\) for some \(x, y \in \mathbb{Z}\). We shall prove that \(d = \gcd(a, n)\). By the division algorithm, \(\exists r, q \in \mathbb{Z}, a = dq + r\) with \(0 \leq r < d\). If \(r > 0\), then

\[
0 < r = a - dq = a - (ax + ny)q = a(1 - xq) - nyq < d
\]

which contradicts the minimality of \(d\). Hence, \(r = 0\) and \(d \mid n\). Similarly, \(d \mid n\). Since \(d = ax + ny\), \(\gcd(a, n) \mid d\), so \(\gcd(a, n) \leq d\). But \(d \mid a\) and \(d \mid n\), so \(d \leq \gcd(a, n)\). Hence, \(d = \gcd(a, n)\). (2) follows from (1) and (3) follows from (2). The converse of (3) is immediate. \(\square\)

1.2.8. **Corollary.** Let \(a, b\) and \(c\) be integers and \(p\) a prime.

1. If \(a \mid bc\) and \(\gcd(a, b) = 1\), then \(a \mid c\).
2. \(p \nmid b \iff \gcd(p, b) = 1\)
3. If \(p \mid bc\) and \(p \nmid b\), then \(p \mid c\). More generally, if \(a_1, a_2, \ldots, a_k\) are integers such that \(p \mid a_1 a_2 \cdots a_k\), then \(p \mid a_i\) for some \(i\).
4. If \(q_1, q_2, \ldots, q_n\) are all primes and \(p \mid q_1 q_2 \cdots q_k\), then \(p = q_i\) for some \(i\).

**Proof.** Since \(\gcd(a, b) = 1\), we have \(1 = ax + by\) for some \(x, y \in \mathbb{Z}\). Then \(c = acx + bcy\). Since \(a \mid bc\), \(a \mid c\). This proves (1). Note that for a prime \(p\), its divisors are 1 and \(p\). Thus, \(\gcd(p, b)\) could be 1 or \(p\). If \(\gcd(p, b) = p\), then \(p \mid b\). If \(\gcd(p, b) = 1\), then \(p \nmid b\). (4) follows from (3) and (3) follows from (1) and (2). \(\square\)

In my opinion, the next theorem is the most important and everyday use result in number theory. Its proof applies the results discussed above.

1.2.9. **Theorem.** **[Fundamental Theorem of Arithmetic]** Every positive integer \(m > 1\) can be expressed as a product of primes; this representation is unique, apart from the order in which the factors occur.

**Proof.** **Expressible:** Assume on the contrary that there exists an integer \(m > 1\) which is not a product of primes. By the well-ordering principle, there is a smallest \(n_0\) such that \(n_0\) is not a product of primes. Then \(n_0\) is composite, so there exist integers \(1 < d_1, d_2 < n_0\) such that \(n_0 = d_1 d_2\). Since \(d_1, d_2 < n_0\), \(d_1\) and \(d_2\) are products of primes, and so is \(n_0\). This gives a contradiction. Hence, every positive integer \(m > 1\) can be expressed as a product of primes.

**Uniqueness:** Assume that

\[
m = p_1 p_2 \cdots p_s = q_1 q_2 \cdots q_t,
\]

where \(1 \leq s \leq t\) and \(p_i\) and \(q_j\) are prime such that

\[
p_1 \leq p_2 \leq \cdots \leq p_s \quad \text{and} \quad q_1 \leq q_2 \leq \cdots \leq q_t.
\]
Corollary 1.2.8 (3) tells us that \( p_1 = q_k \) for some \( k \in \{1, \ldots, t\} \). It makes \( p_1 \geq q_1 \). Similarly, \( q_1 = p_l \) for some \( l \in \{1, \ldots, s\} \). Then \( q_1 \geq p_1 \), so \( p_1 = q_1 \). Thus,
\[
p_2 \ldots p_s = q_2 \ldots q_t.
\]
Now, repeat the process to get \( p_2 = q_2 \), and we obtain
\[
p_3 \ldots p_s = q_3 \ldots q_t.
\]
Continue in this manner. If \( s < t \), we would get
\[
1 = q_{s+1}q_{s+2} \ldots q_t,
\]
which is impossible. Hence, \( s = t \) and
\[
p_1 = q_1, p_2 = q_2, \ldots, p_s = q_s
\]
as desired.

1.2.10. Corollary. Any positive integer \( m > 1 \) can be written uniquely in a canonical form
\[
m = p_1^{k_1}p_2^{k_2} \ldots p_r^{k_r},
\]
where, for \( i = 1, 2, \ldots, r \), each \( k_i \) is a positive integer and each \( p_i \) is a prime, with \( p_1 < p_2 < \ldots < p_r \).

To formulate an important example in group theory, we shall discuss about the set of integers modulo a positive integer.

1.2.11. Definition. Let \( n \) be a fixed positive integer. Two integers \( a \) and \( b \) are said to be congruent modulo \( n \), symbolized by
\[
a \equiv b \pmod{n} \quad \text{or} \quad a \equiv b \pmod{n}
\]
if \( n \) divides the difference \( a - b \); that is, provided that \( a - b = kn \) for some integer \( k \). The number \( m \) is called the modulus of the congruence. When \( n \nmid (a - b) \), then we say that \( a \) is incongruent to \( b \) modulo \( n \) and in this case we write \( a \not\equiv b \pmod{n} \).

1.2.12. Remark. If \( n \mid a \), we may write \( a \equiv 0 \pmod{n} \).

The first theorem is immediate.

1.2.13. Theorem. If \( a \equiv b \pmod{n} \) and \( c \equiv d \pmod{n} \), then we have:
1. \( ax + cy \equiv bx + dy \pmod{n} \) for all integers \( x \) and \( y \),
2. \( ac \equiv bd \pmod{n} \),
3. \( a^m \equiv b^m \pmod{n} \) for every positive integer \( m \).

For \( a \in \mathbb{Z} \), it follows from the division algorithm that there exist unique \( q, r \in \mathbb{Z} \) such that \( a = nq + r \), where \( 0 \leq r < n \). This implies \( a \equiv r \pmod{n} \). Thus, we have shown:

1.2.14. Theorem. For each integer \( a \), there exists a unique integer \( r \), with \( 0 \leq r < n \), such that \( a \equiv r \pmod{n} \).

In terms of congruence, Theorem 1.2.7 (3) may be restated as follows.
1.2.15. **Corollary.** [Inverse Modulo $n$] Let $a$ and $n$ be integers with $n$ positive. Then $\gcd(a, n) = 1$ if and only if there exists an integer $x$ such that $ax \equiv 1 \pmod{n}$. We call $x$ the inverse of $a$ modulo $n$.

We directly obtain the following corollary from Corollary 1.2.8. It gives a condition for canceling integers modulo $n$.

1.2.16. **Corollary.** [Cancellative of Integers Modulo $n$] Let $a, b, c$ and $n$ be integers with $n$ positive and let $p$ be a prime.

1. If $ac \equiv bc \pmod{n}$ and $\gcd(n, c) = 1$, then $a \equiv b \pmod{n}$.
2. If $ac \equiv bc \pmod{p}$ and $p \nmid c$, then $a \equiv b \pmod{p}$.

Finally, we shall define the set of integers modulo a positive integer $n$. It is not difficult to see that the congruence modulo $n$ is an equivalence relation on $\mathbb{Z}$. That is, for $a, b, c \in \mathbb{Z}$, we have:

1. [Reflexivity] $a \equiv a \pmod{n}$,
2. [Symmetry] $a \equiv b \pmod{n}$ implies $b \equiv a \pmod{n}$,
3. [Transitivity] $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$ imply $a \equiv c \pmod{n}$.

1.2.17. **Definition.** For $a \in \mathbb{Z}$, the equivalence class of $a$ is given by

$$\bar{a} = \{b \in \mathbb{Z} : b \equiv a \pmod{n}\} = \{kn + a : k \in \mathbb{Z}\}$$

It also follows from Theorem 1.2.14 that the set of all equivalence classes is

$$\{[0], [1], \ldots, [n-1]\}.$$  

This set is called the set of integers modulo $n$ and is denoted by $\mathbb{Z}_n$. We define two binary operations on $\mathbb{Z}_n$ by

$$[a]_n + [b]_n = [a + b]_n \quad \text{and} \quad [a]_n \times [b]_n = [a \times b]_n \quad \text{for all} \ a, b \in \mathbb{Z}.$$  

By Theorem 1.2.13, they are well-defined. They also satisfy every item in Definition 1.1.1 except $(\times 5)$, where $[0]_n$ and $[1]_n$ are the identities for $+$ and $\times$, respectively. Furthermore, if $n = p$ is a prime, then $(\times 5)$ holds by Corollary 1.2.8 (2) and Corollary 1.2.15.

1.2. **Exercises.**

1. (a) If $a \in \mathbb{Z}$, prove that $a^2$ is of the form $4n$ or $4n + 1$ for some $n$.

(b) Show that no integer $u = 4n + 3$ can be written as $u = a^2 + b^2$, where $a, b \in \mathbb{Z}$.

2. If $p$ is an odd prime, show that $p$ is of the form:

(a) $4n + 1$ or $4n + 3$ for some $n$

(b) $6n + 1$ or $6n + 5$ for some $n$

3. Let $d = \gcd(a, b)$. Prove that:

(a) $\gcd(a/d, b/d) = 1$

(b) $\gcd(a - bq, b) = d$ for all $q \in \mathbb{Z}$.

4. **[Euclidean Algorithm]** Let $a$ and $b$ be positive integers, with $b \leq a$. Repeatedly applications of the division algorithm to $a$ and $b$ give

$$a = bq_1 + r_1,$$

$$b = r_1q_2 + r_2,$$

$$r_1 = q_2r_2 + r_3,$$

$$\vdots$$

$$r_{n-2} = q_{n-1}r_{n-1} + r_n,$$

$$r_{n-1} = q_nr_n.$$  

Prove that $r_n = \gcd(a, b)$. *(Hint. Use 3. (b))*
5. Let \( a, b \in \mathbb{Z} \) and \( n \in \mathbb{N} \). If \( \gcd(a, n) = 1 \), show that there exists an \( x \in \mathbb{Z} \) such that \( ax \equiv b \pmod{n} \).

6. Show that \( \gcd(a, n) = 1 \) and find the inverse of \( a \) modulo \( n \).

   For: (a) \( n = 2559 \) and \( a = 50 \)  
   (b) \( n = 2016 \) and \( a = 21 \).

7. Solve: (a) \( 15x \equiv 4 \pmod{17} \)  
   (b) \( 21x \equiv 12 \pmod{2016} \).

8. The least common multiple (lcm) of two nonzero integers \( a \) and \( b \), denoted by \( \text{lcm}(a, b) \), is the positive integer \( m \) satisfying: (1) \( a \mid m \) and \( b \mid m \),  
   (2) if \( a \mid c \) and \( b \mid c \), with \( c > 0 \), then \( m \leq c \).

   Prove that \( \text{lcm}(a, b) = \frac{ab}{\gcd(a, b)} \) for all \( a, b \in \mathbb{N} \).

9. Let \( a \) and \( b \) be two integers greater than 1 factored as

\[
a = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r} \quad \text{and} \quad b = p_1^{b_1} p_2^{b_2} \cdots p_r^{b_r},
\]

where for \( i = 1, 2, \ldots, r \), each \( p_i \) is a prime with \( p_1 < p_2 < \cdots < p_r \), each \( a_i \) and \( b_i \) are nonnegative integers, and each \( a_i \) or \( b_i \) are positive. Prove that

\[
\gcd(a, b) = p_1^{d_1} p_2^{d_2} \cdots p_r^{d_r}, \quad \text{where} \quad d_i = \min\{a_i, b_i\} \quad \text{for all} \quad i = 1, 2, \ldots, r
\]

and

\[
\text{lcm}(a, b) = p_1^{c_1} p_2^{c_2} \cdots p_r^{c_r}, \quad \text{where} \quad c_i = \max\{a_i, b_i\} \quad \text{for all} \quad i = 1, 2, \ldots, r.
\]

10. For \( d \in \mathbb{N} \), we define \( d\mathbb{Z} = \{dx : x \in \mathbb{Z}\} \subseteq \mathbb{Z} \). Let \( d \in \mathbb{N} \). Prove that:

   (a) \( 0 \in d\mathbb{Z} \) and \( d \in d\mathbb{Z} \).
   (b) If \( a, b \in d\mathbb{Z} \), then \( a + b \in d\mathbb{Z} \).
   (c) If \( a \in d\mathbb{Z} \), then \( -a \in d\mathbb{Z} \).
   (d) If \( m \in \mathbb{Z} \) and \( a \in d\mathbb{Z} \), then \( ma \in d\mathbb{Z} \).

11. Let \( n \in \mathbb{N} \). Define the relation \( r \) on \( \mathbb{Z} \) by

\[
ar \cdot b \iff a - b \in n\mathbb{Z} \quad \text{for all} \quad a, b \in \mathbb{Z}.
\]

   Prove that \( r \) is an equivalence relation on \( \mathbb{Z} \). Find all equivalence classes with respect to \( r \).

12. Let \( m, n \in \mathbb{N} \) with \( m \neq n \). Define \( m\mathbb{Z} + n\mathbb{Z} = \{mx + ny : x, y \in \mathbb{Z}\} \). Prove that:

   (a) \( m\mathbb{Z} + n\mathbb{Z} = d\mathbb{Z} \), where \( d = \gcd(m, n) \).
   (b) \( m\mathbb{Z} \cap n\mathbb{Z} = c\mathbb{Z} \), where \( c = \text{lcm}(m, n) \).

---

### 1.3 Groups

In order to study abstract senses of algebra, we shall begin with the definition of a group which occupies a very important seat in this course.

#### 1.3.1 Definitions and Examples

**1.3.1. Definition.** For a nonempty set \( S \), a function \( \cdot : S \times S \rightarrow S \) is called a **binary operation** and image of \((a, b)\) in \( S \times S \) is denoted by \( a \cdot b \) and it is said to be the **product** or **composition** of \( a \) and \( b \). A **groupoid** is a system \((S, \cdot)\) consisting of a nonempty set \( S \) with binary operation \( \cdot \) on \( S \).

We may write \( S \) for \((S, \cdot)\) and \( ab \) for \( a \cdot b \) where \( a, b \in S \) if there is no ambiguity. Let \( S \) be a groupoid. For nonempty subsets \( A \) and \( B \) of \( S \) and \( x \in S \), let

\[
AB = \{ab : a \in A \text{ and } b \in B\}, \quad xA = \{x\}A \quad \text{and} \quad Ax = A\{x\}.
\]

**1.3.2. Definition.** If \( S \) satisfies the **associative law**, i.e., \( \forall a, b, c \in S, (a \cdot b) \cdot c = a \cdot (b \cdot c) \), we say that \( S \) is a **semigroup**.

Notice that if \( S \) is a semigroup, then any bracketing of \( x_1, \ldots, x_n \) gives the same product, so we can write \( x_1 \cdots x_n \) for this product. In addition, for \( a \in S \) and \( m \in \mathbb{N} \), we may let \( a^m = a \cdots a \) (\( m \) copies).
1.3.3. **Definition.** A groupoid $S$ is said to be **commutative** if $\forall a, b \in S, ab = ba$.

1.3.4. **Definition.** An element $e$ of a groupoid $S$ is a **two-sided identity** or **identity** if $\forall a \in S, ae = a = ea$.

Clearly, $S$ contains at most one identity (if $e$ and $e'$ are identity, then $e = ee' = e$).

1.3.5. **Definition.** A **monoid** is a semigroup with (unique) identity. Let $S$ be a monoid with identity $e$. If $a$ and $b$ in $S$ are such that $ab = e = ba$, then $b$ is called a **two-sided inverse** or **inverse** of $a$.

We have that every element of $S$ has at most one inverse. For, if $b$ and $b'$ are inverses of $a$, then $ab = e = ba$ and $a'b = e = b'a$, so $b = be = b(ab') = (ba)b' = eb' = b'$.

1.3.6. **Definition.** A **group** is a monoid $G$ such that every element of $G$ has an inverse, and for $a \in G$, let $a^{-1}$ denote the (unique) inverse of $a$.

1.3.7. **Remark.** For a nonempty set $G$ with binary operation on $G$ is a **group** if the following axioms are all satisfied:

(G1) [associativity] $\forall a, b, c \in G, (a \cdot b) \cdot c = a \cdot (b \cdot c)$

(G2) [identity] $\exists e \in G, \forall a \in S, ae = a = ea$

(G3) [inverse] $\forall a \in G, \exists b \in G, ab = e = ba$.

1.3.8. **Definition.** A group $G$ is **commutative** if $\forall a, b \in G, ab = ba$. A commutative group is also called an **abelian group**.

1.3.9. **Definition.** The **order** of a group $G$ is the cardinal number $|G|$.

Let $G$ be a group with identity $e$. For $a \in G$ and $m \in \mathbb{N}$, let $a^0 = e$ and $a^{-m} = (a^{-1})^m$.

1.3.10. **Remarks.**

1. For a group $G$, we have:
   (a) $e^{-1} = e$ and $\forall a \in G, (a^{-1})^{-1} = a$,
   (b) $\forall a \in G, \forall m, n \in \mathbb{Z}, a^m a^n = a^{m+n}$ and $(a^m)^n = a^{mn}$, and
   (c) $\forall a, b \in G, (ab)^{-1} = b^{-1} a^{-1}$ because $(ab)(b^{-1} a^{-1}) = e$.

2. In case $G$ is abelian, we may choose to write $G$ additively. This means:
   (a) The binary operation is denoted by $+$.
   (b) $0$ denotes the identity element and $-a$ denotes the inverse of $a$.
   (c) $\forall a \in G, \forall m \in \mathbb{N}, ma = a + \cdots + a$ ($m$ copies).

3. A group $G$ satisfies the **cancellation law**: $\forall a, b, c \in G, ab = ac$ (or $ba = ca$) $\Rightarrow b = c$.

1.3.11. **Examples.**

1. $(\mathbb{Z}, -)$ is a groupoid which is not a semigroup; $(\mathbb{N}, +)$ is a semigroup which is not a monoid; $(\mathbb{N}, \cdot)$ is a monoid which is not a group.

2. $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$, $(\mathbb{Q}^*, \cdot)$, $(\mathbb{R}^*, \cdot)$ and $(\mathbb{C}^*, \cdot)$ are infinite abelian groups. Here, $A^*$ denotes the set of nonzero elements of $A$.

3. Let $X$ be a set and $P(X)$ the power set of $X$. For subsets $A$ and $B$ of $X$, we define $A \triangle B = (A \setminus B) \cup (B \setminus A)$. Then $(P(X), \triangle)$ is an abelian group having the empty set as its identity and $A^{-1} = A$ for all $A \in P(X)$. Also, $(P(X), \cap)$ is a commutative monoid with identity $X$.

4. Write $F$ for any of $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ or other fields. Let $M_n(F)$ be the set of $n \times n$ matrices over $F$ and $\text{GL}_n(F)$ the set of matrices over $F$ with nonzero determinants. Then $(M_n(F), +)$ is an abelian group and $\text{GL}_n(F)$ is a group under multiplication which is not abelian if $n > 1$. The later group is called the **general linear group**.
5. For a nonempty set $X$, a function on $X$ which is 1-1 and onto (a bijection on $X$) is said to be a permutation of $X$. Let $S(X)$ be the set of all permutations of $X$. Then under composition, $(S(X), \circ)$ is a group called the symmetric group on $X$; in case $X = \{1, 2, \ldots, n\}$, we write $S_n$ and call $S_n$ the symmetric group on $n$ letters. It is a group of order $n!$.

6. For $n \in \mathbb{N}$, let $\mathbb{Z}_n = \{\overline{0}, \overline{1}, \ldots, \overline{n-1}\}$ called the set of integers modulo $n$, where $\overline{a} = \{kn + a : k \in \mathbb{Z}\}$ for all $a \in \mathbb{Z}$. Define $+$ and $\cdot$ on $\mathbb{Z}_n$ by

$$\overline{a} + \overline{b} = \overline{a + b} \quad \text{and} \quad \overline{a} \cdot \overline{b} = \overline{a \cdot b} \quad \text{for all} \ a, b \in \mathbb{Z}.$$

Then $(\mathbb{Z}_n, +)$ is an abelian group of order $n$. Moreover, $(\mathbb{Z}_n, \cdot)$ is a commutative monoid with identity $\overline{1}$.

7. For $n \in \mathbb{N}$ and $n \geq 2$, let $\mathbb{Z}^\times_n = \{\overline{a} : \gcd(a, n) = 1\}$. Then $(\mathbb{Z}^\times_n, \cdot)$ is an abelian group of order $\phi(n)$, the Euler $\phi$-function. Note that $\mathbb{Z}^\times_n = \mathbb{Z}_n \setminus \{\overline{0}\} \iff n$ is a prime.

**Proof.** If $n$ is a prime, then $\mathbb{Z}^\times_n = \{\overline{1}, \overline{2}, \ldots, \overline{n-1}\} = \mathbb{Z}_n \setminus \{\overline{0}\}$. Conversely, assume that $n$ is composite. Then $n = bc$ for some $1 < b, c < n$, so $\gcd(b, n)$ and $\gcd(c, n)$ are $> 1$. Thus, $\overline{b}, \overline{c} \notin \mathbb{Z}^\times_n$. Since $b, c < n$, we have $\overline{b}, \overline{c} \neq \overline{0}$. Hence, $\mathbb{Z}^\times_n \not\subseteq \mathbb{Z}_n \setminus \{\overline{0}\}$. \hfill $\Box$

**1.3.12. Remark.** We recall some properties of the Euler’s $\phi$-function as follows.

1. If $p$ is a prime, then $\phi(p) = p - 1$ and $\phi(p^k) = p^k - p^{k-1}$ for all $k \in \mathbb{N}$.
2. $\phi$ is multiplicative, namely, if $\gcd(m, n) = 1$, then $\phi(mn) = \phi(m)\phi(n)$.

Some equivalent definitions of groups are collected in the next theorem.

**1.3.13. Theorem.** Let $G$ be a semigroup. Then the following statements are equivalent.

(i) $G$ is a group.
(ii) (a) $\exists e \in G \forall a \in G, ea = a$ and (b) $\forall a \in G \exists b \in G, ba = e$.
(iii) (a) $\exists e \in G \forall a \in G, ae = a$ and (b) $\forall a \in G \exists b \in G, ab = e$.
(iv) $\forall a, b \in G \exists x, y \in G, ax = b$ and $ya = b$.
(v) $\forall a \in G, aG = G = Ga$.

**Proof.** If (i) holds, (ii)–(v) are clearly true. (iv) $\iff$ (v) is obvious.
(ii) $\Rightarrow$ (i). Assume (ii). Let $a \in G$. Then $\exists b \in G, ba = e$, and so $\exists c \in G, cb = e$. Thus,

$$ab = e(ab) = (eb)(ab) = c(ba)b = c(eb) = cb = e.$$

Moreover, $ae = a(ba) = (ab)a = ea = a$.

“(iii) $\Rightarrow$ (i)” is similar to “(ii) $\Rightarrow$ (i)”.
(iv) $\Rightarrow$ (iii). Assume (iv). Let $a \in G$. Then $\exists e \in G, ae = a$. Let $b \in G$. Then $\exists c \in G, bc = e$ and $\exists y \in G, ya = b$. Thus, $be = (ya)e = ye(ae) = ya = b$. \hfill $\Box$

**1.3.14. Theorem.** If $G$ is a finite cancellative semigroup, then $G$ is a group.

**Proof.** We shall show that $\forall a \in G, aG = G = Ga$. Let $a \in G$. Since $G$ is cancellative, $|aG| = |G| = |Ga|$. Clearly, $aG \subseteq G$ and $Ga \subseteq G$. Since $G$ is finite, $aG = G = Ga$. \hfill $\Box$

**1.3.15. Remark.** $(\mathbb{N}, +)$ is an infinite cancellative semigroup, but it is not a group.
1.3.2 Subgroups

Sometimes a group contains a nonempty subset that is closed under its operation. In this subsection, we discuss a small group in a bigger one with the same operation.

1.3.16. Definition. A nonempty subset $H$ of a group $G$ is said to be a subgroup of $G$ if $H$ is a group under the same operation of $G$ and we write $H \leq G$. Observe that for $\emptyset \neq H \subseteq G$,

$$H \leq G \iff \forall a, b \in H, ab \in H \land a^{-1} \in H \iff \forall a, b \in H, ab^{-1} \in H.$$ 

Moreover, $\{e\}$ and $G$ are always subgroups of $G$. Theorem 1.3.14 gives the following important fact.

1.3.17. Corollary. If $H$ is a finite nonempty subset of a group $G$ which is closed under the operation of $G$, then $H$ is a subgroup of $G$.

Proof. Since $G$ is a cancellative semigroup and $\emptyset \neq H \subseteq G$, $H$ is a finite cancellative semigroup. Hence, $H$ is a group by Theorem 1.3.14. \qed

Next, we investigate the group of symmetries. We begin with the following groups.

1.3.18. Examples. (Group of symmetries)

1. The set of rotations about a point 0 in the plane; composition as usual. If 0 is taken to be the origin, the rotation through an angle $\theta$ can be represented analytically as the map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

For $\theta = 0$, we get the identity map and the inverse of the rotation through the angle $\theta$ is the rotation through $-\theta$. It is called the rotation group.

2. The set of rotations together with the set of reflections in the lines which passes through 0 with slope $\tan \alpha$. The latter are given analytically by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \cos 2\alpha + y \sin 2\alpha \\ x \sin 2\alpha - y \cos 2\alpha \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix}.$$

The product of two reflections is a rotation and the product in either order of a reflection and a rotation is a reflection.

3. Consider the regular $n$-gon (that is, the $n$-sided polygon in which the sides are all the same length and are symmetrically placed about a common center) inscribed in the unit circle in the plane, so that one of the vertices is $(1,0)$. The vertices subtend angles of $0, 2\pi/n, 4\pi/n, \ldots, 2(n-1)\pi/n$ radians with the positive $x$-axis. The subset of rotation maps which maps our figure into itself consists of the $n$ rotations through angles of $0, 2\pi/n, 4\pi/n, \ldots, 2(n-1)\pi/n$ radians, respectively. These elements form a subgroup $R_n$ of the rotation group defined in (1).

4. We now consider the set $D_n$ of rotations and reflections which map the regular $n$-gon, as in (3), into itself. These form a subgroup of the group defined in (2). We shall call the elements of this group the symmetries of the regular $n$-gon. The reflection in the $x$-axis is one of our symmetries. Multiplying this on the right by the $n$ rotational symmetries we obtain $n$ distinct reflectional symmetries. These give them all, for if we let $\sigma$ denote the reflection in the $x$-axis and $\tau$ denote any reflectional symmetry then $\sigma \tau$ is one of the $n$-rotational symmetries $\rho_1, \ldots, \rho_n$, say $\rho_i$. Since $\sigma^2 = 1$, $\sigma \tau = \rho_i$ gives $\tau = \sigma \rho_i$, which is one of those we counted. Thus, $D_n$ consists of $n$ rotations and $n$ reflections and its order is $2n$. The group $D_n$ is called the dihedral group. Note that $D_n = \{\sigma^i \rho_j : i \in \{0,1\} \text{ and } j \in \{0,1,2,\ldots,n-1\}\}$. 

1.3.19. Remark. It is easy to see that the intersection of a family of subgroups of a group $G$ is a subgroup of $G$. If $H$ and $K$ are subgroups of a group $G$, then, in general, $H \cup K$ is not a subgroup of $G$. However, if $H$ and $K$ are subgroups of a group $G$ with $G = H \cup K$, then $H = G$ or $K = G$.

Proof. Assume that there is an $x \in G \setminus H$ and a $y \in G \setminus K$. Since $G = H \cup K$, we have $x \in K$ and $y \in H$. Thus, $xy \notin H$ and $xy \notin K$, a contradiction. 

1.3.20. Definition. Let $G$ be a group and $A$ a subset of $G$. Define $\langle A \rangle$ to be the intersection of all subgroups of $G$ containing $A$. It is the smallest subgroup of $G$ containing $A$ and is called the subgroup of $G$ generated by $A$. The elements of $A$ are called generators. Moreover, we have $\langle \emptyset \rangle = \{e\}$ and

$$\langle A \rangle = \{a_1^{n_1} \cdots a_k^{n_k} : a_i \in A \text{ and } n_i \in \mathbb{Z} \} \quad \text{if } A \neq \emptyset.$$ 

For $a_1, \ldots, a_m \in G$, we write $\langle a_1, \ldots, a_m \rangle$ for $\langle \{a_1, \ldots, a_m\} \rangle$. Then $\forall a \in G, \langle a \rangle = \{a^n : n \in \mathbb{Z}\} = \langle a^{-1} \rangle$ is called the cyclic subgroup of $G$ generated by $a$ and the order of $a$ is $|\langle a \rangle|$ (finite or infinite) and denoted by $|a|$ or $o(a)$. If $G = \langle a \rangle$ for some $a \in G$, then $G$ is said to be the cyclic group generated by $a$.

1.3.21. Definition. A subgroup $N$ of $G$ is normal if $\forall g \in G \forall x \in N, gxg^{-1} \in N$. We write $N \trianglelefteq G$ for a normal subgroup $N$ of $G$.

1.3.22. Examples. (Examples of subgroups and normal subgroups) 1. $\{e\}$ and $G$ are normal subgroups of $G$. They are called the trivial normal subgroups.
2. Every subgroup of an abelian group is normal.
3. Let $SL_n(F)$ be the set of matrices over $F$ with determinant one. Then $SL_n(F)$ is a normal subgroup of $GL_n(F)$ because $\det(PQP^{-1}) = \det Q$ for all $P, Q \in GL_n(F)$.
4. $R_n = \langle \rho_{2\pi/n} \rangle$ and $R_n \trianglelefteq D_n$ because $\sigma \rho_j \sigma = \rho_j^{-1}$ for all $j$.

1.3.23. Remarks. 1. $N \trianglelefteq G$ if and only if $(N \trianglelefteq G \land \forall g \in G, gNg^{-1} = N)$.
2. If $N$ is a subgroup of a group $G$, then $(\forall g \in G, gNg^{-1} = N) \Leftrightarrow (\forall g \in G, gN = Ng)$.

1.3.24. Definition. Let $G$ be a group and $X$ a nonempty subset of $G$. The centralizer of $X$ is the set

$$C_G(X) = \{g \in G : \forall x \in X, gx = xg\}$$

and the normalizer of $X$ is the set

$$N_G(X) = \{g \in G : gX = Xg\}.$$ 

We call $Z(G) = C_G(G) = \{z \in G : \forall z \in G, zx = xz\}$, the center of $G$.

1.3.25. Remarks. 1. The centralizer and normalizer of $X$ are subgroups of $G$.

Proof. Since $\forall x \in G, ex = xe$, we have $e \in C_G(X)$. Let $g, h \in C_G(X)$. Let $x \in X$. Then $gx = xg$ and $hx = xh$, so $xh^{-1} = h^{-1}x$ and $$(gh^{-1})x = g(h^{-1}x) = g(xh^{-1}) = (gx)h^{-1} = (xg)h^{-1} = x(gh^{-1}).$$

Thus, $gh^{-1} \in C_G(X)$. We leave $N_G(X)$ as an exercise. 

2. $Z(G) \trianglelefteq G$ and $Z(G) = \bigcap_{x \in G} C_G(\{x\})$.

Proof. By (1), $Z(G)$ is a subgroup of $G$. To see it is normal, let $g \in G$ and $z \in Z(G)$. Let $x \in G$. Then $zg = gz$ and $xz = zx$, so

$$(gzg^{-1})z = (zg^{-1})z = z = x(zg^{-1}) = x(gzg^{-1}).$$
Hence, \(gzg^{-1} \in Z(G)\). Finally, for \(z \in G\),
\[
z \in Z(G) \iff \forall x \in G, xz = zx \iff \forall x \in G, z \in C_G(x) \iff z \in \bigcap_{x \in G} C_G(x).
\]
This proves the second statement. \(\square\)

3. \(G\) is abelian \(\iff Z(G) = G\). (This is clear.)
4. If \(K \leq G\), then \(K \leq N_G(K)\) and \(N_G(K)\) is the largest subgroup of \(G\) containing \(K\) in which \(K\) is normal (this means \(\forall H \leq G, K \leq H \Rightarrow H \leq N_G(K)\)).

**Proof.** If \(g \in K\), then \(gK = K\). Thus, \(N_G(K)\) is normal in \(G\). Let \(x \in N_G(K)\) and \(g \in K\). Since \(Kx = xK\), we have \(xg \in Kx\), so \(xg = kg\) for some \(k \in K\). Thus \(xgk^{-1} = k \in K\). Next, let \(H\) be a subgroup of \(G\) such that \(K\) is normal in \(H\). Then \(\forall h \in H, Kh = hK\) which implies \(H \leq N_G(K)\). \(\square\)

### 1.3.3 Homomorphisms

We now study a function between two groups that is required to preserve group operations.

#### 1.3.26. Definition.** Let \(G\) and \(H\) be two groups. A homomorphism from \(G\) to \(H\) is a map \(\varphi : G \to H\) which satisfies
\[
\varphi(xy) = \varphi(x)\varphi(y) \quad \text{for all } x, y \in G.
\]
An injective homomorphism is called a monomorphism and a surjective homomorphism is called an epimorphism. An isomorphism is a homomorphism which is both injective and surjective. We write \(G \cong H\) if \(\exists \varphi : G \to H, \varphi\) is an isomorphism. An endomorphism of \(G\) is a homomorphism on \(G\) and an automorphism of \(G\) is an isomorphism of \(G\) onto itself.

#### 1.3.27. Examples. (Examples of group homomorphisms)
1. \(\varphi_1 : \mathbb{Z} \to \mathbb{Z}_n\) given by \(\varphi(a) = \bar{a}\) is a homomorphism.
2. \(\varphi_2 : (\mathbb{R}, +) \to (\mathbb{R}^+, \cdot)\) given by \(\varphi(x) = 2^x\) is a homomorphism.
3. \(\varphi_3 : \text{GL}_n(\mathbb{R}) \to \mathbb{R}\) given by \(\varphi(A) = \det A\) is a homomorphism.

#### 1.3.28. Definition.** The kernel of a homomorphism \(\varphi : G \to H\) is given by the set
\[
\ker \varphi = \{g \in G : \varphi(g) = e_H\}.
\]

#### 1.3.29. Example.** \(\ker \varphi_1 = n\mathbb{Z}\), \(\ker \varphi_2 = \{0\}\) and \(\ker \varphi_3 = \text{SL}_n(\mathbb{R})\).

#### 1.3.30. Remarks.** Let \(\varphi : G \to H\) be a homomorphism of groups.
1. \(\varphi(e_G) = e_H\) and \(\varphi(a^{-1}) = (\varphi(a))^{-1}\) for all \(a \in G\).

**Proof.** Note that \(e_H\varphi(e_G) = \varphi(e_G) = \varphi(e_G)\varphi(e_G) = \varphi(e_G)\varphi(e_G)\). By cancellation in \(H\), we have \(e_H = \varphi(e_G)\). Next, let \(a \in G\). Then \(\varphi(a)\varphi(a^{-1}) = \varphi(aa^{-1}) = \varphi(e_G) = e_H\). Hence, \(\varphi(a^{-1}) = \varphi(a)^{-1}\). \(\square\)

2. \(\varphi\) is 1-1 \(\iff \ker \varphi = \{e_G\}\).

**Proof.** Assume that \(\varphi\) is 1-1. By (1), \(e_G \in \ker \varphi\). Let \(x \in \ker \varphi\). Then \(\varphi(x) = e_H = \varphi(e_G)\). Since \(\varphi\) is 1-1, we have \(x = e_G\). Conversely, suppose that \(\ker \varphi = \{e_G\}\). Let \(x, y \in G\) be such that \(\varphi(x) = \varphi(y)\). Then \(e_H = \varphi(x)^{-1}\varphi(y) = \varphi(x^{-1})\varphi(y) = \varphi(x^{-1}y)\). Thus, \(x^{-1}y \in \ker \varphi\), so \(x^{-1}y = e_G\). Hence, \(x = y\). \(\square\)

#### 1.3.31. Theorem.** Let \(\varphi\) be a homomorphism of a group \(G\) into a group \(H\).
1. \(\ker \varphi\) is a normal subgroup of \(G\).
2. \(\text{im} \varphi = \{\varphi(g) : g \in G\} = \varphi(G)\) is a subgroup of \(H\).
1.3. Groups

**Proof.** Since \( \varphi(e_G) = e_H, \ e_G \in \ker \varphi \) and \( e_H \in \text{im} \varphi \). Let \( x, y \in \ker \varphi \). Then \( \varphi(x) = e_H = \varphi(y) \), so

\[
\varphi(xy^{-1}) = \varphi(x)\varphi(y^{-1}) = \varphi(x)\varphi(y)^{-1} = e_H e_H^{-1} = e_H.
\]

Thus, \( xy^{-1} \in \ker \varphi \). Hence, \( \ker \varphi \) is a subgroup of \( G \). Next, let \( g \in G \) and \( x \in \ker \varphi \). Then \( \varphi(x) = e_H \) and

\[
\varphi(gxg^{-1}) = \varphi(g)\varphi(x)\varphi(g^{-1}) = \varphi(g)\varphi(x)\varphi(g)^{-1} = \varphi(g)e_H\varphi(g)^{-1} = e_H.
\]

Thus, \( gxg^{-1} \in \ker \varphi \), and so \( \ker \varphi \) is normal. Finally, let \( y, z \in \text{im} \varphi \). Then \( \exists x_1, x_2 \in G, \varphi(x_1) = y \) and \( \varphi(x_2) = z \). Thus,

\[
yz^{-1} = \varphi(x_1)\varphi(x_2)^{-1} = \varphi(x_1)\varphi(x_2^{-1}) = \varphi(x_1 x_2^{-1}).
\]

Since \( x_1 x_2^{-1} \in G \), \( yz^{-1} \in \text{im} \varphi \). Hence, \( \text{im} \varphi \) is a subgroup of \( H \).

\( \square \)

1.3.32. **Remark.** If \( \varphi : G \to H \) is a homomorphism and \( H_1 \leq H \), then \( \ker \varphi \subseteq \varphi^{-1}(H_1) \leq G \).

The next result is clear. It gives another way to construct a group for the Cartesian product of groups.

1.3.33. **Theorem.** [Cartesian Product of Groups] If \( G_1 \) and \( G_2 \) are groups, then the Cartesian product

\[
G_1 \times G_2 = \{ (x, y) : x \in G_1, y \in G_2 \}
\]

is a group under coordinatewise multiplication \( (x_1, y_1)(x_2, y_2) = (x_1 x_2, y_1 y_2) \) for all \( x_1, x_2 \in G_1 \) and \( y_1, y_2 \in G_2 \).

1.3. **Exercises.**

1. Let \( G \) be the set of pairs of real numbers \( (a, b) \) with \( a \neq 0 \) and define:

\[
(a, b)(c, d) = (ac, ad + b) \quad \text{and} \quad 1 = (1, 0).
\]

Verify that this defines a group.

2. Let \( H = \{ \sigma \in S_4 : \{ \sigma(1), \sigma(2) \} = \{1, 2\} \text{ or } \{ \sigma(1), \sigma(2) \} = \{3, 4\} \} \). Prove that \( H \) is a subgroup of \( S_4 \) and find \( |H| \). Is \( H \) normal in \( S_4 \)? Justify your answer.

3. Let \( G \) be a semigroup such that \( \forall a \in G, \exists b \in G, a = aba \) and \( \exists! e \in G, e^2 = e \). Prove that \( G \) is a group.

4. Let \( G \) be a semigroup such that \( \forall a, b \in G, a^2 b = b = ba^2 \). Prove that \( G \) is an abelian group.

5. A certain multiplicative operation on a nonempty set \( G \) is associative and allows cancellations on the left, and there exists \( a \in G \) such that \( x^3 = axa \) for all \( x \in G \). Prove that \( G \) endowed with this operation is an abelian group.

6. Let \( G \) be a group with the following properties:

(i) \( G \) has no element of order 2 and \( (xy)^2 = (yx)^2 \), for all \( x, y \in G \).

Prove that \( G \) is abelian. If (i) fails, give an example to support that “\( G \) may not be abelian”.

7. If \( H \) and \( K \) are subgroups of a group \( G \), prove that \( HK \leq G \) if and only if \( HK = KH \).

8. Let \( \varphi : G \to \tilde{G} \) be a group homomorphism, and let \( N \) and \( \tilde{N} \) be a normal subgroup of \( G \) and \( \tilde{G} \), respectively. Show that \( \varphi[N] \) is a normal subgroup of \( \text{im} \varphi \) and \( \varphi^{-1}[\tilde{N}] \) is a normal subgroup of \( G \).

9. Let \( G \) be a group with identity \( e \) and \( \varphi : G \to G \) a function such that

\[
\varphi(g_1)\varphi(g_2)\varphi(g_3) = \varphi(h_1)\varphi(h_2)\varphi(h_3)
\]

whenever \( g_1g_2g_3 = e = h_1h_2h_3 \). Prove that there exists an element \( a \in G \) such that \( \psi(x) = a\varphi(x) \) for all \( x \in G \) is a homomorphism.

10. Let \( D_n \) be the dihedral group of order \( 2n \) where \( n > 2 \). Show that the center of \( D_n \) has one or two elements according as \( n \) is odd or even.
1. Project. (Quaternions) Consider the eight objects $±1, ±i, ±j$ and $±k$ with multiplication rules:

\[
ij = k \quad jk = i \quad ki = j \\
ji = -k \quad kj = -i \quad ik = -j \\
i^2 = j^2 = k^2 = -1
\]

where the minus signs behave as expected and $1$ and $-1$ multiply as expected. (For example, $(-1)j = -j$ and $(-i)(-j) = ij = k$.) Show that these objects form a group containing only one element of order $2$. This group is called the **quaternion group** and is denoted by $\mathbb{H}$. [Hint. The difficulty is to show that the operation is associative. One may transform the elements and operation into $2 \times 2$ matrices and matrix products, respectively.]

2. Project. (Associativity) One of the required properties for $(G, *)$ to be a group is associativity. However, this is the hardest one to check as one can see from the previous item. Consider the set $\{a, b, c\}$ of three distinct elements and the operation $*$ given by

\[
\begin{array}{ccc}
* & a & b & c \\
 a & a & b & c \\
b & b & a & c \\
c & c & c & a \\
\end{array}
\]

To check associativity, we must check every possible instance of the equation $(x * y) * z = x * (y * z)$. That means we must think of every possible combination of what $x, y,$ and $z$ could be. After a while, we find that $(b * c) * c = a$ but $b * (c * c) = b$. Hence, the set $\{a, b, c\}$ under the operation $*$ is not associative. Be careful!

Many students make the mistake of concluding that a set is associative by checking just a few examples. We cannot do this! To determine whether or not a set is associative, you must check every single combination of $3$ elements, unless you have a good general argument for why all combinations will be associative or you have a good reason (such as the existence of an identity element) for limiting the number of cases you must check. (Notice that even when you have the existence of an identity element, you still have to check ALL the cases which do not include the identity element.)

Now, for the set of two elements $\{a, b\}$ the number of different binary operations on this set is $16$. However, the number of associative binary operation on that set is only $8$. This can be checked by writing all out and counting. Unfortunately, for the set of three elements $\{a, b, c\}$, there are $3^{(3 \cdot 3)} = 19683$ binary operations. How to know how many associative binary operations there are on a set of three elements? (There are 113 operations.) Create an efficient algorithm to solve this task. How about the set of $n$ elements?

1.4 Group Actions

Let $G$ be a group. For each $g \in G$, define the left multiplication function $\ell_g : G \to G$ by

\[\ell_g(x) = gx \quad \text{for all } x \in G.\]

By the left cancellation on $G$, $\ell_g$ is a bijection and so $\ell_g \in S(G)$, the symmetric group on $G$. It is easy to see that the map $g \mapsto \ell_g$ is an injective group homomorphism from $G$ into $S(G)$. This proves an important result in group theory.

1.4.1. Theorem. [Cayley] Every group $G$ is isomorphic to a subgroup of $S(X)$ for some set $X$. 

Allowing an abstract group to behave like a group of permutations, as happened in the proof of Cayley’s theorem, is a useful tool. In this section, we talk about how a group acting on a set, called a group action. There are many nice results and applications as we shall see in the following sections.
1.4.2. Definition. Let $G$ be a group with identity element $e$ and $X$ a nonempty set. We say that $G$ acts on $X$ or $X$ is a $G$-set if there is a mapping $G \times X \rightarrow X$ (denoted by $g \cdot x$ or $gx$) which satisfies:
1. $\forall x \in X, e \cdot x = x$ and
2. $\forall g, h \in G, \forall x \in X, g \cdot (h \cdot x) = (gh) \cdot x$.

Assume that a group $G$ act on a set $X$. Then each $g \in G$ determines a set map $\phi_g : X \rightarrow X$ by $\phi_g(x) = gx$.

Moreover, $\forall g \in G, \phi_g$ is a bijection (1-1 and onto). Hence, $\phi_g \in S(X)$, the symmetric group on $X$. The map $g \mapsto \phi_g$ defines a group homomorphism from $G$ to $S(X)$ (i.e., $\phi_{gh} = \phi_g \circ \phi_h$ for all $g, h \in G$). Its kernel is the set $\{g \in G : gx = x \text{ for all } x \in X\}$.

1.4.3. Definition. 1. If $g = e$ is the only element of $G$ such that $gx = x$ for all $x \in X$, then $G$ is said to act faithfully on $X$. In this case, $G \hookrightarrow S(X)$.
2. If $x \in X$, the set $Gx = \{gx : g \in G\}$ is called the orbit of $x$.
3. If $Gx = X$ for some (and hence all) $x \in X$, then $G$ is said to act transitively on $X$.
4. If $Y \subseteq X$, the set $\{g \in G : gx = Y\}$ is called the stabilizer of $Y$, denoted by $\text{Stab}_G Y$.

The stabilizer of $Y$ is a subgroup of $G$.

Proof. Clearly, $e \in \text{Stab}_G Y$. Let $g, h \in \text{Stab}_G Y$. Then $gY = Y$ and $hY = Y$. Thus, $(gh)Y = g(hY) = gY = Y$ and $g^{-1}Y = g^{-1}(gY) = eY = Y$.

1.4.4. Examples. (Examples of group actions) 1. If $X$ is a set, $S(X)$ acts naturally on $X$ by $f \cdot x = f(x)$ for all $f \in S(X)$ and $x \in X$. This action is faithfully if $|X| > 1$. In particular, $S_n$ acts on $\{1, 2, \ldots, n\}$. The orbit of each $i \in \{1, 2, \ldots, n\}$ is all of $\{1, 2, \ldots, n\}$, thus $S_n$ acts transitively on $\{1, 2, \ldots, n\}$. If $Y \subseteq \{1, 2, \ldots, n\}$, the stabilizer of $Y$ is isomorphic to $S(Y) \times S(Z) \cong S_k \times S_{n-k}$ where $Z = \{1, 2, \ldots, n\} \setminus Y$. Hence, the stabilizer of $\{n\}$ is isomorphic to $S_{n-1}$.
2. Let $G$ be any group and let $X = G$, considered as a set. Let $G$ act on $X$ by left multiplication $g \cdot x = gx$.

This action is called the left regular representation. It is faithful and transitive.
3. $GL_n(F)$ acts faithfully on $F^n$, the set of $n \times 1$ column vectors by left multiplication. The orbit of $\bar{0}$ is itself and $GL_n(F)$ acts transitively on the nonzero vectors.
4. Let $G$ be any group and let $X$ be any set. Let $G$ act on $X$ by $g \cdot x = x$ for all $g \in G$ and $x \in X$. This is called the trivial $G$ action. Assuming $g \neq e$ and $X$ has more than one element, this action is not faithful and not transitive. All orbits are singleton and $G$ is the stabilizer of every subset of $X$.
5. Let $G$ be a group and let $X = G$, considered as a set. Let $G$ act on $X$ by conjugation $g \cdot x = gxg^{-1}$.

This action may not be faithful. The center of $G$ acts trivially. The orbit of $x \in G$ is the set of conjugates of $x$, that is $g \cdot X = \{gxg^{-1} : g \in G\}$, called the conjugacy class of $x$. If $|G| > 1$, then this action is not transitive. The number of orbits of the number of conjugacy classes. If $Y$ is a subset of $G$, the stabilizer of $Y$ under the action is its normalizer, i.e., $\text{Stab}_G Y = N_G(Y)$.
6. Let \( G \) be a group and let \( H \) be a subgroup of \( G \). Let \( H \) act on \( G \) by left multiplication. This action is faithful and the orbit of \( x \in G \) is
\[
H \cdot x = \{hx : h \in H\} = Hx.
\]
The action is not transitive unless \( H = G \). Moreover, we can let \( H \) act on \( G \) by \( h \cdot g = gh^{-1} \) for all \( h \in H \) and \( g \in G \). This action is also faithful and the orbit of \( x \in G \) is
\[
H \cdot x = \{xh^{-1} : h \in H\} = \{xh : h \in H\} = xH.
\]

7. Let \( X = \mathbb{C} \cup \{\infty\} \), a set that becomes the Riemann sphere in complex analysis. The group \( \text{GL}_2(\mathbb{C}) \) acts on \( X \) by the linear fractional transformation
\[
\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot z = \frac{az + b}{cz + d}
\]
the understanding being the image of \( \infty \) is \( a/c \) and the image of \( -d/c \) is \( \infty \), just as if we were to pass to a limit in each case.

8. Let \( \text{SL}_2(\mathbb{R}) \) be the subgroup of real matrices in \( \text{GL}_2(\mathbb{R}) \) of determinant one, and let \( \mathcal{H} \) be the subset of \( \mathbb{C} \cup \{\infty\} \) in which \( \text{Im} z > 0 \), called the Poincaré upper half plane. Then \( \text{SL}_2(\mathbb{R}) \) acts on \( \mathcal{H} \) by linear fractional transformations.

Now, we present another way to view the orbits.

**1.4.5. Theorem.** Let \( G \) be a group and suppose \( G \) acts on a nonempty set \( X \). Define a relation \( \sim \) on \( X \) by
\[
x \sim y \iff \exists g \in G, y = g \cdot x.
\]

Then
1. \( \sim \) is an equivalence relation on \( X \).
2. The equivalence class of \( x \in X \) under \( \sim \) is \( Gx = \{gx : g \in G\} \), the orbit of \( x \). Thus, \( X \) is a disjoint union of orbits under the action of \( G \).

**Proof.** It is routine to show that \( \sim \) is an equivalence relation on \( X \). The equivalence class of \( x \) is
\[
[x]_\sim = \{y \in X : x \sim y\} = \{y \in X : \exists g \in G, y = gx\} = \{gx : x \in X\} = Gx
\]
and hence \( X = \bigcup_{x \in X} Gx \).

**1.4.6. Definition.** If \( H \) is a subgroup of \( G \) and \( x \in G \), the set
\[
Hx = \{hx : h \in H\} \quad \text{and} \quad xH = \{xh : h \in H\}
\]
are called a left coset of \( H \) in \( G \) and a right coset of \( H \) in \( G \), respectively.

From Example 1.4.4 (6), we can let \( H \) act on \( G \) in two ways. Then \( Hx \) \([xH]\) is an orbit of \( x \), and so \( G \) is a disjoint union of left [right] cosets of \( H \) in \( G \). If we choose a subset \( \{x_\alpha\} \) of \( G \) such that \( G \) is the disjoint union of the left cosets \( Hx_\alpha \), then \( \{x_\alpha\} \) is called a right transversal or system of left coset representatives of \( H \) in \( G \) and if we choose a subset \( \{x_\alpha\} \) of \( G \) such that \( G \) is the disjoint union of the right cosets \( y_\alpha H \), then \( \{y_\alpha\} \) is called a left transversal or system of right coset representatives of \( H \) in \( G \).

**1.4.7. Remarks.**
1. By Theorem 1.4.5,
   (a) \( G = \bigcup_{x \in G} Hx \) \([G = \bigcup_{x \in G} xH]\),
   (b) \( \forall x, y \in G, Hx = Hy \) or \( Hx \cap Hy = \emptyset \) \([\forall x, y \in G, xH = yH \) or \( xH \cap yH = \emptyset\]),
(c) \( \forall x, y \in G, Hx = Hy \Leftrightarrow xy^{-1} \in H \) \( \land \forall x, y \in G, xH = yH \Leftrightarrow y^{-1}x \in H \).

2. \( \forall a \in G, \lvert H \rvert = \lvert Ha \rvert = \lvert aH \rvert \) by cancellation on \( H \).

3. The map \( aH \mapsto Ha^{-1} \) for all \( a \in G \) is a 1-1 correspondence between the sets \( \{xH : x \in G\} \) and \( \{Hx : x \in G\} \).

**Proof.** For \( a, b \in G \), \( aH = bH \Leftrightarrow b^{-1}a \in H \Leftrightarrow a^{-1}b \in H \Leftrightarrow Ha^{-1} = Hb^{-1} \). Then this map is 1-1, well defined and clearly onto. \( \square \)

**1.4.8. Definition.** The **index of \( H \) in \( G \)**, denoted by \( [G : H] \), is the cardinal number of distinct right (or left) cosets of \( H \) in \( G \), that is,

\[
[G : H] = \lvert \{Hx : x \in G\} \rvert = \lvert \{xH : x \in G\} \rvert.
\]

Next, we show that a subgroup of index two is always a normal subgroup.

**1.4.9. Theorem.** If \( H \) is a subgroup of \( G \) of index two, then \( H \) is normal in \( G \).

**Proof.** Since \( [G : H] = 2 \), \( G \) has exactly two right (or left) cosets. Then \( H \cap G = G \setminus H \) and \( gH = G \setminus H \) for all \( g \in G \) not in \( H \). Hence, \( \forall g \in G, Hg = gH \), so \( H \) is normal in \( G \). \( \square \)

Let \( I \) and \( A \) be sets. Define \( A_i = \{(i, a) : a \in A\} \) for all \( i \in I \). Then \( |A_i| = |A| \) for all \( i \in I \), \( A_i \cap A_j = \emptyset \) if \( i \neq j \), and

\[
\bigcup_{i \in I} A_i = \sum_{i \in I} |A_i| = \sum_{i \in I} |A| = |I||A|.
\]

Lagrange observed an important property of subgroups of \( G \), namely, its order must be a divisor of the order of \( G \).

**1.4.10. Theorem.** [Lagrange] If \( H \) is a subgroup of \( G \), then \( |G| = [G : H]|H| \). In particular, if \( G \) is finite and \( H < G \), then \( |H| \) divides \( |G| \), and so \( |a| = |\langle a \rangle| \) divides \( |G| \) for all \( a \in G \).

**Proof.** Since \( G \) is a disjoint union of distinct left cosets \( Hx_\alpha, \alpha \in \Lambda \), and \( |\Lambda| = [G : H] \),

\[
|G| = \bigcup_{\alpha \in \Lambda} Hx_\alpha = \sum_{\alpha \in \Lambda} |Hx_\alpha| = \sum_{\alpha \in \Lambda} |H| = |\Lambda||H| = [G : H]|H|.
\]

If \( G \) is finite, then \( |H| \) divides \( |G| \). In addition, \( \forall a \in G, \langle a \rangle \leq G \), so \( |a| = |\langle a \rangle| \) divides \( |G| \) for all \( a \in G \). \( \square \)

**1.4.11. Corollary.** If \( G \) is a group of prime order, then \( \{e\} \) and \( G \) are the only two subgroups of \( G \) and \( G \) must be cyclic.

**Proof.** Let \( H \leq G \). Then \( |H| \) divides \( |G| = p \), so \( |H| = 1 \) or \( |H| = p \). Thus, \( H = \{e\} \) and \( H = G \). Also, if \( a \neq e \), then \( \langle a \rangle \neq \langle e \rangle \). Hence, \( \langle a \rangle \) = \( G \) and so \( G \) is cyclic. \( \square \)

A relationship between the stabilizer of \( x \) in a group \( G \) and the number of elements in the orbit \( G \cdot x \) is recorded in the next theorem.

**1.4.12. Theorem.** [Orbit-Stabilizer Theorem] Let a group \( G \) act on a set \( X \) and suppose \( x \in X \). Then \( [G : Stab_G x] = |G \cdot x| \), that is, the index of the stabilizer of \( x \) in \( G \) is the number of elements in the orbit of \( x \).
Proof. Let \( x \in X \). Note that for all \( g_1, g_2 \in G \),

\[
g_1 x = g_2 x \iff (g_2^{-1} g_1) x = x \iff g_2^{-1} g_1 \in \text{Stab}_G \{x\} \iff g_1 \text{Stab}_G \{x\} = g_2 \text{Stab}_G \{x\}.
\]

Then \( |\{gx : g \in G\}| = |\{g \text{Stab}_G x : g \in G\}| \). Hence, \( |G : x| = |G : \text{Stab}_G x| \). □

This theorem is most useful when this index is finite but it is true in general. We see some applications of this theorem in the following results.

### 1.4.13. Theorem.
Let \( G \) be a group and \( x \in G \). Then the following statements hold.

1. \( |\{gxg^{-1} : g \in G\}| = [G : C_G(x)] \), i.e., the number of conjugates of \( x \) is \( [G : C_G(x)] \).
2. If \( G \) is finite, then the number of conjugates of \( x \) is a divisor of \( |G| \).

Proof. It follows directly from the Orbit-Stabilizer Theorem if we consider the action of \( G \) on \( G \) by conjugation. □

Observe that for each \( x \in G \),

\[
|\{gxg^{-1} : g \in G\}| = 1 \iff \{gxg^{-1} : g \in G\} = \{x\} \iff \forall g \in G, gx = xg \iff x \in Z(G).
\]

### 1.4.14. Corollary. [Class Equation]
Let \( G \) be a finite group and let \( x_1, \ldots, x_s \) represent the conjugacy classes of \( G \) which contains more than one element. Then

\[
|G| = |Z(G)| + \sum_{i=1}^s [G : C_G(x_i)].
\]

Now, let \( G \) act on the set of all subsets of \( G \) by conjugation, i.e., if \( Y \subset G \), then \( g \cdot Y = gYg^{-1} \). Under this action the stabilizer of \( Y \) is \( \{g \in G : gYg^{-1} = Y\} = N_G(Y) \), the normalizer of \( Y \), and the orbit of \( Y \) is \( \{gYg^{-1} : g \in G\} \), the set of conjugates of \( Y \). Thus, the number of conjugates of \( Y \) is the index of the normalizer of \( Y \). Hence, we have shown:

### 1.4.15. Theorem.
Let \( G \) be a finite group and \( Y \) a subset of \( G \). Then the number of conjugates of \( Y \) is \( [G : N_G(Y)] \) where \( N_G(Y) \) is the normalizer of \( Y \). In particular, the number of conjugates of \( Y \) divides the group order.

### 1.4.16. Remark.
If \( H \) is a subgroup of \( G \), then \( H \trianglelefteq N_G(H) < G \). Hence, if \( G \) is finite, then the number of conjugates of \( H \) is

\[
[G : N_G(H)] = \frac{|G : H|}{[N_G(H) : H]} \leq |G : H|.
\]

Burnside's theorem gives the number of orbits in \( X \) under the action of a finite group \( G \).

### 1.4.17. Theorem. [Burnside]
Let a finite group \( G \) act on a finite set \( X \). For each \( g \in G \), let \( X_g = \{x \in X : gx = x\} \), the set of points in \( X \) fixed by \( g \). Then the number of orbits in \( X \) is

\[
N = \frac{1}{|G|} \sum_{g \in G} |X_g|.
\]
Proof. Let $U$ be the subset of $G \times X$ defined by

$$U = \{(g, x) \in G \times X : gx = x\}.$$ 

For $h \in G$, let

$$U(h) = U \cap \{(h) \times X\} = \{(h, x) : x \in X \text{ and } hx = x\}.$$ 

For $y \in X$, let

$$U[y] = U \cap \{(G \times \{y\})\} = \{(g, y) : g \in G \text{ and } gy = y\} = \{\text{Stab}_G\{y\}\} \times \{y\}.$$ 

Now, $U = \bigcup_{g \in G} U(g) = \bigcup_{x \in X} U[x]$ and these unions are “disjoint”. Note that for each $g \in G$, $|U(g)| = |X_g|$ and for each $x \in X$, $|U[x]| = |\text{Stab}_G\{x\}| = |G : G \cdot x| = |G|/|G \cdot x|$. Thus,

$$\sum_{g \in G} |X_g| = \sum_{g \in G} |U(g)| = |U| = \sum_{x \in X} |U[x]| = \sum_{x \in X} \frac{|G|}{|G \cdot x|} = |G| \sum_{x \in X} \frac{1}{|G \cdot x|} = |G|N$$

as desired. \hfill \Box

1.4.18. Corollary. Let a finite group $G$ act transitively on a finite set $X$. Then $|G| = \sum_{g \in G} |X_g|$. Moreover, if $|X| > 1$, then there exists a $g \in G$ fixing no point of $X$.

Proof. Since $G$ acts transitively on $X$, $N = 1$, and so $|G| = \sum_{g \in G} |X_g|$. Assume that $|X| > 1$ and no $g \in G$ fixing no point of $X$. Then $\forall g \in G, \exists x \in X, gx = x$ which implies that $|X_g| \geq 1$ for all $g \in G$. Thus,

$$|G| \leq \sum_{g \in G} |X_g| = |G|.$$ 

This forces that $|X_g| = 1$ for all $g \in G$. But $|X_e| = |X| > 1$, a contradiction. Hence, there exists a $g \in G$ fixing no point of $X$. \hfill \Box

We have known from Lagrange's theorem that the order of any subgroup of a group $G$ is a divisor of $|G|$. The next theorem implies that if $|G|$ has a prime divisor $p$, then $G$ has a subgroup of order $p$. Its proof is another application of group actions.

1.4.19. Theorem. [Cauchy] Suppose $G$ is a finite group and a prime $p$ divides $|G|$. Then the number of solutions of $g^p = e$ in $G$ is a multiple of $p$. Hence, $G$ contains an element of order $p$. In particular, if $G$ is a finite group and a prime $p$ divides $|G|$, then $G$ has a subgroup of order $p$.

Proof. Consider the set $Y = G \times G \times \cdots \times G$ ($p$ copies) and let $X = \{(g_1, g_2, \ldots, g_p) \in Y : g_1g_2 \cdots g_p = e\}$. Then $|X| = |G|^{p-1}$ since $g_p = (g_1g_2 \cdots g_{p-1})^{-1}$. Let $R_p = \langle \rho_{2n/p} \rangle$ act on $X$ by

$$\rho_{2n/p}(g_1, g_2, \ldots, g_p) = (g_2, g_3, \ldots, g_p, g_1).$$

Note that the orbit of $(g_1, g_2, \ldots, g_p) \in X$ under this action has either

(a) length $p$, in case $g_1, g_2, \ldots, g_p$ are not all equal, or

(b) length 1, in case $g_1 = g_2 = \cdots = g_p = g$, i.e., $g^p = e$.

Thus, # of orbits of length 1 = # of solutions of $g^p = e$ in $G$. By Theorem 1.4.5 (2),

$$|G|^{p-1} = |X| = p(\# \text{ of orbits of length } p) + 1(\# \text{ of orbits of length } 1).$$

Since $p$ divides $|G|$, it follows that $|\{g \in G : g^p = e\}| > 1$ is a multiple of $p$. \hfill \Box
1.4. Exercises. 1. Let $G$ act on $S$, $H$ act on $T$ and assume $S \cap T = \emptyset$. Let $U = S \cup T$ and define $(g, h)s = gs$ and $(g, h)t = ht$ for all $g \in G$, $h \in H$, $s \in S$, $t \in T$. Show that this gives an action of the group $G \times H$ on $U$.

2. Let $H$ and $K$ be subgroups of a group $G$.

(a) If $H$ and $K$ are finite, then $HK$ is a finite set and $|HK| = \frac{|H||K|}{|H \cap K|}$.

(b) For $x$ and $y$ in $G$, prove that $xH \cap yK$ is empty or is a coset of $H \cap K$.

(c) Deduce from (b) that if $H$ and $K$ have finite index in $G$, then so does $H \cap K$.

(d) If $[G : H]$ and $[G : K]$ are finite and relatively prime, prove that $G = HK$.

3. Let $\alpha$ be an automorphism of a finite group $G$ which leaves only the identity fixed. Prove that $G = \{x^{-1}\alpha(x) | x \in G\}$.

4. Let a group $G$ act on a set $X$ transitively. Prove that

(a) $\forall x, y \in X, \exists g \in G, gx = y$, and

(b) $\forall x, y \in X, \exists g \in G, gGxg^{-1} = Gy$, i.e., all stabilizers are conjugate.

5. Let $H$ be a subgroup of a group $G$ and $N = \bigcap_{x \in G} xHx^{-1}$. Prove that

(a) $N$ is a normal subgroup of $G$, and

(b) if $[G : H]$ is finite, then $[G : N]$ is finite.

6. Determine the number of conjugacy classes in a non-abelian group $G$ of order $p^3$ where $p$ is a prime.

7. Let $S$ and $T$ be sets and let $M(S, T)$ denote the set of all functions of $S$ into $T$. Let $G$ be a finite group acting on $S$. For each map $f : S \to T$ and $x \in G$ define the map $\pi_x f : S \to T$ by $(\pi_x f)(s) = f(x^{-1}s)$.

(a) Prove that $x \mapsto \pi_x$ is an action of $G$ on $M(S, T)$.

(b) Assume that $S$ and $T$ are finite. Let $n(x)$ denote the number of orbits of the cyclic group $\langle x \rangle$ on $S$. Prove that the number of orbits of $G$ in $M(S, T)$ is equal to

$$\frac{1}{|G|} \sum_{x \in G} |T|^{n(x)}.$$

8. Two actions of a group $G$ on sets $X$ and $Y$ are called equivalent if there is a bijection $f : X \to Y$ such that $f(gx) = gf(x)$ for all $g \in G$ and $x \in X$. Let $H$ and $K$ be subgroups of a group $G$. Let $G$ act by left multiplication on the sets of left cosets $G/H$ and $G/K$. Prove that these actions are equivalent if and only if $H$ and $K$ are conjugate (i.e., $K = aHa^{-1}$ for some $a \in G$).

3. Project. (Semi-direct product) A group $H$ is said to act on a group $K$ by automorphisms if we have an action of $H$ on $K$ and for every $h \in H$ the map $k \mapsto hk$ of $K$ is an automorphism. Suppose this is the case and let $G$ be the product set $K \times H$. Define a binary operation in $K \times H$ by

$$(k_1, h_1)(k_2, h_2) = (k_1h_1k_2, h_1h_2)$$

and define $1 = (1, 1)$ – the units of $K$ and $H$, respectively. Verify that this defines a group such that $h \mapsto (1, h)$ is a monomorphism of $H$ into $K \times H$ and $k \mapsto (k, 1)$ is a monomorphism of $K$ into $K \times H$ whose image is a normal subgroup. This group is called a semi-direct product of $K$ by $H$ and is denoted by $K \rtimes H$.

1.5 Quotient Groups and Cyclic Groups

This section contains a construction of a new group using a normal subgroup, called a quotient group. We also work on an important kind of subgroups of $G$ which are generated by a single element. We conclude this section by studying the group of automorphisms of $G$.

1.5.1 Quotient Groups

Suppose $G$ is any group and $N$ is a normal subgroup of $G$. Then for any $g \in G$,

$$N = gNg^{-1} \quad \text{or} \quad Ng = gN.$$
In other words, every left coset of $N$ in $G$ is also a right coset of $N$ in $G$. If we have two left cosets of $N$ in $G$,

$$N_x = \{ax : a \in N\} \quad \text{and} \quad N_y = \{by : b \in N\},$$

then $N_x N_y = \{axby : a, b \in N\} = N(xN)y = N(Nx)y = Nxy$ is again a left coset of $N$ in $G$. Thus

$$N_x N_y = Nxy$$

defines a binary operation on the set $G/N = \{Nx : x \in G\}$ of left cosets of $N$ in $G$.

**1.5.1. Theorem.** [Quotient Groups] Suppose $G$ is a group and $N$ is a normal subgroup of $G$. Let $G = \bigcup N_{x_a}$ be a decomposition of $G$ as a disjoint union of left (or right) cosets. Then the binary operation

$$N_{x_a} N_{x_{\beta}} = N_{x_a x_{\beta}}$$

makes the set of left cosets of $N$ into a group, called the **quotient or factor group of $G$ by the normal subgroup** $N$. This group is denoted by $G/N$. The map $\pi : G \to G/N$ defined by

$$\pi(x) = Nx$$

is a group homomorphism whose kernel is $N$, called the **canonical projection**.

We have the following observations on the above construction:

1. If $H$ is a subgroup of $G$ which is not normal, then the set of left cosets of $H$ in $G$ does not form a group in any natural way. For example, if $G = S_3$ and $H = \langle (12) \rangle = \{(1), (12)\}$, then $H$ is not normal in $G$ and $\{H, H(13), H(23)\}$ is not a group because

$$H(13)H(23) = \{(13), (132)\}\{(23), (123)\} = \{(132), (12), (13), (1)\}$$

which is not one of the cosets.

2. $|G/N| = |G : N|$, the index of $N$ in $G$.

3. If $G$ is abelian written additively, then

$$G/N = \{N + x : x \in G\}$$

and the binary operation on $G/N$ is given by

$$(N + x) + (N + y) = N + (x + y)$$

for all $x, y \in G$.

We now present three group isomorphism theorems.

**1.5.2. Theorem.** [First Isomorphism Theorem] Suppose $\varphi : G \to H$ is a group homomorphism. Then $G/(\ker \varphi) \cong \im \varphi$.

**Proof.** By Theorem 1.3.31, $\im \varphi$ is a subgroup of $H$ and $\ker \varphi$ is a normal subgroup of $G$, and so $G/(\ker \varphi)$ is a group. Let $\overline{\varphi} : G/\ker \varphi \to \im \varphi$ be $\varphi : x(\ker \varphi) \mapsto \varphi(x)$. Then for all $x, y \in G$,

$$\overline{\varphi}(x(\ker \varphi)) = \overline{\varphi}(y(\ker \varphi)) \iff \varphi(x) = \varphi(y) \iff xy^{-1} \in \ker \varphi \iff x(\ker \varphi) = y(\ker \varphi),$$

so $\overline{\varphi}$ is well defined and 1-1. In addition, for all $x, y \in G$,

$$\overline{\varphi}(x(\ker \varphi)y(\ker \varphi)) = \overline{\varphi}(xy(\ker \varphi)) = \varphi(xy) = \varphi(x)\varphi(y) = \overline{\varphi}(x(\ker \varphi))\overline{\varphi}(y(\ker \varphi)).$$

Moreover, $\overline{\varphi}$ is clearly onto. Hence, $\varphi$ is an isomorphism. \qed
1.5.3. Examples. \( \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n, D_n/R_n \cong \mathbb{Z}_2, \mathbb{R}/\mathbb{Z} \cong S^1 \) and \( \text{GL}_n(F)/\text{SL}_n(F) \cong F \setminus \{0\} \).

1.5.4. Definition. Let \( G \xrightarrow{\theta} H \xrightarrow{\varphi} K \) be a sequence of group homomorphisms. We say that it is exact at \( H \) if \( \text{im} \theta = \ker \varphi \). A short exact sequence of groups is a sequence of groups and homomorphisms

\[
1 \longrightarrow G \xrightarrow{\theta} H \xrightarrow{\varphi} K \longrightarrow 1
\]

which is exact at \( G, H \) and \( K \). That is, \( \theta \) is injective, \( \varphi \) is surjective and \( \text{im} \theta = \ker \varphi \). Here, \( 1 \) stands for the smallest group of order one.

1.5.5. Remark. If \( N \) is a normal subgroup of \( G \), then

\[
1 \longrightarrow N \xrightarrow{\iota} G \xrightarrow{\pi} G/N \longrightarrow 1
\]

is exact. Here \( \iota \) denotes the inclusion map. On the other hand, if \( N \leq G \) and

\[
1 \longrightarrow N \xrightarrow{\iota} G \longrightarrow H \longrightarrow 1
\]

is exact, then \( N \) is normal in \( G \) and \( H \cong G/N \). Thus short exact sequences are just another notation for normal subgroups and factor groups.

1.5.6. Theorem. [Second Isomorphism Theorem] Suppose \( G \) is a group and \( H \) and \( N \) are subgroups of \( G \), with \( N \) normal in \( G \). Then \( HN = NH \) is a subgroup of \( G \) in which \( N \) is normal, \( H \cap N \) is normal in \( H \) and \( H/(H \cap N) \cong HN/N \).

Proof. Since \( N \) is normal in \( G \), \( hN = Nh \) for all \( h \in H \), so \( HN \subseteq NH \) and \( NH \subseteq HN \). Thus, \( NH = HN \). It is routine to show that \( NH \) is a subgroup of \( G \). Since \( N \not\leq G \), \( N \not\leq NH \). The theorem follows from exactness of the sequence

\[
1 \longrightarrow H \cap N \xrightarrow{\iota} H \xrightarrow{\varphi} HN/N \longrightarrow 1,
\]

where the homomorphism \( \varphi : h \mapsto hN \) for all \( h \in H \) and \( \ker \varphi = \{ h \in H : hN = N \} = \{ h \in H : h \in N \} = H \cap N \). \( \square \)

1.5.7. Remark. If \( H \) and \( N \) are not normal in \( G \), then \( HN \) may not be a subgroup of \( G \). E.g., in \( S_3 \), the subgroups \( H = \{(1), (12)\} \) and \( N = \{(1), (13)\} \) are not normal in \( S_3 \) and \( HN = \{(1), (12), (13), (132)\} \) is not a subgroup of \( S_3 \).

1.5.8. Theorem. [Third Isomorphism Theorem] Suppose \( G \) is a group and \( N \) is a normal subgroup of \( G \). Then the map

\[
\theta : H \mapsto H/N
\]

gives a 1-1 correspondence

\[
\{ \text{subgroups of } G \text{ containing } N \} \longleftrightarrow \{ \text{subgroups of } G/N \}.
\]

This correspondence carries normal subgroups to normal subgroups. Moreover, if \( H \) is normal in \( G \) containing a subgroup \( N \), then

\[
G/H \cong (G/N)/(H/N).
\]
1.5. Quotient Groups and Cyclic Groups

Proof. Let $H_1$ and $H_2$ be subgroups of $G$ containing $N$ and assume that $H_1/N = H_2/N$. Let $x \in H_1$. Then $Nx \in H_1/N = H_2/N$, so $Nx = Ny$ for some $y \in H_2$. Thus, $xy^{-1} \in N \subseteq H_2$. Since $y \in H_2$, $x \in H_2$. Hence, $H_1 \subseteq H_2$. By symmetry, $H_2 \subseteq H_1$. Therefore, $H_1 = H_2$ and $\theta$ is 1-1. Next, let $\mathcal{H} \leq G/N$. Then $\{N\} \subseteq \mathcal{H}$. Choose $H = \bigcup \mathcal{H}$, the union of cosets in $\mathcal{H}$. Thus, $N \subseteq H$. Let $x, y \in H$. Then $Nx, Ny \in \mathcal{H}$, so $Nx y^{-1} \in \mathcal{H}$ which implies $xy^{-1} \in H$. Thus, $H$ is a subgroup of $G$ containing $N$ and $\mathcal{H} = H/N$. Hence, $\theta$ is onto.

Assume that $H$ is a normal subgroup of $G$ containing $N$. Let $g \in G$ and $x \in H$. Then $g x g^{-1} \in H$, so $g N x g^{-1} N = g x g^{-1} N \in H/N$. Hence, $H/N$ is normal subgroup of $G/N$.

The final isomorphism follows from exactness of the sequence

$$1 \rightarrow H/N \rightarrow G/N \rightarrow G/H \rightarrow 1,$$

where the homomorphism $\varphi : gN \mapsto gH$ for all $g \in G$ which is well defined because $N \subseteq H$, and $\ker \varphi = \{gN : g \in G$ and $gH = H\} = \{gN : g \in H\} = H/N$. \hfill \square

1.5.2 Cyclic Groups

Recall that a cyclic subgroup of $G$ is a subgroup of $G$ generated by a singleton. It has a simple structure and is easy to construct. Its properties depend mostly on the group of integer modulo $n$. We shall go deep inside the groups $\mathbb{Z}_n$ and $\mathbb{Z}_n^\times$ in this section. We recall from Examples 1.3.11 that:

1.5.9. Theorem. Let $n \geq 2$ and $\mathbb{Z}_n^\times = \{a : \gcd(a, n) = 1\}$. Then $(\mathbb{Z}_n^\times, \cdot)$ is an abelian group of order $\phi(n)$, the Euler $\phi$-function.

1.5.10. Example. $\mathbb{Z}_{10}^\times = \{1, 3, 7, 9\}$ and $\mathbb{Z}_p^\times = \{1, 2, \ldots, p-1\}$ where $p$ is a prime.

Now we study cyclic subgroups of a group $G$. Recall that if $G$ is a group and $a \in G$, then $\langle a \rangle = \{a^n : m \in \mathbb{Z}\}$ and the order of $a$ is $|a| = |\langle a \rangle|$.

1.5.11. Theorem. Let $G$ be a group and $a \in G$. Then

1. $\forall n \in \mathbb{N}$, $a^n = e \Rightarrow \langle a \rangle = \{e, a, a^2, \ldots, a^{n-1}\}$.
2. $|a|$ is finite $\iff \exists i, j \in \mathbb{Z} (i \neq j \land a^i = a^j) \iff \exists n \in \mathbb{N}, a^n = e$.
3. $|a|$ is infinite $\iff \forall i, j \in \mathbb{Z} (i \neq j \Rightarrow a^i \neq a^j) \iff \exists a \in \mathbb{Z}$.
4. If $G = \langle a \rangle$ is infinite, then $a$ and $a^{-1}$ are only two generators of $G$.
5. If $G$ is finite, then $|a| = \min\{n \in \mathbb{N} : a^n = e\}$ and $\langle a \rangle = \{e, a, a^2, \ldots, a^{|a|-1}\} \cong \mathbb{Z}_n$.
6. $\forall n \in \mathbb{Z}$, $a^n = e \Rightarrow |a|$ divides $n$.
7. If $G$ is finite, then $a^{|G|} = e$.

Proof. (1)–(3) are clear.

(4) Assume that $G = \langle a \rangle$ is infinite and $a^m$ is a generator for some $m \in \mathbb{Z}$. Then $\langle a^n \rangle = \langle a \rangle$, so $a = (a^m)^k$ for some $k \in \mathbb{Z}$. Since $|a|$ is infinite, $mk = 1$. Thus, $m | 1$, so $m = \pm 1$.

(5) Assume that $G$ is finite. Then $a^n = e$ for some $n \in \mathbb{N}$. Choose $n_0$ to be the smallest such $n$. Thus, $a^{n_0} = e$. We shall show that $\langle a \rangle = \{e, a, a^2, \ldots, a^{n_0-1}\}$. Clearly, $\{e, a, a^2, \ldots, a^{n_0-1}\} \subseteq \langle a \rangle$.

Let $j \in \mathbb{Z}$. Then $j = n_0 q + r$ for some $q, r \in \mathbb{Z}$ and $0 \leq r < n_0$, so $a^j = a^{n_0 q + r} = a^r \in \{e, a, a^2, \ldots, a^{n_0-1}\}$. Hence, $|a| = n_0 = \min\{n \in \mathbb{N} : a^n = e\}$ and $a^{|a|} = e$. Finally, an isomorphism is given by $a^r \rightarrow j$ for all $j \in \mathbb{Z}$.

(6) Let $n \in \mathbb{Z}$ and $a^n = e$. By the division algorithm, $n = |a|q + r$ for some $q, r \in \mathbb{Z}$ and $0 \leq r < |a|$. If $r > 0$, then $e = a^n = a^{|a|q + r} = a^r$ which contradicts the minimality of $|a|$. Hence, $r = 0$ and so $|a|$ divides $n$.

(7) By Lagrange, $|a|$ divides $|G|$, so $|G| = |a|q$ for some $q \in \mathbb{Z}$. Then $a^{|G|} = a^{|a|q} = e$. \hfill \square
The above results for the group $\mathbb{Z}_n^\times$ yield famous results below.

1.5.12. Corollary. 1. [Euler] If $a \in \mathbb{Z}$ and $\gcd(a, n) = 1$, then $a^{\varphi(n)} \equiv 1 \pmod{n}$.  
2. [Fermat] If $p$ is a prime, then $a^p \equiv a \pmod{p}$ for all $a \in \mathbb{Z}$.

Proof. (1) Apply the above theorem to $G = \mathbb{Z}_n^\times$. 
(2) If $\gcd(a, p) = 1$, then by (1), $a^{p-1} \equiv 1 \pmod{p}$, so $a^p \equiv a \pmod{p}$. If $\gcd(a, p) > 1$, then $p \mid a$, so $p \mid (a^p - a)$. Hence, $a^p \equiv a \pmod{p}$ for all $a \in \mathbb{Z}$. 

1.5.13. Theorem. Any two cyclic groups of the same orders (finite or infinite) are isomorphic.

Proof. Assume that $G$ is cyclic. Then $G = \langle a \rangle$ for some $a \in G$. By Theorem 1.5.11, if $G$ is infinite, then $G \cong \mathbb{Z}$, and if $G$ is finite, then $|G| = |a|$ and $G = \{e, a, \ldots, a^{[a]}-1\} \cong \mathbb{Z}_{|a|}$.

Next, we study subgroups of a cyclic group.

1.5.14. Theorem. [Subgroups of a Cyclic Group] Let $G$ be a cyclic group generated by $a$, and let $H$ be a subgroup of $G$. Then $H$ is also a cyclic group generated by $a^k$ where $k = \min\{m \in \mathbb{N} : a^m \in H\}$ or $H = \{e\}$. Consequently, every subgroup of a cyclic group is cyclic.

Proof. Since $a^k \in H$, $\langle a^k \rangle \subseteq H$. Let $x \in H$. Then $x \in G$, so $x = a^t$ for some $t \in \mathbb{Z}$. By the division algorithm, $t = kq + r$ for some $q, r \in \mathbb{Z}$ and $0 \leq r < k$. Thus, 

$$a^r = a^{t-kr} = a^t a^{-kr} = x(a^k)^{-q} \in H.$$ 

But $r < k$, so $r = 0$. Hence, $x = a^{kr} = (a^k)^q \in \langle a^k \rangle$.

1.5.15. Corollary. All distinct subgroups of $\mathbb{Z}$ are $k\mathbb{Z} = \{kq : q \in \mathbb{Z}\}$ where $k \in \mathbb{N} \cup \{0\}$.

1.5.16. Theorem. [Generators of a Subgroup of a Finite Cyclic Group] Let $G$ be a finite cyclic group of order $n$. Then $G$ has exactly one subgroup $H$ of order $d$ for each divisor $d$ of $n$, and no other subgroups. Moreover, if $G$ is generated by $a$, then $H$ is generated by $a^{n/d}$.

Proof. Let $d \mid n$. Since $(a^{n/d})^d = e$, $|a^{n/d}| \leq d$. If $|a^{n/d}| = r < d$, then $a^{nr/d} = e$ and $nr/d < n$ which contradicts $|a| = n$. Thus, $|a^{n/d}| = d$. Let $H$ be a subgroup of $G$ of order $d$. If $d = 1$, then $H = \{e\}$. Assume that $d > 1$. By Theorem 1.5.14, $H = \langle a^k \rangle$, where $k = \min\{m \in \mathbb{N} : a^m \in H\}$. Since $|H| = d$, $(a^k)^d = e$, so $n \mid kd$ which implies $\frac{n}{d} \mid k$. Thus, $k = \frac{n}{d} q$ for some $q \in \mathbb{Z}$. Hence, $a^k = (a^{n/d})q \in \langle a^{n/d} \rangle$. It follows that $H \subseteq \langle a^{n/d} \rangle$. However, $|H| = d = |\langle a^{n/d} \rangle|$, so $H = \langle a^{n/d} \rangle$. 

1.5.17. Example. All subgroups of the cyclic group $G = \langle a \rangle$ of order 12 are shown in the following diagram.

![Diagram](attachment:image.png)
The order of an element in a cyclic group and its generators are studied in the next theorem.

**1.5.18. Theorem.** [Order of an Element] Let $G$ be a finite cyclic group of order $n$ generated by $a$ and $m \in \mathbb{Z}$. Then

1. $\langle a^m \rangle = \langle a^d \rangle$, where $d = \gcd(m, n)$.
2. $|a^m| = \frac{n}{\gcd(m, n)}$.
3. $a^m$ is a generator of $G$ if and only if $\gcd(m, n) = 1$, and so $G$ contains precisely $\phi(n)$ elements of order $n$.

**Proof.** (1) Since $d \mid m$, $\langle a^m \rangle \subseteq \langle a^d \rangle$. Since $d = \gcd(m, n)$, $d = mx + ny$ for some $x, y \in \mathbb{Z}$, so

$$a^d = a^{mx + ny} = a^{mx}a^{ny} = a^{mx} \in \langle a^m \rangle.$$

(2) $|a^m| = |\langle a^m \rangle| = |\langle a^d \rangle| = |\langle a^{n/(n/d)} \rangle| = \frac{n}{d}$ from Theorem 1.5.16.

(3) $\langle a^m \rangle = G \iff |a^m| = n \iff \frac{n}{d} = n \iff d = 1$. \qed

**1.5.19. Remark.** Since $\mathbb{Z}_n = \langle \bar{1} \rangle$ and $\bar{m} \cdot \bar{1} = \bar{m}$, we have $\langle \bar{m} \rangle = \mathbb{Z}_n \iff \gcd(m, n) = 1 \iff \bar{m} \in \mathbb{Z}_n^\times$.

**1.5.20. Theorem.** Let $G$ be a group and $a \in G$. Then $|a| = n \iff (\forall k \in \mathbb{N}, a^k = e \iff n \mid k)$.

**Proof.** Assume that $\forall k \in \mathbb{N}, a^k = e \iff n \mid k$. Since $n \mid n$, $a^n = e$. Let $k \in \mathbb{N}$ be such that $a^k = e$. Then $n \mid k$, so $n \leq k$. Hence, $|a| = n$. Another direction follows from Theorem 1.5.11 (6). \qed

Recall that an automorphism of a group $G$ is an isomorphism on $G$. The set of all automorphisms of a group $G$ is denoted by $\text{Aut} G$ and is called the automorphism group of $G$. We shall close this section by studying the group of automorphisms of $G$ and determining the automorphism group of cyclic groups.

**1.5.21. Theorem.** [Inner Automorphisms]

1. With group operation composition of functions, $\text{Aut} G$ is a group.
2. Each $g \in G$ determines an automorphism $\phi_g : G \to G$ defined by

$$\phi_g(x) = gxg^{-1} \text{ for all } x \in G,$$

and $\phi_g$ is called an inner automorphism. The subgroup of $\text{Aut} G$ consisting of the $\{\phi_g : g \in G\}$ is called the inner automorphism group of $G$ and is denoted by $\text{Inn} G$.
3. The map $\theta : g \mapsto \phi_g$ is a group homomorphism from $G$ into $\text{Aut} G$.
4. The kernel of $\theta$ is $Z(G)$, the center of $G$, and the image of $\theta$ is $\text{Inn} G$. Consequently,

$$G/Z(G) \cong \text{Inn} G \leq \text{Aut} G.$$

**1.5.22. Example.** $\text{Aut} \mathbb{Z} \cong \mathbb{Z}_2$ and $\text{Aut} \mathbb{Z}_n \cong \mathbb{Z}_n^\times$.

**Proof.** Let $\varphi \in \text{Aut} \mathbb{Z}$. Note that for each $k \in \mathbb{N}$, $\varphi(k) = \varphi(k \cdot 1) = \varphi(\overbrace{1 + \cdots + 1}^k) = k \cdot \varphi(1)$ and $\varphi(-k) = -\varphi(k) = -(k \cdot \varphi(1))$, so $\varphi$ is completely determined by $\varphi(1)$. Since $\varphi$ is onto, $\text{im} \varphi = \varphi(1)\mathbb{Z} = \mathbb{Z}$. Thus, $\varphi(1) \mid 1$, so $\varphi(1) = \pm 1$. Hence, $\text{Aut} \mathbb{Z} = \{\pm 1\} \cong \mathbb{Z}_2$.

Let $\varphi \in \text{Aut} \mathbb{Z}_n$. Similarly, $\varphi$ is completely determined by $\varphi(\bar{1})$. Since $\varphi$ is onto, $\text{im} \varphi = \langle \varphi(\bar{1}) \rangle = \mathbb{Z}_n$. By Remark after Theorem 1.5.18, $\varphi(1) \in \mathbb{Z}_n^\times$. Therefore, $\text{Aut} \mathbb{Z}_n \cong \mathbb{Z}_n^\times$ with isomorphism $\varphi \mapsto \varphi(\bar{1})$. \qed
1.5. Exercises.  
1. Prove that if $G$ is a group for which $G/Z(G)$ is cyclic, then $G$ is abelian.
2. Let $G$ be a group of order $2k$ where $k$ is odd. Show that $G$ contains a subgroup of index 2.
3. Let $H$ be a proper subgroup of a finite group $G$. Show that $G \neq \bigcup_{g \in G} gHg^{-1}$.
4. Let $G$ be a group and $a \in G$. If $(a) \triangleleft G$ and $H < (a)$, prove that $H$ is normal in $G$.
5. Let $G$ be a group and $N$ a subgroup contained in the center of $G$. Suppose that $G/N$ is cyclic. Prove that $G$ is necessarily abelian.
6. Let $G$ be a group. If $a, b \in G$ are of finite order such that $ab = ba$ and $\forall m \in \mathbb{N}$, $a^m b^m = e \implies a^m = b^m = e$, prove that $[ab] = \text{lcm}([a],[b])$.
7. Let $m$ and $n$ be integers. Prove the following statements.
   (a) $m\mathbb{Z} + n\mathbb{Z} = d\mathbb{Z}$ and $m\mathbb{Z} \cap n\mathbb{Z} = \mathbb{Z}$ where $d = \gcd(m,n)$ and $l = \text{lcm}(m,n)$.
   (b) If $\gcd(m,n) = 1$, then $\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$. This is called the “Chinese remainder theorem”. Is the converse true?
8. Let $G$ be a group, $K$ a normal subgroup of $G$ of index $r$, and let $g \in G$ be an element of order $n$. Prove that if $r$ and $n$ are relatively prime, then $g \in K$.
9. Prove the following statements.
   (a) If $\gcd(m,n) = 1$, then $\text{Aut}(\mathbb{Z} \times \mathbb{Z}) \cong \mathbb{Z}^\times \times \mathbb{Z}_n^\times$.
   (b) $\text{Aut}(\mathbb{Z}_p \times \mathbb{Z}_p) \cong \text{GL}_2(\mathbb{Z}_p)$.
10. Prove Theorem 1.5.21.
11. Let $H < G$. Prove that $G/H \triangleleft N_G(H)$ and $N_G(H)/G/H$ is isomorphic to a subgroup of $N_G(H)$.
12. If $G$ is a group for which $\text{Aut}(G) = \{1\}$, prove that $|G| \leq 2$.


   (a) Let $G$ be a finite group. For each positive integer $a$, let $[a]^G$ be the set of functions from $G$ to $\{1, 2, \ldots, a\}$. Prove that
   \[(g \cdot f)(h) = f(g^{-1}h) \quad \text{for all } g, h \in G \text{ and } f \in [a]^G\]


   (b) Show that
   \[\sum_{g \in G} a^{[G]/[a]} \equiv 0 \pmod{|G|}.
   \]
   [Hint. Use Burnside’s Theorem with the action in (a) to conclude that $\frac{1}{|G|} \sum_{g \in G} a^{[G]/[a]}$ is a positive integer.]

   (c) Taking $G = \mathbb{Z}_m$, deduce that $\sum_{k=1}^m a^{\text{ord}(k,m)} \equiv 0 \pmod{m}$.

1.6 The Symmetric Group

In this section, we study the symmetric group on $n$ letters, $S_n$. Recall that $S_n$ is the group of permutations (1-1 and onto maps) on $\{1,2,\ldots,n\}$ under composition. Its order is $n!$.

1.6.1. Definition. A permutation $\gamma$ of $\{1,2,\ldots,n\}$ which permutes a sequence of distinct elements $i_1, i_2, \ldots, i_r$, $r > 1$, cyclically in the sense that

   \[\gamma(i_1) = i_2, \gamma(i_2) = i_3, \ldots, \gamma(i_{r-1}) = i_r, \text{ and } \gamma(i_r) = i_1\]

and fixed (that is, leaves unchanged) the other numbers in $\{1,2,\ldots,n\}$ is called a cycle or an $r$-cycle. We denote this as

   \[\gamma = (i_1 i_2 \ldots i_r).\]
It is clear that we can equally well write
\[ \gamma = (i_2 i_3 \ldots i_r i_1) = (i_3 i_4 \ldots i_r i_1 i_2), \] etc.

**1.6.2. Definition.** Two cycles \( \gamma \) and \( \gamma' \) are said to be disjoint if their symbols contain no common letters.

In this case, it is clear that any number moved by one of these transformations is fixed by the other, i.e., \( \forall i \), \( \gamma(i) \neq i \Rightarrow \gamma'(i) = i \). Hence, if \( i \) is any number such that \( \gamma(i) \neq i \), then \( \gamma \gamma'(i) = \gamma(i) \), and since also \( \gamma^2(i) \neq \gamma'(i) \), \( \gamma \gamma'(i) \). Similarly, if \( \gamma'(i) \neq i \), then \( \gamma \gamma'(i) = \gamma'(i) = \gamma \gamma'(i) \). Also if \( \gamma(i) = i = \gamma'(i) \), then \( \gamma \gamma'(i) = \gamma'(i) \). Thus \( \gamma \gamma' = \gamma \gamma' \), that is, we have proved (1) of the following theorem.

**1.6.3. Theorem.** [Order of a Cycle]

1. Any two disjoint cycles commute.
2. If \( \gamma = (i_1 i_2 \ldots i_r) \) is an \( r \)-cycle, then the order of \( \gamma \) is \( r \).
3. If \( \alpha = (i_1 i_2 \ldots i_{r_1})(j_1 j_2 \ldots j_{r_2}) \ldots (k_1 k_2 \ldots k_{r_s}) \) is a product of disjoint cycles, then the order of \( \alpha \) is the least common multiple of \( r_1, r_2, \ldots, r_s \).

**Proof.** For (2), clearly, \( \gamma^r = (1) \). Let \( 1 \leq s < r \). Then \( \gamma^s(i_1) = i_{s+1} \neq i_1 \), so \( \gamma^s \neq (1) \). (3) follows from (2) and the fact that \( |ab| = \text{lcm}(|a|, |b|) \) for all \( a, b \in G \) such that \( ab = ba \) and \( \forall m \in \mathbb{N}, a^m b^m = e \Rightarrow a^m = b^m = e \) (see Exercises 1.5).

It is convenient to extend the definition of cycles and the cycle notation to 1-cycles where we adopt the convention that for any \( i \), \( (i) \) is the identity mapping. With this convention, we can see that:

**1.6.4. Theorem.** [Decomposition of a Permutation] Every permutation is a product of disjoint cycles. Moreover, the product is unique up to rearranging its cycles and cyclically permuting the numbers within each cycle.

**Proof.** Let \( \sigma \in S_n \). If \( \sigma = (1) \), we are done. Assume that \( \sigma \neq 1 \). Let \( G = \langle \sigma \rangle \) act on \( \{1, 2, \ldots, n\} \) naturally as in Examples 1.4.4 (1). Let \( B_1, B_2, \ldots, B_r \) be distinct orbits of \( \{1, 2, \ldots, n\} \) under this action. For each \( j \in \{1, 2, \ldots, r\} \), we define the cycle \( \mu_j \) by
\[
\mu_j(x) = \begin{cases} 
\sigma(x), & \text{if } x \in B_j; \\
x, & \text{if } x \in \{1, 2, \ldots, n\} \setminus B_j.
\end{cases}
\]

Since \( B_i, i = 1, 2, \ldots, r \), are disjoint, \( \mu_i \) are disjoint cycles, and clearly, \( \sigma = \mu_1 \mu_2 \ldots \mu_r \).

**1.6.5. Remark.** The above two theorems tell us how to find the order of an element in \( S_n \).

Next, we shall discuss the cycle structure and the conjugacy class of a permutation.

**1.6.6. Lemma.** If \( \alpha \in S_n \) is a permutation, then
\[
\alpha(i_1 i_2 \ldots i_r) \alpha^{-1} = (\alpha(i_2) \alpha(i_3) \ldots \alpha(i_r)) \alpha(i_1).
\]

**Proof.** For \( x \in \{1, 2, \ldots, n\} \),
\[
\alpha(i_1 i_2 \ldots i_r)(x) = \begin{cases} 
\alpha(i_{m+1} \mod r), & \text{if } x = i_m \mod r; \\
\alpha(x), & \text{if } x \notin \{i_1, i_2, \ldots, i_r\} \\
\alpha(i_{m+1} \mod r), & \text{if } \alpha(x) = \alpha(i_m \mod r); \\
\alpha(x), & \text{if } \alpha(x) \notin \{\alpha(i_1), \alpha(i_2), \ldots, \alpha(i_r)\} \\
\end{cases}
\]
\[
= (\alpha(i_1) \alpha(i_2) \ldots \alpha(i_r))(\alpha(x)).
\]

Hence, \( \alpha(i_1 i_2 \ldots i_r) = (\alpha(i_1) \alpha(i_2) \ldots \alpha(i_r)) \alpha \).
1.6.7. **Definition.** If $\sigma \in S_n$ is the product of disjoint cycles of lengths $r_1, r_2, \ldots, r_s$ with $r_1 \leq r_2 \leq \ldots \leq r_s$ (including its 1-cycles) then the integers $r_1, r_2, \ldots, r_s$ are called the cycle structure of $\sigma$.

1.6.8. **Definition.** A partition of a positive integer $n$ is any nondecreasing sequence of positive integers whose sum is $n$.

For example, 5 has seven partitions, namely, $1+1+1+1+1$, $1+1+1+2$, $1+2+2$, $1+1+3$, $1+4$, $2+3$ and 5. The following result shows that the number of conjugacy classes of $S_n$ and the number of partitions of $n$ are coincide.

1.6.9. **Theorem.** Two elements of $S_n$ are conjugate if and only if they have the same cycle structure. The number of conjugacy classes of $S_n$ equals the number of partitions of $n$.

**Proof.** Assume that $\sigma$ and $\tau$ are conjugate. Then $\tau = \alpha \sigma \alpha^{-1}$ for some $\tau \in S_n$. Write

$$\sigma = (i_1 i_2 \ldots i_{r_1}) (j_1 j_2 \ldots j_{r_2}) \ldots (k_1 k_2 \ldots k_{r_s})$$

as a product of disjoint cycles. Thus,

$$\tau = \alpha \sigma \alpha^{-1} = \alpha (i_1 i_2 \ldots i_{r_1}) \alpha^{-1} \alpha (j_1 j_2 \ldots j_{r_2}) \alpha^{-1} \alpha (k_1 k_2 \ldots k_{r_s}) \alpha^{-1}$$

$$= (\alpha (i_1) \alpha (i_2) \ldots \alpha (i_{r_1})) (\alpha (j_1) \alpha (j_2) \ldots \alpha (j_{r_2})) (\alpha (k_1) \alpha (k_2) \ldots \alpha (k_{r_s})).$$

Hence, $\sigma$ and $\tau$ have the same cycle structure.

Conversely, suppose that $\sigma$ and $\tau$ have the same cycle structure written as a product of $s$ disjoint cycles (including 1-cycles) as

$$\sigma = (a_1 a_2 \ldots a_{r_1}) (a_{r_1+1}a_{r_1+2} \ldots a_{r_1+r_2}) \ldots (a_{r_1+r_2+\ldots+r_{s-1}+1} \ldots a_{n-1}a_n)$$

and

$$\tau = (b_1 b_2 \ldots b_{r_1}) (b_{r_1+1}b_{r_1+2} \ldots b_{r_1+r_2}) \ldots (b_{r_1+r_2+\ldots+r_{s-1}+1} \ldots b_{n-1}b_n).$$

Define $\alpha \in S_n$ by $\alpha (a_i) = b_i$ for all $i \in \{1, 2, \ldots, n\}$. Then $\alpha \sigma \alpha^{-1} = \tau$. □

1.6.10. **Example.** The number of conjugacy classes of $S_5$ is 7 and $|\{\alpha (12)(345)\alpha^{-1} : \alpha \in S_5\}| = \binom{5}{2} (3)! = 20.$

1.6.11. **Example.** The Klein group $V_4 = \{(1), (12)(34), (13)(24), (14)(23)\}$ is normal in $S_4$ because $V_4$ contains all products of disjoint 2-cycles and so $\forall \alpha \in S_4, \alpha V_4 \alpha^{-1} = V_4$ by Theorem 1.6.9. Moreover, since the group $(1), (12)(34)$ is of index two in $V_4$, it is normal in $V_4$ by Theorem 1.4.9. However, $(1), (12)(34)$ is not normal in $S_4$. Thus, normality of subgroups is not transitive.

1.6.12. **Corollary.** [Center of $S_n$] If $n \geq 3$, then the center of $S_n$ is trivial, i.e., $Z(S_n) = \{(1)\}$.

**Proof.** We wish to prove that $\forall \alpha \in S_n |\forall \beta \in S_n, \beta \alpha \beta^{-1} = \alpha \Rightarrow \alpha = (1)|$. By Theorem 1.6.9, $\forall \alpha \in S_n |\alpha \neq (1) \Rightarrow |\{\beta \alpha \beta^{-1} : \beta \in S_n\}| > 1 \Rightarrow \alpha \notin Z(S_n)|$. Hence, let $\alpha \in Z(S_n)$. Then $|\{\beta \alpha \beta^{-1} : \beta \in S_n\}| = 1$, so $\alpha = (1)$. □

To define an important subgroup of $S_n$, namely the alternating group, we shall need some results on 2-cycles. A cycle of the form $(ab)$, where $a \neq b$, is called a transposition. It is easy to verify that

$$(i_1 i_2 \ldots i_r) = (i_1 i_r) \ldots (i_1 i_3)(i_1 i_2),$$

a product of $r - 1$ transpositions. It follows that any $\alpha \in S_n$ is a product of transpositions. Also, a transposition $(ab)$ has order two in $S_n$. 

1.6.13. Theorem. 1. (1) is always a product of even number of transpositions.
2. If \( \alpha \in S_n \) is written as a product of transpositions, then either the number of transpositions in any product is always odd or always even.

Proof. (1) Assume that
\[
(1) = (x_1y_1)(x_2y_2) \cdots (x_ky_k) = (1x_1)(1y_1)(1x_2)(1y_2)(1x_2) \cdots (1x_k)(1y_k)(1x_k)
\]
with \( x_i < y_i \) for all \( i \in \{1, 2, \ldots, k\} \). Consider any \((1u), u > 1, \) in the right hand side. Since the opposite side is \((1), (1u)\) must occur twice (or even number of times) in the right hand side. Note that \((1 \rightarrow u \text{ and } u \rightarrow 1)\) will give \( u \rightarrow u\). Thus each transposition in the right hand side occurs even numbers of times, which implies that the right hand side should have even number of transpositions. Hence, \( k \) is even.

(2) Assume that we have two transposition decompositions
\[
\alpha = (x_1y_1)(x_2y_2) \cdots (x_ky_k) = (w_1z_1)(w_2z_2) \cdots (w_lz_l)
\]
for some \( x_i \neq y_i, w_j \neq z_j \) and \( k, l \in \mathbb{N} \). Since \(|(w_1z_1)| = 2\) for all \( i \),
\[
(x_1y_1)(x_2y_2) \cdots (x_ky_k)(w_1z_1)^{-1}(w_2z_2)^{-1} \cdots (w_lz_l)^{-1} = (1)
\]
so \( k + l \) is even. Hence, \( k \) and \( l \) have the same parity. \( \square \)

The previous theorem leads to the definition of parity of a permutation.

1.6.14. Definition. We call the permutation \( \alpha \) even or odd according as \( \alpha \) factors as a product of an even or an odd number of transpositions.

1.6.15. Remarks. Let \( \alpha, \beta \in S_n \).
1. \( \alpha \beta \) is even \( \iff \alpha \) and \( \beta \) have the same parity.
2. Since \( \alpha \alpha^{-1} = (1) \) which is even, \( \alpha \) and \( \alpha^{-1} \) have the same parity.

1.6.16. Theorem. Let \( n > 1 \). The set \( A_n \) of all even permutations forms a normal subgroup of \( S_n \) of index two. It is called the alternating group of degree \( n \) and \(|A_n| = n!/2\).

Proof. By Theorem 1.6.13, (1) is even. It is clear that the product of even permutations is even. Since a transposition has order two, the inverse of an even permutation is even. Hence, \( A_n \) is a subgroup of \( S_n \). Since \( n > 1 \), let \((ab)\) be a transposition in \( S_n \). Clearly, \((ab)\) is an odd permutation. We will show that \( S_n = A_n \cup (ab)A_n \). Let \( \alpha \in S_n \). If \( \alpha \) is even, then \( \alpha \in A_n \). On the other hand, assume that \( \alpha \) is odd. Then \((ab)\alpha \) is even, so \((ab)\alpha \in A_n \), i.e., \( \alpha \in (ab)A_n \). Thus, \(|S_n : A_n| = 2\). In addition, since \( \alpha \) and \( \alpha^{-1} \) have the same parity, \( \alpha A_n \alpha^{-1} \subseteq A_n \). Hence, \( A_n \) is normal in \( S_n \). \( \square \)

The above proof also shows that if \( n > 1 \), then \( S_n = A_n \cup (ab)A_n \) and the number of even permutations and odd permutations are the same.

1.6.17. Corollary. Let a group \( G \) act on a finite set \( X \), and assume that some element \( h \in G \) induces an odd permutation on \( X \). Then there exists a normal subgroup \( N \) of \( G \) with \(|G : N| = 2\) and \( h \notin N \).
Proof. Consider the diagram

\[ G \xrightarrow{\theta} S(X) \xrightarrow{\pi} S(X)/A(X), \]

where \(A(X)\) is the alternating group of even permutations on \(X\), \(\theta : g \mapsto \phi_g\) and \(\pi\) is the canonical map. Since \(\phi_h\) is an odd permutation, \(\pi \circ \theta\) is onto. Choose \(N = \ker \pi \circ \theta\). Then \(N \triangleleft G\) and \(G/N \cong S(X)/A(X)\). Thus, 
\(|G : N| = [S(X) : A(X)] = 2\). Since \((\pi \circ \theta)(h) = \phi_hA(X) \neq A(X)\), we have \(h \notin N\).

Finally, we talk about the simplicity of \(A_n\).

1.6.18. Definition. A group is simple if it has no nontrivial normal subgroup. That is, all normal subgroups of \(G\) are \(\{e\}\) and \(G\).

For example, \(\mathbb{Z}_p\) is simple for all primes \(p\).

1.6.19. Corollary. Let \(|G| = 2m\), where \(m\) is odd. Then \(G\) has a normal subgroup of order \(m\). In particular, if \(m > 1\), then \(G\) is not simple.

Proof. Since \(|G|\) is even, let \(g\) be an element of order two in \(G\). Let \(G\) act on \(G\) by left multiplication and consider \(G \xrightarrow{\theta} S(G)\). Since the action is faithful, \(\theta\) is 1-1, so \(|\theta(g)| = 2\) Thus, \(\theta(g) = \phi_g\) is an odd permutation. By the previous corollary, there exists a normal subgroup \(N\) of \(G\) such that \(|G : N| = 2\). Hence, \(G\) is not simple.

1.6.20. Example. Since the Klein group \(V_4 = \{(1), (12)(34), (13)(24), (14)(23)\}\) is normal in \(A_4\), it follows that \(A_4\) is not simple.

In general, we have the next theorem.

1.6.21. Theorem. \(A_n\) is simple for all \(n \neq 4\).

Proof. Clearly, \(A_2\) and \(A_3\) are simple. For \(n \geq 5\), we give a step-by-step guideline in Project 5.

1.6.22. Corollary. If \(n \neq 4\), then the only normal subgroups of \(S_n\) are \(\{(1)\}\), \(A_n\) and \(S_n\).

1.6. Exercises. 
1. Prove that \(A_4\) is the only subgroup of \(S_4\) of order 12.
2. Prove that \(A_4\) has no subgroup of order six.
3. Determine all normal subgroups of \(S_4\). (Hint. Use conjugacy classes.)
4. (a) Find the largest positive integer \(n\) such that \(S_{10}\) has a permutation of order \(n\).
   (b) The exponent of a finite group \(G\) is the smallest positive integer \(n\) such that \(g^n = 1\) for all \(g \in G\). Find the exponent of \(S_{30}\), the symmetric group on 30 letters.
5. Show that if \(H\) is any subgroup of \(S_n\), \(n \geq 2\), then either all permutations in \(H\) are even or exactly half are even.
6. Let \(G\) be a group of order 360 having a maximal subgroup isomorphic to \(A_5\). Prove that \(G \cong A_6\).

5. Project. (Simplicity of \(A_n\)) Prove that \(A_n\) is simple for \(n \geq 5\), following the steps and hints given.
   (a) Show that \(A_n\) contains every 3-cycle if \(n \geq 3\).
   (b) Show \(A_n\) is generated by the 3-cycles for \(n \geq 3\). [Hint. Note that \((a, c)(a, b) = (a, b, c)\) and \((a, b)(c, d) = (a, c, b)(a, c, d)\).]
(c) Let $r$ and $s$ be fixed elements of $\{1, 2, \ldots, n\}$ for $n \geq 3$. Show that $A_n$ is generated by the $n$ “special” 3-cycles of the form $(r, s, i)$ for $1 \leq i \leq n$. \textbf{[Hint.] Show} every 3-cycle is the product of “special” 3-cycles by computing

\[(r, s, i)^2, (r, s, j)(r, s, i)^2, (r, s, j)^2(r, s, i) \text{ and } (r, s, i)^2(r, s, k)(r, s, j)^2(r, s, i).\]

Observe that these products give all possible types of 3-cycles.

(d) Let $N$ be a normal subgroup of $A_n$ for $n \geq 3$. Show that if $N$ contains a 3-cycle, then $N = A_n$. \textbf{[Hint.] Show} that $(r, s, i) \in N$ implies that $(r, s, j) \in N$ for $j = 1, 2, \ldots, n$ by computing

\[(r, s, i)(i, j)(r, s, i)^{-1}.\]

(e) Let $N$ be a nontrivial normal subgroup of $A_n$ for $n \geq 5$. Show that one of the following cases must hold, and conclude in each case that $N = A_n$.

\textbf{Case 1.} $N$ contains a 3-cycle.

\textbf{Case 2.} $N$ contains a product of disjoint cycles, at least one of which has length greater than 3. \textbf{[Hint.] Suppose} $N$ contains the disjoint product $\sigma = \mu(a_1, a_2, \ldots, a_r)$. Show that $\sigma^{-1}(a_1, a_2, a_3)\sigma(a_1, a_2, a_3)^{-1}$ is in $N$, and compute it.

\textbf{Case 3.} $N$ contains a disjoint product of the form $\sigma = \mu(a_4, a_5, a_6)(a_1, a_2, a_3)$. \textbf{[Hint.] Show} that $\sigma^{-1}(a_1, a_2, a_3)\sigma(a_1, a_2, a_3)^{-1}$ is in $N$, and compute it.

\textbf{Case 4.} $N$ contains a disjoint product of the form $\sigma = \mu(a_1, a_2, a_3)$ where $\mu$ is a product of disjoint 2-cycles. \textbf{[Hint.] Show} $\sigma^2 \in N$ and compute it.

\textbf{Case 5.} $N$ contains a disjoint product of the form $\sigma = \mu(a_3, a_4)(a_1, a_2)$, where $\mu$ is a product of an even number of disjoint 2-cycles.

\textbf{[Hint.] Show} that $\sigma^{-1}(a_1, a_2, a_3)\sigma(a_1, a_2, a_3)^{-1}$ is in $N$, and compute it to deduce that $\sigma = (a_2, a_4)(a_1, a_3)$ is in $N$. Using $n \geq 5$ for the first time, find $i \in \{1, 2, \ldots, n\}$, where $i \neq a_1, a_2, a_3, a_4$. Let $\beta = (a_1, a_3, i)$. Show that $\beta^{-1}\alpha\beta \in N$, and compute it.

\section*{6. Project. (Wilson)} Let $p$ be a prime. Taking $G = S_p$ in the proof of Cauchy’s theorem (Theorem 1.4.19), we see that the set $\{\sigma \in S_p : \sigma^p = 1\}$ is of cardinality a multiple of $p$. Count the number of elements in this set and deduce that $(p - 1)! = -1 \pmod p$. This provides another proof of Wilson’s theorem.

\section*{1.7 Sylow Theorems}

We know the order of a subgroup of a finite group $G$ must divide $|G|$. If $|G|$ is cyclic (even only abelian), then there exist subgroups of every order dividing $|G|$. A natural question is: If $k$ divides $|G|$ is there always a subgroup of $G$ of order $k$? A little experimenting shows that this is not so. For example, the alternating group $A_4$, whose order is 12, contain no subgroup of order six. Moreover, $A_n$ for $n \geq 5$ is simple, that is, contains no normal subgroup $\neq A_n$. Since any subgroup of index two is normal, it follows that $A_n$, $n \geq 5$, contains no subgroup of order $n!/4$.

\subsection*{1.7.1 Sylow $p$-subgroups}

The main positive result of the type we are discussing was discovered by Sylow. Its proof provides us another application of action of a group on a set. Unless specified otherwise, $p$ denotes a prime.

\textbf{1.7.1. Definition.} A group $G$ is said to be a $p$-\textbf{group} if $|a|$ is a power of $p$ for all $a \in G$.

\textbf{1.7.2. Example.} The group $\mathbb{Z}_{p^n}$ is a $p$-group. If $X$ is a set, then $(P(X), \Delta)$ is a 2-group.

Since we mainly study finite groups, the following corollary will be useful. Lagrange theorem and Cauchy theorem imply each direction, respectively.
1.7.3. Corollary. Let $G$ be a finite group. Then $G$ is a $p$-group $\iff |G|$ is a power of $p$.

1.7.4. Remark. Let $P$ and $Q$ be subgroups of $G$. If $P$ is a $p$-group and $Q$ is a $q$-group, where $p$ and $q$ are distinct primes, then $P \cap Q = \{e\}$.

1.7.5. Theorem. Let $G$ be a finite $p$-group and $|G| > 1$. Then the following statements hold.
1. $|Z(G)| > 1$.
2. If $|G| = p^2$, then $G$ is abelian.

Proof. By Corollary 1.7.3, $|G| = p^l$ for some $l \in \mathbb{N}$. Recall from Corollary 1.4.14 that

$$|G| = |Z(G)| + \sum_{i=1}^{s} |\{gx_i g^{-1} : g \in G\}| = |Z(G)| + \sum_{i=1}^{s} |G : C_G(x_i)|,$$

where $x_1, \ldots, x_s$ represent the conjugacy classes of $G$ which contains more than one element. Since $|G : C_G(x_i)| = |G|/|C_G(x_i)| > 1$ for all $i$ and $|G| = p^l$, $p$ divides $|\{gx_i g^{-1} : g \in G\}|$ for all $i \in \{1, 2, \ldots, s\}$. Hence, $p \mid |Z(G)|$, so $|Z(G)| > 1$. This proves (1). For the second part, assume that $|G| = p^2$. We know that $Z(G)$ is a normal subgroup of $G$ and $|Z(G)| > 1$. By Lagrange Theorem, $|Z(G)| = p$ or $|Z(G)| = p^2$. If $|Z(G)| = p^2$, we have $Z(G) = G$ and so $G$ is abelian. Suppose that $|Z(G)| = p$. Then $G/Z(G)$ is of order $p$ and so a cyclic group. This implies that $G$ is abelian. \hfill \Box

Because all subgroups of a $p$-group have $p$-power index, the length of an orbit under an action by a $p$-group is a multiple of $p$ unless the point is a fixed point, when its orbit has length one. This leads to an important congruence modulo $p$ when a $p$-group is acting.

1.7.6. Lemma. [Fixed Point Congruence] Let $G$ be a finite $p$-group. If $G$ acts on a finite set $X$ and $X_0 = \{x \in X : gx = x \text{ for all } g \in G\}$, then $|X_0| \equiv |X| \mod p$. Here, $X_0$ is called the set of fixed points.

Proof. We observe first that $X_0 = \{x \in X : |G \cdot x| = 1\}$. Let $x_1, \ldots, x_s$ represent the orbits of $X$ which contains more than one element. Then

$$|X| = |X_0| + \sum_{i=1}^{s} |G \cdot x_i|.$$ 

By Orbit-Stabilizer Theorem, for each $i \in \{1, 2, \ldots, s\}$, $1 < |G \cdot x_i| = |G : \text{Stab}_G x_i|$ which is divisible by $p$. Hence, $|X_0| \equiv |X| \mod p$ as desired. \hfill \Box

1.7.7. Lemma. Let $G$ be a finite group and $H, P \leq G$. If $H$ is a $p$-group, then
1. $|\{xP : x \in G \text{ and } H \subseteq xPx^{-1}\}| \equiv |G : P| \mod p$,
2. $|N_G(H) : H| \equiv |G : H| \mod p$, and
3. if $p \mid |G : H|$, then $p \mid |N_G(H) : H|$ and $N_G(H) \neq H$.

Proof. Let $X = \{xP : x \in G\}$ and let $H$ act on $X$ by $h \cdot xP = hxP$ for all $x \in G$ and $h \in H$. Clearly, $|X| = |G : P|$ and

$$X_0 = \{xP : x \in G \text{ and } hxP = xP \text{ for all } h \in H\} = \{xP : x \in G \text{ and } x^{-1}hx \in P \text{ for all } h \in H\} = \{xP : x \in G \text{ and } x^{-1}hx \subseteq P\} = \{xP : x \in G \text{ and } H \subseteq xPx^{-1}\},$$
so \( \{ XP : x \in G \text{ and } H \subseteq xP x^{-1} \} \equiv [G : P] \mod p \) by Lemma 1.7.6. Furthermore, if \( P = H \), we have
\[
X_0 = \{ xH : x \in G \text{ and } H \subseteq xH x^{-1} \}.
\]
Since \( \forall x \in G, |xH x^{-1}| = |H| \) and \( H \) is finite, we have
\[
|X_0| = |\{ xH : x \in G \text{ and } x^{-1} H x = H \}| = |\{ xH : x \in N_G(H) \}| = |N_G(H) : H|,
\]
so \( |N_G(H) : H| \equiv [G : H] \mod p \). The final result clearly follows from (2).

We now discuss three theorems due to Sylow. The first theorem shows the existence of a maximal \( p \)-subgroup of a finite group \( G \).

### 1.7.8. Theorem. [First Sylow Theorem] Let \( G \) be a group of order \( p^n m \) where \( n \geq 1 \) and \( p \) does not divide \( m \). Then the following statements hold.
1. \( G \) contains a subgroup of order \( p^i \) for all \( 1 \leq i \leq n \).
2. For each \( i \), where \( 1 \leq i < n \), every subgroup \( H \) of \( G \) of order \( p^i \) is a normal subgroup of a subgroup of order \( p^{i+1} \).

### Proof. Since \( p \) divides \( |G| \), by Cauchy theorem, \( G \) has a subgroup \( H_1 \) of order \( p \). Assume that \( k \in \{1, 2, \ldots, n - 1\} \) and \( G \) has a subgroup \( H_k \) of order \( p^k \). Then the index \( [G : H_k] = p^{n-k} m \) and \( n - k \geq 1 \). By Lemma 1.7.7, \( p \) divides \( [N_G(H_k) : H_k] = |N_G(H_k) / H_k| \). Again, by Cauchy theorem, \( N_G(H_k) / H_k \) has a subgroup \( \mathcal{H} \) of order \( p \). By the Third Isomorphism Theorem, \( \mathcal{H} = H_{k+1} / H_k \) for some subgroup \( H_{k+1} \) of \( N_G(H_k) \) containing \( H_k \). Moreover, \( H_k \triangleleft H_{k+1} \) and \( |H_{k+1}| = |\mathcal{H}| |H_k| = p^{k+1} \). Hence, there are subgroups \( H_1, H_2, \ldots, H_{n} \) of \( G \) such that \( |H_i| = p^i \) for \( i = 1, 2, \ldots, n \) and \( H_1 \triangleleft H_2 \triangleleft \cdots \triangleleft H_n \).

### 1.7.9. Definition. A maximal \( p \)-subgroup of a group \( G \) is called a Sylow \( p \)-subgroup of \( G \).

By Zorn’s lemma, we have the following statements.
1. A Sylow \( p \)-subgroup of a group \( G \) always exists and it may be trivial.
2. Every \( p \)-subgroup of a group \( G \) is contained in a Sylow \( p \)-subgroup of \( G \).

By Corollary 1.7.3 and Theorem 1.7.8, if \( G \) is a finite group and \( p \) is a prime such that \( p \) divides \( |G| \), then \( G \) has a Sylow \( p \)-subgroup of order \( p^{|G|} |G| \). (Here, \( p^{|G|} |G| \) means \( n \) is the highest power of \( p \) dividing \( |G| \).) That is, \( [G : P] \) is not divisible by \( p \). Moreover, we have:

### 1.7.10. Corollary. Let \( G \) be a group of order \( p^n m \) where \( n \geq 1 \) and \( p \) does not divide \( m \).
1. \( G \) has a Sylow \( p \)-subgroup of order \( p^n \).
2. For \( H < G \), \( H \) is a Sylow \( p \)-subgroup of \( G \) if and only if \( |H| = p^n \).
3. Every conjugate of a Sylow \( p \)-subgroup of \( G \) is a Sylow \( p \)-subgroup of \( G \).
4. If \( P \) is the only Sylow \( p \)-subgroup of \( G \), then \( P \) is normal in \( G \).

### Proof. (1) and (2) follow from the definition and the above discussion. Since a conjugate of a subgroup of \( G \) is of the same order as the subgroup, (2) implies (3). Finally, (4) follows from (3).

The second and third Sylow theorems determine all Sylow \( p \)-subgroups and possible numbers of Sylow \( p \)-subgroups, respectively. Also, they give the converse of the above results.

### 1.7.11. Theorem. [Second Sylow Theorem] Let \( G \) be a finite group.
1. If \( P \) is a Sylow \( p \)-subgroup of \( G \) and \( H \) is a \( p \)-subgroup of \( G \), then \( H \subseteq xP x^{-1} \) for some \( x \in G \).
2. Any two Sylow \( p \)-subgroups of \( G \) are conjugate.
Proof. By Lemma 1.7.7 \(|\{xP : x \in G \text{ and } H \subseteq xPx^{-1}\}| \equiv |G : P| \mod p. Since P is a Sylow
p-subgroup of G, \(p \nmid |G : P|\), so \(\{xP : x \in G \text{ and } H \subseteq xPx^{-1}\} \neq \emptyset\). Thus, there exists an \(x \in G\)
such that \(H \subseteq xPx^{-1}\).

Next, we let \(P_1\) and \(P_2\) be Sylow p-subgroups of G. Then there exists an \(x \in G\) such that
\(P_1 \subseteq xP_2x^{-1}\). But \(|P_1| = |P_2| = |xP_2x^{-1}|\) and G is finite, \(P_1 = xP_2x^{-1}\). \(\square\)

1.7.12. Corollary. Let G be a finite group and let P be a Sylow p-subgroup of G.
1. \(\{xP^{-1} : x \in G\}\) is the set of all Sylow p-subgroups of G.
2. The number of Sylow p-subgroups of G is \(|G : NG(P)|\) and it divides \(|G : P|\) and \(|G|\).
3. P is normal in G \(\iff P\) is the only one Sylow p-subgroup of G.

For a finite group G and a prime \(p\) divides \(|G|\), we write \(n_p(G)\) for the number of Sylow p-subgroups of G.

1.7.13. Theorem. [Third Sylow Theorem] If G is a finite group and a prime \(p\) divides \(|G|\), then
\[n_p(G) \equiv 1 \mod p.\]

Proof. Let P be a Sylow p-subgroup of G. Then the set \(X = \{xP^{-1} : x \in G\}\) consists of all
Sylow p-subgroups of G. Let P act on X by conjugation, namely, \((g, xP^{-1}) \mapsto gxP^{-1}g^{-1}\) for
all \(g \in P\) and \(x \in G\). Since \(gp^g^{-1} = P\) for all \(g \in P, P \in X_G\). Let \(Q \in X_G\). Then \(gQg^{-1} = Q\)
for all \(g \in P\), so \(P \subseteq NG(Q)\). Since P and Q are Sylow p-subgroups of \(NG(Q)\) and Q is normal
in \(NG(Q)\), \(P = Q\) by the uniqueness of normal Sylow p-subgroup. This proves \(X_0 = \{P\}\).
By Lemma 1.7.6, we have \(n_p(G) = |X| \equiv |X_0| = 1 \mod p\) as desired. \(\square\)

1.7.2 Applications of Sylow Theorems

Here, we present some applications of Sylow theorems on a finite group. The proofs use basic
properties of subgroups, quotient groups, cyclic groups and symmetric groups studied previously.
We shall see many techniques in group theory in this subsection.

1.7.14. Theorem. Let G be a finite group. If P is a Sylow p-subgroup of G, then
\[NG(NG(P)) = NG(P).\]

Proof. Since \(P \triangleleft NG(P)\), P is the only Sylow p-subgroup of \(NG(P)\). Let \(x \in NG(NG(P))\). Then
\(xNG(P)x^{-1} = NG(P)\). Since \(P \subseteq NG(P)\), \(xPx^{-1} \subseteq NG(P)\). Thus, \(xPx^{-1} = P\) since \(xPx^{-1}\) is a
Sylow p-subgroup of G. Hence, \(x \in NG(P)\). \(\square\)

1.7.15. Theorem. [Group of Order pq] Let G be a group of order pq where \(p\) and \(q\) are primes
and \(p < q\). Then G is a cyclic group, or G has q Sylow p-subgroups and \(p \mid (q - 1)\).

Proof. Since the number of Sylow p-subgroups divides \(|G| = pq\), it is 1, \(p\), \(q\) or \(pq\). But this number
is \(\equiv 1 \mod p\), so it is 1 or \(q\). If G has \(q\) Sylow p-subgroups, then we are done. Assume that G
has only one Sylow p-subgroup, say P. Then P is normal in G. Consider the number of Sylow
q-subgroups of G. It is again 1, \(p\), \(q\) or \(pq\), and \(\equiv 1 \mod q\), so the only possibility is 1 since \(p < q\).
Thus, G also has a unique Sylow q-subgroup, say Q, and Q is normal in G. Since the orders of
P and Q are prime, both P and Q are cyclic. Let a and b be generators of P and Q, respectively.
Note that \(aba^{-1}b^{-1} \in P \cap Q = \{e\}\). Thus, \(ab = ba\), so \(|ab| = pq = |G|\). Hence, G = \(\langle ab \rangle\). \(\square\)
1.7.16. Remark. Theorem 1.7.15 demonstrates the power of the Sylow theorems in classifying the finite groups whose orders have small numbers of prime factor. Results along this lines of this theorem exist for groups of order \( p^aqp2q^2p^3p^4 \), where \( p < q \) are primes.

1.7.17. Example. There can be no simple groups of order 200 and of order 280.

Proof. Let \( H \) be a group of order 200. Let \( P \) be a Sylow 5-subgroup of \( H \). Then \( n_5(H) \) divides \([H : P] = 8\) and \( n_5(H) \equiv 1 \pmod{5} \), so \( n_5(H) = 1 \). Hence, \( P \) is normal in \( H \).

Next, let \( G \) be a group of order 280. By Corollary 1.7.12 and Theorem 1.7.13, we have \( n_2(G) = 1, 5, 7 \) or 35, \( n_5(G) = 1 \) or 56 and \( n_7(G) = 1 \) or 8. If \( n_5(G) = 1 \) or \( n_7(G) = 1 \), we are done. Assume that \( n_5(G) = 56 \) and \( n_7(G) = 8 \). Then we have \( 56 \cdot 4 = 224 \) elements of order 5, and \( 8 \cdot 6 = 48 \) elements of order 7. Hence, \( G \) has a unique Sylow 2-subgroup.

1.7.18. Example. Let \( G \) be a group of order 30. Then

1. Either a Sylow 3-subgroup or a Sylow 5-subgroup is normal in \( G \).
2. \( G \) has a normal subgroup of order 15.
3. Both a Sylow 3-subgroup and a Sylow 5-subgroup are normal in \( G \).

Proof. Assume that neither a Sylow 3-subgroup nor a Sylow 5-subgroup are normal in \( G \). By Corollary 1.7.12, \( n_3(G) \) and \( n_5(G) \) are more than one and are factors of \( |G| \). By Third Sylow Theorem, \( n_3(G) \geq 10 \) and \( n_5(G) \geq 6 \), so \( G \) contains at least 20 elements of order three and at least 24 elements of order five. This exceeds the number of elements of \( G \), a contradiction. Thus, we have (1). Now, let \( P_3 \) and \( P_5 \) be a Sylow 3-subgroup and a Sylow 5-subgroup of \( G \), respectively. By (1), we see that \( P_3 \) or \( P_5 \) is normal in \( G \), so \( P_3P_5 \) is a subgroup of \( G \). Since \( P_3 \cap P_5 = \{e\} \), \( |P_3P_5| = 15 \), so the index \( [G : P_3P_5] \) is two. Hence, \( P_3P_5 \) is normal in \( G \). This proves (2).

Finally, we assume that \( P_3 \) is normal while \( P_5 \) is not. Thus, \( G \) has two elements of order three at least 24 elements of order five. By Theorem 1.7.15, \( P_3P_5 \) is cyclic, so \( G \) has \( \phi(15) = 8 \) elements of order 15. Hence, \( G \) contains more than 30 elements, a contradiction. On the other hand, we assume that \( P_3 \) is normal while \( P_5 \) is not. Thus, \( G \) has four elements of order five at least 20 elements of order three. Again, \( G \) also contains 8 elements of order 15. This leads to a contradiction, so \( P_3 \) and \( P_5 \) are normal in \( G \) as desired.

1.7.19. Example. Every group \( G \) of order 12 that is not isomorphic with \( A_4 \) contains an element of order 6.

Proof. If \( A \) is a Sylow 3-subgroup, then \( A = \langle a \rangle \) and \( |a| = 3 \). Let \( G \) act on \( \{A, xA, x_2A, x_4A\} \) by \( (g, xA) \rightarrow gxA \). This action induces a homomorphism \( \theta : G \rightarrow S_4 \) whose kernel \( K \) is a subgroup of \( A \). Then \( K = \{e\} \) or \( K = A \). If \( K = \{e\} \), then \( G \) is isomorphic to a subgroup of \( S_4 \) of order 12, so \( G \cong A_4 \) which is excluded by hypothesis. Thus, \( A = K \) is normal in \( G \) which implies that \( A \) is a unique Sylow 3-subgroup of \( G \). Hence, \( a \) and \( a^2 \) are only two elements of order 3 in \( G \). Since \( |G : CG(a)| \) is the number of conjugates of \( a \) which is 1 or 2, \( |CG(a)| = 12 \) or 6, so there is a \( b \in CG(a) \) of order two. Since \( ab = ba \), \( |ab| = 6 \).

1.7.20. Example. Recall that \( V_4 = \{(1), (12)(34), (13)(24), (14)(23)\} \) is a normal subgroup of \( A_4 \). Since \( |A_4| = 12 = 2^2 \cdot 3 \), \( V_4 \) is the unique Sylow 2-subgroup of \( A_4 \). Moreover \( V_4 \) has three subgroups of order two, namely \( \langle (12)(34) \rangle, \langle (13)(24) \rangle \) and \( \langle (14)(23) \rangle \). Next, we analyze the Sylow 3-subgroups of \( A_4 \). They are cyclic subgroups of order three generated by a 3-cycle. Note that there are eight 3-cycles in \( A_4 \), so we have four subgroups of order three, which are \( \langle (123) \rangle, \langle (124) \rangle, \langle (134) \rangle \) and \( \langle (234) \rangle \). By Exercises 1.6, \( A_4 \) has no subgroup of order six. Hence, the diagram below
shows all subgroups of $A_4$.

1.7. Exercises. 1. If $G$ is a finite $p$-group where $p$ is a prime, $N$ is normal in $G$ and $N \neq \{e\}$, prove that $N \cap Z(G) \neq \{e\}$.

2. Prove that if $|G| = pn$ with $p > n$, $p$ is a prime, and $H$ is a subgroup of $G$ of order $p$, then $H \triangleleft G$.

3. Let $p$ be the smallest prime dividing the order of a finite group $G$. Show that any subgroup $H$ of $G$ of index $p$ is normal.

4. Let $G$ be a group of order $p^n$ where $p$ is a prime and $n \in \mathbb{N}$. Prove that there exist normal subgroups $N_1, \ldots, N_n$ of $G$ such that $N_1 < N_2 < \cdots < N_n$ with $|N_i| = p^i$ for all $i \in \{1, 2, \ldots, n\}$.

5. Let $G$ be a group, $M \triangleleft G$ and $N \triangleleft G$. Prove the following statements.
   (a) If $M \cap N = \{e\}$, then $xy = yx$ for all $x \in M$ and $y \in N$.
   (b) If $M$ and $N$ are finite cyclic subgroups of $G$ and $\gcd(|M|, |N|) = 1$, then $MN$ is a cyclic subgroup of $G$ of order $|M||N|$.

6. Let $P$ be a Sylow $p$-subgroup of a finite group $G$ and $N$ a normal subgroup of $G$. Show that:
   (a) $P \cap N$ is a Sylow $p$-subgroup of $N$.
   (b) $PN/N$ is a Sylow $p$-subgroup of $G/N$.

7. Show that there are no simple groups of order 148 or of order 56.

8. How many elements of order 7 are there in a simple group of order 168?

9. Let $G$ be a group of order 153. Prove that $G$ is abelian.

10. Let $G$ be a group of order 231. Show that $n_{11}(G) = 1$ and the Sylow 11-subgroup of $G$ is contained in $Z(G)$.

11. Show that there is no non-abelian finite simple group of order less than 60. (Hint. We may focus on groups of the following orders: 24, 30, 40, 48, 54 and 56.)

12. Let $G$ be a group of order 385. Show that a Sylow 11-subgroup of $G$ is normal and a Sylow 7-subgroup of $G$ is contained in $Z(G)$.

13. Let $p$ be a prime and $P$ a Sylow $p$-subgroup of a finite group $G$. Suppose that, for all $g \in G$, if $P \neq gPg^{-1}$, then $P \cap gPg^{-1} = \{e\}$. Show that $n_p(G) \equiv 1 \mod |P|$.

14. Let $G$ be a group of order 2013. Prove that $G$ has a proper normal subgroup $N$ such that $G/N$ is cyclic.

15. (a) Let $G$ be a finite group and $N$ a normal subgroup of $G$. If $N$ contains a Sylow $p$-subgroup of $G$, prove that the number of Sylow $p$-subgroups of $N$ is the same as that of $G$ (i.e., $n_p(N) = n_p(G)$).

16. Show that if $G$ is a group of order 130, then $G$ has a normal subgroup of order 5.

16. Let $G$ be a finite group acting transitively on a finite set $X$. Let $x \in X$ and $G_x$, the stabilizer of $x$. Let $P$ be a Sylow $p$-subgroup of $G_x$. Show that the subgroup $N_G(P) = \{z \in G : zPz^{-1} = P\}$ of $G$ acts transitively on $Y = \{y \in X : hy = y \text{ for all } h \in P\}$.

7. Project. (Simple groups of small order) We have learned from the above exercise that the smallest non-abelian simple group is of order 60 by using Sylow's theorems to eliminate the groups of smaller order. Write a computer program that uses Sylow's theorems to eliminate all orders between 1 and say 1,000 (or more) for which group that cannot be simple. For any order that could have a simple group $G$, list $n_p(G)$ for all primes $p$ dividing the order.
8. Project. (Lucas’ congruence) Let \( p \) be a prime and let \( n \geq m \) be non-negative integers. Write \( n = pn' + a_0 \) and \( m = pm' + b_0 \) where \( 0 \leq a_0, b_0 \leq p - 1 \). Decompose \( \{1, 2, \ldots, n\} \) into a union of \( p \) blocks on \( n' \) consecutive integers, from 1 to \( pn' \), followed by a final block of length \( a_0 \). That is, let

\[
A_i = \{in' + 1, in' + 2, \ldots, (i + 1)n'\}
\]

for \( 0 \leq i \leq p - 1 \), so

\[\{1, 2, \ldots, n\} = A_0 \cup A_1 \cup \cdots \cup A_{p-1} \cup \{pn' + 1, pn' + 2, \ldots, pn' + a_0\}.\]

For \( 1 \leq t \leq n' \), let \( \sigma_t \) be the \( p \)-cycle

\[
\sigma_t = (t, n' + t, 2n' + t, \ldots, (p - 1)n' + t).
\]

This cycle cyclically permutes the numbers in \( A_0, A_1, \ldots, A_{p-1} \) that are \( \equiv t \pmod{n'} \). The \( \sigma_t \)'s for different \( t \) are disjoint, so they commute. Set \( \sigma = \sigma_1 \sigma_2 \cdots \sigma_n \). Then \( \sigma \) has order \( p \) as a permutation of \( \{1, 2, \ldots, n\} \) (fixing all numbers above \( pn' \)). Let \( X \) be the set of \( m \)-element subsets of \( \{1, 2, \ldots, n\} \). Then \( |X| = \binom{n}{m} \). Let the group \( \langle \sigma \rangle \) act on \( X \).

(a) Show that the number of fixed points of this action is \( \binom{a_0}{m} \binom{n'}{m'} \pmod{p} \). Deduce, by Lemma 1.7.6, that

\[\binom{a_0}{m} = \binom{a_0}{m'} \pmod{p} \]

(b) Prove Lucas’ congruence: if \( n = a_0 + a_1p + a_2p^2 + \cdots + a_kp^k \) and \( m = b_0 + b_1p + b_2p^2 + \cdots + b_kp^k \) with \( 0 \leq a_i, b_i \leq p - 1 \), then

\[\binom{n}{m} \equiv \binom{a_0}{b_0} \binom{a_1}{b_1} \cdots \binom{a_k}{b_k} \pmod{p}.
\]

1.8 Finite Abelian Groups

The study of finite non-abelian groups is complicated as we have learned that the Sylow theorems give us some important information about them. This section gives us complete information about all finite abelian groups. We start with formal definitions of the direct product of groups.

1.8.1. Definition. Let \( A \) and \( B \) be groups. The direct product of \( A \) and \( B \) is defined as:

1. a set \( A \times B = \{(a, b) : a \in A \text{ and } b \in B\} \) is the Cartesian product of \( A \) and \( B \),
2. multiplication is coordinatewise, namely, \( (a, b)(c, d) = (ac, bd) \).

More generally, if \( \{A_i : i \in I\} \) is a family of groups, then \( \prod_{i \in I} A_i \) is a group with coordinatewise multiplication. It is called the direct product of the groups \( A_i \). The subgroup

\[\prod_{i \in I}^w A_i = \left\{(a_i) \in \prod_{i \in I} A_i : a_i = e \text{ for all but finitely many } i\right\}\]

of \( \prod_{i \in I} A_i \) is called the external weak direct product of the groups \( A_i \). Note that it is normal in \( \prod_{i \in I} A_i \). In case \( A_i \) are all additive abelian groups, we may write \( \sum_{i \in I} A_i \) for \( \prod_{i \in I} A_i \) and \( \sum_{i \in I} A_i \) for \( \prod_{i \in I} A_i \).

Let \( G \) be a group. It is easy to show that:

1. If \( N_1, \ldots, N_m \) are normal subgroups of \( G \), then

\[N_1N_2 \cdots N_m = \langle N_1 \cup N_2 \cup \cdots \cup N_m \rangle.
\]

2. If \( \{N_i : i \in I\} \) is a family of normal subgroups of \( G \), then

\[\left\langle \bigcup_{i \in I} N_i \right\rangle = \bigcup_{i_1, \ldots, i_m \in I, m \in \mathbb{N}} N_{i_1} \cdots N_{i_m}.
\]
1.8.2. Theorem. Let \( \{N_i : i \in I\} \) be a family of normal subgroups of \( G \) such that
1. \( G = \bigcup_{i \in I} N_i \) and 2. \( \forall k \in I, N_k \cap \left( \bigcup_{i \in I \setminus \{k\}} N_i \right) = \{e\} \).
Then \( G \cong \prod_{i \in I} N_i \).

1.8.3. Definition. The group \( G \) satisfies conditions of Theorem 1.8.2 is called the internal weak direct product of \( \{N_i : i \in I\} \) and we write \( G = \prod_{i \in I}^w N_i \). If \( G \) is additive abelian, then \( G \) is called the internal direct sum of \( \{N_i : i \in I\} \) and we write \( G = \bigoplus_{i \in I} N_i \).

1.8.4. Corollary. Let \( N_1, N_2, \ldots, N_m \) be normal subgroups of \( G \). If \( G = N_1 \cdots N_m \) and \( N_k \cap (N_1 \cdots N_{k-1} N_{k+1} \cdots N_m) = \{e\} \) for all \( k \in \{1, \ldots, m\} \), then \( G \cong N_1 \times \cdots \times N_m \).

Proof of Theorem 1.8.2. From (2), for each \( i, j \in I \) with \( i \neq j \), we have \( N_i \cap N_j = \{e\} \), this implies that \( xy = yx \) for all \( x \in N_i \) and \( y \in N_j \) because \( N_i \) and \( N_j \) are normal in \( G \).
Define \( \varphi : \prod_{i \in I}^w N_i \to G \) by
\[
\varphi(\{a_i\}) = \prod_{i \in I} a_i
\]
which is a finite product since \( a_i = e \) for all but a finite number of \( i \in I \) and it is a well defined homomorphism by the previous observation. To show that this map is onto, let \( x \in G \). Since \( G \) is generated by \( \bigcup_{i \in I} N_i \), \( G = \bigcup_{i, j \in I, i \neq j} N_i \cdot N_j \), so there are distinct \( k_1, \ldots, k_l \in I \) and \( a_1 \in N_{k_1}, \ldots, a_l \in N_{k_l} \) such that \( x = a_{k_1} \cdots a_{k_l} \). Let \( \{x_i\} \) in \( \prod_{i \in I}^w N_i \) be defined by \( x_i = a_i \) if \( i \in \{k_1, \ldots, k_l\} \) and \( x_i = e \) for other \( i \). Then \( \varphi(\{x_i\}) = \prod_{i \in I} x_i = a_{k_1} \cdots a_{k_l} = x \) as required.
Finally, we show that \( \varphi \) is injective. Let \( \{a_i\} \in \prod_{i \in I}^w N_i \) be such that \( \prod_{i \in I} a_i = e \). Then for each \( k \in I \), \( a_k^{-1} = \prod_{i \in I \setminus \{k\}} a_i \) is in \( N_k \cap \bigcup_{i \in I \setminus \{k\}} N_i \). This implies that \( a_k = e \) for all \( k \in I \), and hence \( \varphi \) is an isomorphism. \( \square \)

The proof of injectivity above also implies the following theorems.

1.8.5. Theorem. Let \( N_1, N_2, \ldots, N_m \) be normal subgroups of \( G \). Then the following statements are equivalent.
(i) \( G \) is the internal weak direct product of \( N_1, \ldots, N_m \).
(ii) \( \forall x \in G, \exists a_1 \in N_1, \ldots, a_m \in N_m, x = a_1 \cdots a_m \).

1.8.6. Theorem. Let \( \{N_i : i \in I\} \) be a family of normal subgroups of a group \( G \). Then the following statements are equivalent.
(i) \( G \) is the internal weak direct product of \( \{N_i : i \in I\} \).
(ii) \( \forall x \in G \setminus \{e\}, \exists i_1, \ldots, i_m \in I, \exists a_{i_1} \in N_{i_1} \setminus \{e\}, \ldots, a_{i_m} \in N_{i_m} \setminus \{e\}, x = a_{i_1} \cdots a_{i_m} \).

1.8.7. Corollary. [Internal Direct Product] Let \( G \) be a group. Suppose that \( A \) and \( B \) are normal subgroups of \( G \) such that
1. \( A \cap B = \{e\} \)  2. \( AB = G \)  3. \( \forall a \in A, b \in B, ab = ba \).
Then \( G \cong A \times B \). In this case, we say that \( G \) is the internal direct product of \( A \) and \( B \).

An application of the first isomorphism theorem with the natural map gives the next results.

1.8.8. Theorem. Let \( \{G_i : i \in I\} \) be a family of groups, and for \( i \in I \), let \( N_i \) be normal in \( G_i \). Then \( \prod_{i \in I} N_i \) is normal in \( \prod_{i \in I} G_i \) and \( \prod_{i \in I} G_i / \prod_{i \in I} N_i \cong \prod_{i \in I} (G_i / N_i) \). Similarly, \( \prod_{i \in I}^w N_i \) is normal in \( \prod_{i \in I} G_i \) and \( \prod_{i \in I}^w G_i / \prod_{i \in I}^w N_i \cong \prod_{i \in I}^w (G_i / N_i) \).
1.8.9. Corollary. Let $G_1, \ldots, G_m$ be abelian groups, and for $1 \leq j \leq m$, let $H_j$ be a subgroup of $G_j$. Then $(G_1 \oplus \cdots \oplus G_n)/(H_1 \oplus \cdots \oplus H_n) \cong (G_1/H_1) \oplus \cdots \oplus (G_n/H_n)$.

Next, we study the structure of a finite abelian group. Results on elements of finite order are presented in the next theorem and we recall the Chinese Remainder Theorem in a group theoretic language. Their proof are routine and left as exercises.

1.8.10. Theorem. Let $A$ be an abelian group and $n \in \mathbb{N}$. Then the following statements hold.
1. The mapping $\varphi_n : A \to A$ defined by $\varphi_n(a) = a^n$ is a group homomorphism.
2. $A^n = \{a^n : a \in A\} = \text{im} \varphi_n$ is a subgroup of $A$.
3. $A(n) = \{a \in A : a^n = e\} = \ker \varphi_n$ is a subgroup of $A$.
4. $\tau(A) = \{a \in A : \exists k \in \mathbb{N}, a^k = e\} = \bigcup_{n \in \mathbb{N}} A(n)$ is a subgroup of $A$. It is called the torsion subgroup of $A$.

1.8.11. Theorem. [Chinese Remainder Theorem] Suppose $m_1, \ldots, m_k$ are pairwise relatively prime (i.e., if $i \neq j$, then $\gcd(m_i, m_j) = 1$), and let $n_1, \ldots, n_k$ be any integers. Then there exists a unique integer $v$ modulo $m = m_1 \ldots m_k$ such that

$$v \equiv n_i \mod m_i$$

for all $i = 1, \ldots, k$.

1.8.12. Remark. The Chinese remainder theorem may be restated as: if $m_1, \ldots, m_k$ are pairwise relatively primes and $m = m_1 \ldots m_k$, then

$$\mathbb{Z}_m \cong \mathbb{Z}_{m_1} \oplus \cdots \oplus \mathbb{Z}_{m_k}.$$

1.8.13. Corollary. If $m = p_1^{n_1} \ldots p_k^{n_k}$ where $n_1, \ldots, n_k \in \mathbb{N}$ and $p_1, \ldots, p_k$ are distinct primes, then

$$\mathbb{Z}_m \cong \mathbb{Z}_{p_1^{n_1}} \oplus \cdots \oplus \mathbb{Z}_{p_k^{n_k}}.$$

We shall use the Chinese remainder theorem to prove our first lemma.

1.8.14. Lemma. Let $g \in G$ be an element such that $g^m = 1$ where $m = m_1 \ldots m_k$ and $m_i$ are pairwise relatively prime. Then there exist unique elements $g_1, \ldots, g_k$ of $G$ satisfying the following conditions:
(a) $g_i^{m_i} = e$ for all $i \in \{1, \ldots, k\}$
(b) $g_1, \ldots, g_k$ commute pairwise and
(c) $g = g_1 \cdots g_k$.

Proof. First we show existence, then uniqueness. The $g_i$ are in fact powers of $g$.
Existence: By the Chinese Remainder Theorem, choose $v_1, \ldots, v_k$ satisfying

$$v_i \equiv 1 \mod m_i \text{ and } v_i \equiv 0 \mod m/m_i.$$ 

For each $i$, let $v_i = \lambda_i(m/m_i)$ for some $\lambda_i \in \mathbb{Z}$ and set $g_i = g^{v_i}$. Then we have
(i) $g_i^{m_i} = g_i^{v_i m_i} = g_i^{\lambda_i(m/m_i)m_i} = g_i^{\lambda_i m}$ and
(ii) $g_1, \ldots, g_k$ are powers of $g$ and hence commute pairwise.
(iii) Note that \( v_1 + \cdots + v_k - 1 \equiv 0 \mod m_i \) for \( i = 1, 2, \ldots, k \), that is, \( m_i | (v_1 + \cdots + v_k - 1) \). Since \( m_1, \ldots, m_k \) are pairwise relatively prime, \( v_1 + \cdots + v_k \equiv 1 \mod m_1 \cdots m_k \), so

\[
g_1 \cdots g_k = g^{v_1} \cdots g^{v_k} = g^{v_1 + \cdots + v_k} = g.
\]

**Uniqueness:** Suppose \( g = g_1 \cdots g_k \) where \( g = g_1 \cdots g_k \) and \( g_1, \ldots, g_k \) satisfy (i), (ii) and (iii). Then for each \( i \),

\[
g_i^{v_i} = g_i^{v_1} \cdots g_i^{v_k} = g_i,
\]

that is, \( g_i = g_i^{v_i} \) is the only possibility. \( \square \)

**1.8.15. Example.** Consider \( g \in G \) with \( g^{60} = e \). Then \( m = 3 \cdot 4 \cdot 5 \), so \( v_1 = 45, v_2 = 40 \) and \( v_3 = 36 \). Thus, \( g = g^{45}g^{40}g^{36} = g_1g_2g_3 \).

In case \( g \) has order \( m = p_1^{a_1} \cdots p_k^{a_k} \) where \( m_i = p_i^{a_i} \) and \( p_1, \ldots, p_k \) are distinct primes, \( g_i \) is called the \( p_i \)-primary part of \( g \). We have the first step of our decomposition.

**1.8.16. Theorem.** Let \( A \) be a finite abelian group of order \( m = m_1 \cdots m_k \) where \( m_1, \ldots, m_k \) are pairwise relatively prime. For each \( i \in \{1, 2, \ldots, k\} \), let \( A_i = \{ g \in A : g^{m_i} = e \} \). Then \( A \cong A_1 \times \cdots \times A_k \). Moreover, \( |A_i| = m_i \) for all \( i \).

**Proof.** Define \( \phi : A_1 \times \cdots \times A_k \to A \) by \( \phi(g_1, \ldots, g_k) = g_1 \cdots g_k \). Clearly, \( \phi \) is a group homomorphism. By Lemma 1.8.14, \( \phi \) is 1-1 and onto. Finally, \( m_1 \cdots m_k = |A| = |A_1 \times \cdots \times A_k| = |A_1| \cdots |A_k| \). Let

\[
\begin{align*}
m_1 &= p_1^{u_{11}} \cdots p_{1r_1}^{u_{1r_1}} \\
m_2 &= p_2^{u_{21}} \cdots p_{2r_2}^{u_{2r_2}} \\
&\vdots \\
m_k &= p_k^{u_{k1}} \cdots p_{kr_k}^{u_{kr_k}}.
\end{align*}
\]

Here, the \( p_{ij} \) are distinct primes and \( u_{ij} \geq 1 \). Since every element of \( A_i \) satisfies \( g^{m_i} = e \), \( |A_i| \) involves only those primes occurring in \( m_i \) by Cauchy’s theorem. This forces \( |A_i| = m_i \) for all \( i \). \( \square \)

If \( m = p_1^{a_1} \cdots p_k^{a_k} \), then \( A_i \) in Theorem 1.8.16 is just the Sylow \( p_i \)-subgroup of \( A \). It now suffices to study each factor \( A_i \) which is a Sylow \( p_i \)-subgroup of \( A \). To investigate them, we shall need the following definition.

**1.8.17. Definition.** The positive integer \( n \) is an **exponent** for a group \( G \) if for each \( g \in G \), \( g^n = 1 \). In this case, \( G \) is said to have finite exponent and the least such \( n \) is called the **exponent** of \( G \).

For example, 12 is an exponent of \( \mathbb{Z}_6 \) but 6 is the exponent of \( \mathbb{Z}_6 \). We denote the exponent of a group \( G \) (if it exists) by \( \exp G \). Note that \( \mathbb{Z} \) has no exponent and \( \exp G \) divides \( |G| \). The exponent of a finite abelian \( p \)-group gives its structure as follows.

**1.8.18. Theorem.** Let \( A \) be an abelian group with \( |A| = p^n \) where \( p \) is a prime. Suppose \( A \) has the exponent \( p \), (that is, \( a^p = e \) for all \( a \in A \)). Then

\[
A \cong \underbrace{\mathbb{Z}_p \times \cdots \times \mathbb{Z}_p}_{u \text{ copies}} = (\mathbb{Z}_p)^u.
\]
Proof. Note that if \( a \in A \setminus \{ e \} \), \( \langle a \rangle \cong \mathbb{Z}_p \). For any subset \( \{a_1, \ldots, a_k\} \subseteq A \setminus \{ e \} \), we can define a group homomorphism \( \theta : \langle a_1 \rangle \times \cdots \times \langle a_k \rangle \to A \) by
\[
\theta(a_1^{i_1}, \ldots, a_k^{i_k}) = a_1^{i_1} \cdots a_k^{i_k}.
\]

We shall say that \( a_1, \ldots, a_k \) are "linearly independent" if \( \theta \) is 1-1. This is equivalent to saying: If \( \theta(a_1^{i_1}, \ldots, a_k^{i_k}) = a_1^{i_1} \cdots a_k^{i_k} = e \), then \( i_1 = \cdots = i_k = 0 \mod p \).

Now there exists subsets of \( A \setminus \{ e \} \) for which \( \theta \) is 1-1, e.g., the empty set, a singleton set. Choose a subset \( \{a_1, \ldots, a_k\} \) for which \( \theta \) is 1-1 and \( k \) is as large as possible. We claim that in this case
\[
\langle a_1 \rangle \times \cdots \times \langle a_k \rangle \xrightarrow{\theta} A
\]
is onto, and hence is an isomorphism.

To see that \( \theta \) is onto, let \( b \in A \). If \( b = e \), clearly \( b \in \text{im} \theta \). If \( b \neq e \), consider
\[
\langle a_1 \rangle \times \cdots \times \langle a_k \rangle \times \langle b \rangle \xrightarrow{\overline{\theta}} A.
\]
By the maximal choice of \( \{a_1, \ldots, a_k\} \), \( \overline{\theta} \) is not 1-1. Thus,
\[
eq \overline{\theta}(a_1^{i_1}, \ldots, a_k^{i_k}, b^j) = a_1^{i_1} \cdots a_k^{i_k} b^j
\]
and \( j \neq 0 \mod p \) since \( \theta \) is not 1-1. Hence, there is a \( \lambda \) such that \( j \lambda \equiv 1 \mod p \), so
\[e = a_1^{\lambda i_1} \cdots a_k^{\lambda i_k} b^\lambda = a_1^{\lambda i_1} \cdots a_k^{\lambda i_k} b\]
which implies \( b = a_1^{-\lambda i_1} \cdots a_k^{-\lambda i_k} \in \text{im} \theta \). Therefore, \( \theta \) is onto as claimed, and we have an isomorphism
\[
\langle a_1 \rangle \times \cdots \times \langle a_k \rangle \xrightarrow{\theta} A.
\]
Thus, \( p^k = |\langle a_1 \rangle| \cdots |\langle a_k \rangle| = |\langle a_1 \rangle \times \cdots \times \langle a_k \rangle| = |A| = p^u \), so \( k = u \) and the theorem is proved. \( \square \)

1.8.19. Remark. If we write \( A \) in Theorem 1.8.18 additively, we see that it is just a vector space over the field \( \mathbb{Z}_p \). Since \( A \) is finite, it is a finite dimensional vector space over \( \mathbb{Z}_p \). All we were doing in Theorem 1.8.18 is finding a basis for \( A \) as a vector space over \( \mathbb{Z}_p \).

1.8.20. Theorem. [Burnside Basis Theorem for Abelian \( p \)-groups] Suppose \( A \) is an abelian group of exponent \( p^k \) where \( p \) is a prime. Let \( A^p = \{ a^p : a \in A \} \). If \( H \) is a subgroup of \( A \) and \( H A^p = A \), then \( H = A \). Equivalently, if the cosets \( a^p a_1, \ldots, a^p a_k \) of \( A/A^p \) generate \( A/A^p \), then \( a_1, \ldots, a_k \) generate \( A \).

Proof. Observe that \( H A^p = A \) implies \( H^p A^{p^2} = A^p \), so
\[
A = H A^p = H(H^p A^{p^2}) = H A^{p^2}.
\]
Also, \( H A^{p^2} = A \) implies \( H^p A^{p^3} = A^p \), so
\[
A = H A^p = H(H^p A^{p^2}) = H A^{p^3}.
\]
Continue inductively, we have \( A = H A^{p^r} \) for all \( r \). But \( A^{p^k} = \{ e \} \), so \( A = H A^{p^k} = H \). This completes the proof. \( \square \)

1.8.21. Theorem. Let \( A \) be a finite abelian group of exponent \( p \) where \( p \) is a prime, and let \( H \) be a subgroup of \( A \). Then there exists a subgroup \( K \) of \( A \) such that \( H \cap K = \{ e \} \) and \( HK = A \). In other words, \( A \) is the internal direct product of \( H \) and \( K \).
Proof. Let $K$ be a subgroup of $A$ satisfying $H \cap K = \{e\}$ and among all subgroups $K$ of $A$ satisfying $H \cap K = \{e\}$, $K$ is as large as possible. We claim that $HK = A$ which proves the theorem.

For, suppose conversely that $a \in A$ and $a \notin HK$. Then $H \cap \langle K, a \rangle \neq \{e\}$ by the maximal choice of $K$, so there is a nontrivial element

$$e \neq h = ka^i \in H \cap \langle K, a \rangle$$

where $h \in H \in <k \in K$.

If $p | i$, $a^i = e$ and $h = k \in H \cap K = \{e\}$, a contradiction. If $p \nmid i$, there is a $\lambda$ with $a^{i\lambda} = a$ ($i \lambda \equiv 1 \mod p$) and then $a = a^{i\lambda} = (hk^{-1})^\lambda \in HK$, a contradiction. Hence, $HK = A$ as required. \qed

1.8.22. Remark. As with Theorem 1.8.18, the above theorem can be regarded as a statement about vector spaces over $\mathbb{Z}_p$ as follows: If $V$ is a finite dimensional vector space over $\mathbb{Z}_p$ and $U$ is a subspace, then there is a subspace $W$ such that $V = U \oplus W$.

We are now ready to prove the structure theorem for a finite abelian $p$-group.

1.8.23. Theorem. Let $A$ be a finite abelian $p$-group. Then $A$ is (isomorphic to) a direct product of cyclic groups.

Proof. We use induction on $|A|$. If $|A| = 1$, the result is clear. Now suppose $|A| = p^n > 1$. We assume inductively that any $p$-group where order is less than $p^n$ is a direct product of cyclic groups. Consider the group

$$A^p = \{a^p : a \in A\}.$$  

Claim $A \neq A^p$. For, suppose $A = A^p$. Then

$$A = A^p = A^{p^2} = \cdots = A^{p^n} = \{e\}$$

since $|A| = p^n$. As $|A| > 1$, we must have $A \neq A^p$.

Thus, $A > A^p$, so $|A| > |A^p|$ and by the inductive hypothesis, $A^p$ is the (internal) direct product of cyclic subgroup. But every element of $A^p$ has the form $a^p$. Hence, so do the generators of these cyclic factors, so there exist $a_1, \ldots, a_k \in A$ such that

$$A^p = \langle a_1^p \rangle \times \cdots \times \langle a_k^p \rangle.$$  

More precisely, the map $(a_1^{\pi_1}, \ldots, a_k^{\pi_k}) \mapsto a_1^{\pi_1} \cdots a_k^{\pi_k}$ is an isomorphism.

Now let $H = \langle a_1, \ldots, a_k \rangle$ be the subgroup of $A$ generated by $a_1, \ldots, a_k$. We claim that it is the (internal) direct product of the groups $\langle a_1 \rangle, \ldots, \langle a_k \rangle$, or that

$$\langle a_1 \rangle \times \cdots \times \langle a_k \rangle \cong H$$

$$(a_1^{i_1}, \ldots, a_k^{i_k}) \mapsto a_1^{i_1} \cdots a_k^{i_k}$$

is an isomorphism. Since $H = \langle a_1, \ldots, a_k \rangle$, $\theta$ is onto. To see that $\theta$ is 1-1, suppose

$$e = \theta(a_1^{i_1}, \ldots, a_k^{i_k}) = a_1^{i_1} \cdots a_k^{i_k}.$$  

Then $e = (a_1^{i_1} \cdots a_k^{i_k})^p = a_1^{pi_1} \cdots a_k^{pi_k}$, so

$$a_1^{pi_1} = \cdots = a_k^{pi_k} = e$$

since the map $\theta_0$ above is 1-1. Consider the integers $i_1, \ldots, i_k$. If $p | i_j$ for some of these integers, then $a_j^{pi_j} = e$ implies $a_j^p = e$. For, $(a_j^{pi_j})^{\lambda_i} = (a_1^{\lambda_i i_1})^p = a_j^p$, where $\lambda_i i_1 \equiv 1 \mod |a_i|$. Thus, $p | i_1, \ldots, i_k$, so $i_1 = p j_1, \ldots, i_k = p j_k$. But then (1.8.1) becomes

$$e = \theta(a_1^{p j_1}, \ldots, a_k^{p j_k}) = a_1^{p j_1} \cdots a_k^{p j_k}.$$  

(1.8.1)
so \( a_{p1}^{p1} = \cdots = a_k^{pk} = e \) since \( \theta_0 \) is 1-1. Hence, \( \theta \) is 1-1.

\[
\begin{tikzpicture}
    \node (A) at (0,0) {$A$};
    \node (H) at (-2,-2) {$H$};
    \node (Ap) at (2,-2) {$A_p$};
    \node (K) at (0,-4) {$\{e\}$};
    \node (HAp) at (0,-2) {$H \cap A_p$};
    \node (ApH) at (0,-3) {$A_p^p \cong H^p$};
    \draw (A) to (H); \draw (A) to (Ap); \draw (H) to (HAp); \draw (Ap) to (HAp); \draw (HAp) to (K); \draw (HAp) to (ApH); \draw (ApH) to (ApH); \draw (ApH) to (K);
\end{tikzpicture}
\]

Next, let \( A_p = \{ a \in A : a^p = e \} \). Then \( A_p \) is a finite group of exponent \( p \) and contains \( H \cap A_p \) as a subgroup. Therefore, \( H \cap A_p \) has a component in \( A_p \) by Theorem 1.8.21. More precisely, there is a subgroup \( K \) of \( A_p \) such that

(a) \( (H \cap A_p) \cap K = \{ e \} \) and (b) \( (H \cap A_p) K = A_p \).

Note that since \( K \) is a group of exponent \( p \), \( K \) is a direct product of copies of \( \mathbb{Z}_p \) by Theorem 1.8.18. Finally, we claim that

(I) \( H \cap K = \{ e \} \) and (II) \( HK = A \).

They imply that \( A \) is a direct product of \( H \) and \( K \) which are both direct products of cyclic groups.

(I) Suppose \( H \cap K \neq \{ e \} \). Thus, there is some \( x \in H \cap K \) with \( x \neq e \) and \( x^p = e \). But then \( x \in A_p \), so \( (H \cap A_p) \cap K = (H \cap K) \cap A_p \neq \{ e \} \), contradicting (i) above.

(II) Suppose \( a \in A \). Then \( a^p \in A_p = \langle a_1^p \rangle \times \cdots \times \langle a_k^p \rangle \), so \( a^p = a_{p1}^{p1} \cdots a_{pk}^{pk} = (a_1^1 \cdots a_k^k)^p = b^p \) where \( b = a_1^1 \cdots a_k^k \in H = \langle a_1, \ldots, a_k \rangle \). Thus, \( b^{-1} a \in A_p = (H \cap A_p) K \subseteq HK \) by (ii) above. Hence, \( a = b(b^{-1} a) \in HK \) and \( A = HK \) as required. This completes the proof. \( \square \)

In addition, the above decomposition is unique. Hence, we are able to count the number of non-isomorphic abelian \( p \)-groups.

1.8.24. Theorem. Suppose

\[
A = \mathbb{Z}_{p_1} \times \cdots \times \mathbb{Z}_{p_{u_1}} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_{u_2}} \times \cdots \times \mathbb{Z}_{p_m} \times \cdots \times \mathbb{Z}_{p_{u_m}}
\]

is isomorphic to

\[
B = \mathbb{Z}_{q_1} \times \cdots \times \mathbb{Z}_{q_{v_1}} \times \mathbb{Z}_{q_2} \times \cdots \times \mathbb{Z}_{q_{v_2}} \times \cdots \times \mathbb{Z}_{q_m} \times \cdots \times \mathbb{Z}_{q_{v_m}}
\]

where \( u_i, v_i \geq 1 \). Then \( u_i = v_i \) for all \( i = 1, \ldots, m \). In other words, the orders and multiplicities of the factors in a decomposition of a finite abelian \( p \)-group uniquely determine the group up to isomorphism.

Proof. Since \( A \cong B \), it follows that for any positive integer \( n \),

\[\# \text{ of solutions of } x^n = e \text{ in } A = \# \text{ of solutions of } x^n = e \text{ in } B.\]

Consider the following table.
\[
\begin{array}{|c|c|c|}
\hline
n & \text{# of solutions of } x^n = e \text{ in } A & \text{# of solutions of } x^n = e \text{ in } B \\ \hline
p & p^{u_1+v_2+u_3+\cdots+u_m} & p^{v_1+v_2+v_3+\cdots+v_m} \\ p^2 & p^{u_1+2u_2+u_3+\cdots+2u_m} & p^{v_1+2v_2+2v_3+\cdots+2v_m} \\ p^3 & p^{u_1+2u_2+3u_3+\cdots+3u_m} & p^{v_1+2v_2+3v_3+\cdots+3v_m} \\ \vdots & \vdots & \vdots \\ p^{m-1} & p^{u_1+2u_2+3u_3+\cdots+(m-1)u_{m-1}+(m-1)u_m} & p^{v_1+v_2+\cdots+(m-1)v_{m-1}+(m-1)v_m} \\ p^m & p^{u_1+2u_2+3u_3+\cdots+(m-1)u_{m-1}+mu_m} & p^{v_1+v_2+\cdots+(m-1)v_{m-1}+mv_m} \\
\hline
\end{array}
\]

Then we have
\[
\begin{align*}
&u_1 + u_2 + u_3 + \cdots + u_m = v_1 + v_2 + v_3 + \cdots + v_m \\
u_1 + 2u_2 + 2u_3 + \cdots + 2u_m = v_1 + 2v_2 + 2v_3 + \cdots + 2v_m \\
u_1 + 2u_2 + 3u_3 + \cdots + 3u_m = v_1 + 2v_2 + 3v_3 + \cdots + 3v_m \\
\vdots \\
u_1 + 2u_2 + 3u_3 + \cdots + (m-1)u_{m-1} + (m-1)u_m = v_1 + 2v_2 + 3v_3 + \cdots + (m-1)v_{m-1} + (m-1)v_m \\
u_1 + 2u_2 + 3u_3 + \cdots + (m-1)u_{m-1} + mu_m = v_1 + 2v_2 + 3v_3 + \cdots + (m-1)v_{m-1} + mv_m.
\end{align*}
\]

It is easy to see that the above equations force \( u_1 = v_1, u_2 = v_2, \ldots, u_m = v_m \) as required.

1.8.25. Theorem. Let \( p \) be any prime and \( n \) a positive integer. Then
\[
\{ r_1 \leq r_2 \leq \cdots \leq r_k \} \longleftrightarrow \mathbb{Z}_{p^{r_1}} \times \mathbb{Z}_{p^{r_2}} \times \cdots \times \mathbb{Z}_{p^{r_k}}.
\]
defines a 1-1 correspondence between partitions of \( n \) and isomorphism classes of abelian groups of order \( p^n \). In particular, the number of isomorphism classes of abelian groups of order \( p^n \) is the number of partitions of \( n \).

1.8.26. Examples. 1. Abelian groups of order \( p^3 \).

<table>
<thead>
<tr>
<th>partitions of 3</th>
<th>corresponding abelian groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>{1, 1, 1}</td>
<td>\mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p</td>
</tr>
<tr>
<td>{1, 2}</td>
<td>\mathbb{Z}<em>p \times \mathbb{Z}</em>{p^2}</td>
</tr>
<tr>
<td>{3}</td>
<td>\mathbb{Z}_{p^3}</td>
</tr>
</tbody>
</table>

2. Abelian groups of order \( p^5 \).

<table>
<thead>
<tr>
<th>partitions of 5</th>
<th>corresponding abelian groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>{1, 1, 1, 1, 1}</td>
<td>\mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p</td>
</tr>
<tr>
<td>{1, 1, 1, 2}</td>
<td>\mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}<em>p \times \mathbb{Z}</em>{p^2}</td>
</tr>
<tr>
<td>{1, 2, 2}</td>
<td>\mathbb{Z}<em>p \times \mathbb{Z}</em>{p^2} \times \mathbb{Z}_{p^2}</td>
</tr>
<tr>
<td>{1, 1, 3}</td>
<td>\mathbb{Z}_p \times \mathbb{Z}<em>p \times \mathbb{Z}</em>{p^3}</td>
</tr>
<tr>
<td>{2, 3}</td>
<td>\mathbb{Z}<em>{p^2} \times \mathbb{Z}</em>{p^3}</td>
</tr>
<tr>
<td>{1, 4}</td>
<td>\mathbb{Z}<em>p \times \mathbb{Z}</em>{p^4}</td>
</tr>
<tr>
<td>{5}</td>
<td>\mathbb{Z}_{p^5}</td>
</tr>
</tbody>
</table>

Suppose \( A \) is a finite abelian group of order \( p_1^{a_1} \cdots p_k^{a_k} \) where \( p_1, \ldots, p_k \) are distinct primes. Let
\[
A_i = \{ g \in A : g^{p_i^{a_i}} = e \}.
\]

By Theorem 1.8.16, \( A \cong A_1 \times \cdots \times A_k \) where \( |A_i| = p_i^{a_i} \). Since each \( A_i \) is a direct product of cyclic group by Theorem 1.8.24, this yields the following theorem.
1.8.27. Theorem. A finite abelian group is (isomorphic to) a direct product of cyclic groups.

1.8.28. Corollary. If $m$ is a square free integer, then every abelian group of order $m$ is cyclic.

Proof. Assume that an abelian group $A$ is of order $m = p_1 \ldots p_r$ where $p_i$ are distinct primes. Then

$$A \cong \mathbb{Z}_{p_1} \times \ldots \times \mathbb{Z}_{p_r} \cong \mathbb{Z}_m$$

by Theorem 1.8.16 and the Chinese remainder theorem, respectively. \qed

Let $A \cong A_1 \times \ldots \times A_k$ as above. It is clear that $A_i$ is the unique largest $p_i$-subgroup of $A$. Moreover, if $B$ is finite abelian, and $B \cong B_1 \times \ldots \times B_k$ where $B_i$ is a $p_i$-group, then

$$A \cong B \iff (A_1 \cong B_1 \land \ldots \land A_k \cong B_k).$$

Recall that if $m$ and $n$ are relatively prime then $\mathbb{Z}_m \times \mathbb{Z}_n \cong \mathbb{Z}_{mn}$. It follows that if $n = p_1^{v_1} \cdots p_k^{v_k}$ where $p_1, \ldots, p_k$ are distinct primes, then

$$\mathbb{Z}_{p_1^{v_1}} \times \cdots \times \mathbb{Z}_{p_k^{v_k}} \cong \mathbb{Z}_n. \quad (1.8.2)$$

This gives rise to a second way of writing a finite abelian group $A$ as a direct product of cyclic groups. Namely, let $p_1, \ldots, p_k$ be the primes dividing $|A|$, and let

$$A = A_1 \times \cdots \times A_k,$$

where $A_i$ is the $p_i$-primary part of $A$. Express each $A_i$ as a direct product of cyclic factors and assume that $t$ is the largest number of factors occurring in any $A_i$. Write

$$A_1 = \mathbb{Z}_{p_1^{v_{i1}}} \times \mathbb{Z}_{p_1^{v_{i2}}} \times \cdots \times \mathbb{Z}_{p_1^{v_{it}}},$$

$$A_2 = \mathbb{Z}_{p_2^{v_{i1}}} \times \mathbb{Z}_{p_2^{v_{i2}}} \times \cdots \times \mathbb{Z}_{p_2^{v_{it}}},$$

$$\vdots$$

$$A_k = \mathbb{Z}_{p_k^{v_{i1}}} \times \mathbb{Z}_{p_k^{v_{i2}}} \times \cdots \times \mathbb{Z}_{p_k^{v_{it}}},$$

where

$$0 \leq v_{i1} \leq v_{i2} \leq \cdots \leq v_{it} \quad (1.8.3)$$

and we have allowed (for notational convenience) some $v_{ij}$ to be zero. Let

$$n_1 = p_1^{v_{i1}} p_2^{v_{i2}} \cdots p_k^{v_{i1}},$$

$$n_2 = p_1^{v_{i2}} p_2^{v_{i2}} \cdots p_k^{v_{i2}},$$

$$\vdots$$

$$n_t = p_1^{v_{it}} p_2^{v_{it}} \cdots p_k^{v_{it}}.$$

Condition (1.8.3) guarantees that $n_1 \mid n_2, n_2 \mid n_3, \ldots, n_{t-1} \mid n_t$. Then, using (1.8.2) gives

$$A = A_1 \times \cdots \times A_k$$

$$= (\mathbb{Z}_{p_1^{n_1}} \times \cdots \times \mathbb{Z}_{p_1^{n_1}}) \times (\mathbb{Z}_{p_1^{n_2}} \times \cdots \times \mathbb{Z}_{p_k^{n_2}}) \times \cdots \times (\mathbb{Z}_{p_1^{n_t}} \times \cdots \times \mathbb{Z}_{p_k^{n_t}})$$

$$\cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_t}.$$
The integers \( n_i \) are completely determined by the decomposition of \( A \) into a direct of cyclic \( p_i \)-groups. Conversely, given that
\[
A = \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_t},
\]
where \( n_1 \mid n_2, n_2 \mid n_3, \ldots, n_{t-1} \mid n_t \), the decomposition of \( A \) into a direct product of cyclic \( p_i \)-groups is completely determined. Thus, we have:

**1.8.29. Theorem.** [Structure Theorem for Finite Abelian Groups] Let \( A \) be a finite abelian group. Then there exist integers \( n_1, \ldots, n_t > 1 \) such that \( n_1 \mid n_2, n_2 \mid n_3, \ldots, n_{t-1} \mid n_t \) and
\[
A \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_t},
\]
where these integers are uniquely defined by \( A \). More precisely, if \( m_1, \ldots, m_s \) are positive integers greater than 1 such that \( m_1 \mid m_2, m_2 \mid m_3, \ldots, m_{s-1} \mid m_s \), and
\[
A \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_t} \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_s},
\]
then \( t = s \), and \( n_1 = m_1, \ldots, n_t = m_t \).

**1.8.30. Example.** Find all non-isomorphic abelian groups of order:

1. 6 
2. 12 
3. 27 
4. 500.

**Solution.** By Example 1.8.26 and Theorem 1.8.29. We have the following answers.

\[
6 = 2 \cdot 3: \mathbb{Z}_2 \times \mathbb{Z}_3 \cong \mathbb{Z}_6,
\]

\[
12 = 2^2 \cdot 3: \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \cong \mathbb{Z}_2 \times \mathbb{Z}_6 \text{ and } \mathbb{Z}_{2^2} \times \mathbb{Z}_3 \cong \mathbb{Z}_{12},
\]

\[
27 = 3^3: \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \cong \mathbb{Z}_{27},
\]

\[
500 = 2^2 \cdot 5^3: \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \cong \mathbb{Z}_2 \times \mathbb{Z}_{10} \times \mathbb{Z}_{10}, \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \cong \mathbb{Z}_{10} \times \mathbb{Z}_{50}, \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{250}, \mathbb{Z}_2 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \cong \mathbb{Z}_6 \times \mathbb{Z}_6 \times \mathbb{Z}_{20}, \mathbb{Z}_2 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \cong \mathbb{Z}_5 \times \mathbb{Z}_{100} \text{ and } \mathbb{Z}_{2^2} \times \mathbb{Z}_{5^3} \cong \mathbb{Z}_{600}.
\]

By Cauchy's theorem, if \( p \) is a divisor of the order of a group \( G \), then \( G \) has a subgroup of order \( p \). We also see that \( A_4 \) is a group of order 12 but it has no subgroup of order six. Hence, it may not hold that \( G \) will have a subgroup of order \( m \) when \( m \) is a divisor of \( |G| \). However, if \( G \) is an abelian group, we have our final results.

**1.8.31. Corollary.** Let \( A \) be a finite abelian group. If \( m \) divides the order of \( A \), then \( A \) has a subgroup of order \( m \).

**Proof.** Write \( A = \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_t} \) as in the above theorem. Then \( |A| = n_1 n_2 \cdots n_t \). Since \( m \) divides \( |A| \), \( m = l_1 l_2 \cdots l_t \) with \( l_i \mid n_i \) for all \( i \in \{1, 2, \ldots, t\} \). Then \( (n_i/l_i)\mathbb{Z}_{n_i} \) is a subgroup of \( \mathbb{Z}_{n_i} \) of order \( l_i \) for all \( i \). Thus,
\[
(n_1/l_1)\mathbb{Z}_{n_1} \times (n_2/l_2)\mathbb{Z}_{n_2} \times \cdots \times (n_t/l_t)\mathbb{Z}_{n_t}
\]
is a subgroup of \( A \) of order \( l_1 l_2 \cdots l_t = m \) as desired.

**1.8.32. Corollary.** Let \( A \) be a finite abelian group. Then there exists \( g \in A \) such that the order of \( g \) is the exponent of \( A \).

**Proof.** By Theorem 1.8.29, there exist positive integers \( n_1, n_2, \ldots, n_t \geq 1 \) such that \( n_1 \mid n_2 \mid \cdots \mid n_t \) and
\[
A \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_t}.
\]
Thus, \( \exp A = n_t \) and \( (0, 0, \ldots, 0, 1) \) in the rightmost group has order \( n_t \).
1.8.33. Remark. The above corollary is false if \( A \) is non-abelian. For example, the exponent of \( S_3 \) is 6, however \( S_3 \) contains no elements of order 6.

1.8. Exercises.  
1. Suppose \( G_1 \) and \( G_2 \) are finite groups of relatively prime orders. Show that every subgroup of \( G_1 \times G_2 \) is of the form \( H_1 \times H_2 \) for some subgroups \( H_1 \) and \( H_2 \) of \( G_1 \) and \( G_2 \), respectively.  
2. Let \( G_1 \) and \( G_2 \) be simple groups. Show that every nontrivial normal subgroup of \( G = G_1 \times G_2 \) is isomorphic to either \( G_1 \) or \( G_2 \).  
3. Prove Theorem 1.8.10.  
4. Find the order of torsion subgroup of \( \mathbb{Z}_4 \times \mathbb{Z} \times \mathbb{Z} \) and of \( \mathbb{Z}_{12} \times \mathbb{Z}_{12} \times \mathbb{Z} \).  
5. Find the torsion subgroup of the multiplicative group \( \mathbb{R}^* \).  
6. Let \( G \) be an abelian group of order 72.  
   (a) Can you say how many subgroups of order eight \( G \) has? Why?  
   (b) Can you say how many subgroups of order four \( G \) has? Why?  
7. Show that a finite abelian group is not cyclic if and only if it contains a subgroup isomorphic to \( \mathbb{Z}_p \times \mathbb{Z}_p \) for some prime \( p \).  
8. Find the exponent of the following groups:  
   (a) \( \mathbb{Z}_6 \times \mathbb{Z}_9 \) \hspace{1em} (b) \( \mathbb{Z}_6 \times S_4 \).  
9. Find all non-isomorphic abelian groups of order \( 35 \) \hspace{1em} (b) \( 48 \) \hspace{1em} (c) \( 360 \).  
10. List (up to isomorphism) all abelian groups of order 108 and express each in the following two ways:  
   (a) As a direct sum of cyclic groups of prime power order.  
   (b) As a direct sum of cyclic groups of order \( d_1, d_2, \ldots, d_k \) where \( d_i \mid d_{i+1} \) for \( i = 1, 2, \ldots, k - 1 \).  
11. List all groups of order 99 and of order 1225 up to isomorphism. (Hint. Show that they must be abelian.)  
12. Let \( G \) be a finite group. Prove the following statements.  
   (a) The exponent of \( G \) divides its order \( |G| \).  
   (b) If \( H \) is a subgroup of \( G \), then the exponent of \( H \) divides the exponent of \( G \).  
   (c) If \( G \) is cyclic, then \( \exp G = |G| \).  
13. Let \( G = S_3 \times \mathbb{Z}_4 \).  
   (a) Find \( \exp G \).  
   (b) Determine all Sylow 2-subgroups and Sylow 3-subgroups of \( G \).  
14. Let \( G \) be a group of order 2156 = \( 2^2 \cdot 7^2 \cdot 11 \).  
   (a) If \( G \) is abelian, list all \( G \) up to isomorphism.  
   (b) Prove that \( G \) cannot be simple.  
15. List all finite groups \( G \) which have the property: \( \forall g, h \in G, g \) is a power of \( h \) or \( h \) is a power of \( g \).

9. Project. (Characters of a group) Let \( S^1 = \{ z \in \mathbb{C} : |z| = 1 \} \) be the unit circle in the complex plane. It is a subgroup of \( S^\times \). A character of a finite group \( G \) is a homomorphism from \( G \) to \( S^1 \). The character sending \( G \) to 1 is called the trivial character. Let \( G \) be a finite group. Prove the following statements.  
(a) All characters of \( G \) form an abelian group under pointwise multiplication, called the dual group of \( G \) and denoted by \( \hat{G} \).  
(b) Prove that if \( G \) is a finite cyclic group, then \( G \) and \( \hat{G} \) are isomorphic.  
(c) If \( G \) is a finite abelian group, then \( G \) is isomorphic to its dual group \( \hat{G} \). (Hint. Use (b) and Theorem 1.8.27.)
Rings and Fields

Rings and fields are the most common algebraic structures for students. They have learned addition together with multiplication since elementary schools. The abstract treatments using groups are presented in the first section. Ideals and factorizations are discussed in details. Finally, we talk about polynomials over a ring and which will be used in a construction of field extensions.

2.1 Basic Concepts

2.1.1 Rings

2.1.1. Definition. A ring is a triple \((R, +, \cdot)\) where

\begin{itemize}
  \item \((R, +)\) is an abelian group,
  \item \((R, \cdot)\) is a semigroup, and
  \item \(+\) and \(\cdot\) satisfy the distributive laws, namely,
  \[
  \forall a, b, c \in R, a \cdot (b + c) = (a \cdot b) + (a \cdot c) \quad \text{and} \quad \forall a, b, c \in R, (b + c) \cdot a = (b \cdot a) + (c \cdot a).
  \]
\end{itemize}

The binary operations \(+\) and \(\cdot\) are called the addition and the multiplication of the ring \((R, +, \cdot)\), respectively, \(0\), the identity of \((R, +)\) is called the zero of \(R\), for \(a, b \in R\), \(ab\) (juxtaposition) may denote \(a \cdot b\) and \((R, +, \cdot)\) may be denoted by \(R\). For \(a \in R\), \(-a\) is called the additive inverse of \(a\) in \(R\).

2.1.2. Definition. A ring \(R\) is said to be commutative if \(\forall a, b \in R, ab = ba\). If \(1\) is the identity of \((R, \cdot)\), then \(1\) is called the identity or unity of the ring \(R\). If \(R\) contains an identity, i.e., \((R, \cdot)\) is a multiplicative monoid, then \(R\) is called a ring with identity.

Unless the contrary is explicitly stated “ring” will mean “ring with identity”.

2.1.3. Definition. A subset \(S\) of a ring \(R\) is a subring if \(S\) is a subgroup of the additive group and also a submonoid of the multiplicative monoid of \(R\). Clearly the intersection of any set of subrings of \(R\) is a subring. Hence, if \(A\) is a subset of \(R\), we may define the subring generated by \(A\) to be the intersection of all subrings of \(R\) which contain \(A\).

2.1.4. Examples. (Examples of rings)  
1. \(\mathbb{Z}, \mathbb{Q}, \mathbb{R}\) and \(\mathbb{C}\) are commutative rings under usual addition and multiplication.
2. For \(n \in \mathbb{N}\), \((\mathbb{Z}_n, +, \cdot)\) is a commutative ring.
3. Recall that for a nonempty set \(X\) and \(P(X)\) the power set of \(X\), we define \(A \triangle B = (A \setminus B) \cup (B \setminus A)\) for all subsets \(A\) and \(B\) of \(X\). Then \((P(X), \triangle, \cap)\) is a commutative ring with identity \(X\).
4. If \(A\) is an abelian group, then \(\text{End}(A)\), the set of all homomorphisms on \(A\), is a ring with the addition and multiplication are given by

\[
(f + g)(a) = f(a) + g(a) \quad \text{and} \quad (f \cdot g)(a) = f(g(a))
\]

for all \(f, g \in \text{End}(A)\) and \(a \in A\). Its identity is the identity map.
5. If \( X \) is a topological space, then \( C(X) \), the set of all continuous functions from \( X \) to \( \mathbb{R} \), is a commutative ring with pairwise operations. That is,

\[(f + g)(x) = f(x) + g(x) \quad \text{and} \quad (f \cdot g)(x) = f(x)g(x)\]

for all \( f, g \in C(X) \) and \( x \in X \).

6. Let \( d \) be a square free integer. The set \( \mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} : a, b \in \mathbb{Z}\} \) is a subring of \( \mathbb{C} \).

A number of elementary properties of rings are consequences of the fact that a ring is an abelian group relative to addition and a monoid relative to multiplication. For example, we have \(- (a + b) = -a - b := (-a) + (-b)\) and if \( na \) is defined for \( n \in \mathbb{Z} \) as before, then the rules for multiples (or powers) in an abelian group,

\[n(a + b) = na + nb, \quad (n + m)a = na + ma, \quad \text{and} \quad (nm)a = n(ma)\]

hold. There are also a number of simple consequences of the distributive laws which we now note. In the first place, induction on \( m \) and \( n \) gives the generalization

\[ \left( \sum_{i=1}^{m} a_i \right) \left( \sum_{j=1}^{n} b_j \right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_ib_j. \]

We note next that \( a0 = 0a = 0 \) for all \( a \in R \); for we have \( a0 = a(0 + 0) = a0 + a0 \). Addition of \(-a0\) gives \( a0 = 0 \). Similarly, \( 0a = 0 \). We have the equation

\[ 0 = 0b = (a + (-a))b = ab + (-a)b, \]

which shows that \((-a)b = -ab\). Similarly, \( a(-b) = -ab\); consequently

\[ (-a)(-b) = a(-b) = -(ab) = ab. \]

If \( a \) and \( b \) commute, that is, \( ab = ba \), then \( a^mb^n = b^n a^m \). Also, by induction we can prove the binomial theorem

\[(a + b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n-1}ab^{n-1} + b^n\]

for all \( n \in \mathbb{N} \).

2.1.5. Remark. In the case \( 1 = 0 \) in a ring \( R \), we have that

\[ ab = (a \cdot 1)b = a(1 \cdot b) = a(0 \cdot b) = a \cdot 0 = 0 \quad \text{for all} \ a, b \in R. \]

A ring with this property is called a zero ring.

2.1.6. Definition. If \( R \) and \( S \) are rings, so is \( R \times S \), with coordinatewise operations:

\[(r, s) + (r', s') = (r + r', s + s') \quad \text{and} \quad (r, s)(r', s') = (rr', ss').\]

Note that \((1_R, 1_S)\) is the identity of \( R \times S \). More generally, if \( R_\alpha \) is a family of rings, then \( \prod_\alpha R_\alpha \) is a ring with coordinatewise operations.

2.1.7. Definition. Let \( R \) be a ring. An element \( x \in R \) is said to be invertible or a unit if there is a \( y \in R \) such that \( xy = yx = 1 \). In this case, \( y \) is called the inverse of \( x \).
2.1.8. Remark. If \( x \) is invertible, its inverse is unique. The invertible elements of \( R \) form a group under multiplication, called the **group of units of** \( R \) and denoted by \( \mathcal{U}(R) \) or \( R^\times \).

2.1.9. Definition. A ring \( D \) is a **division ring** or **skew field** if every nonzero element of \( D \) is invertible. A commutative division ring is called a **field**.

2.1.10. Examples. 1. \( \mathbb{Q}, \mathbb{R} \) and \( \mathbb{C} \) are fields. \( \mathbb{Z} \) is not a field.
2. By Example 1.311 (5), we have \( \forall n \in \mathbb{N}, \mathbb{Z}_n \) is a field \( \iff n \) is a prime.

2.1.11. Example. Let \( n \in \mathbb{N}, R \) a ring and \( M_n(R) \) the set of all \( n \times n \) matrices over \( R \). Then \( (M_n(R), +, \cdot) \) is a ring under the usual addition and multiplication of matrices with identity \( I_n \), the identity matrix. If \( n > 1 \), then \( M_n(R) \) is not commutative. The group of invertible elements of \( M_n(R) \) is called the **general linear group** and denoted by \( \text{GL}_n(R) \). For the case \( R \) is commutative, we can derive the determinant criterion for a matrix \( A \) to be invertible. We have the following results.

2.1.12. Theorem. Let \( R \) be a commutative ring and \( A \in M_n(R) \). Then
\[
A(\text{adj } A) = (\det A)I_n = (\text{adj } A)A.
\]
In particular, \( A \) is invertible if and only if its determinant is invertible in \( R \).

A noteworthy special case of the theorem is the next corollary.

2.1.13. Corollary. If \( F \) is a field, \( A \in M_n(F) \) is invertible if and only if \( \det A \neq 0 \).

Some rings do not have the property that the product of two nonzero elements is always nonzero. If so, it leads to the cancellation property in the rings.

2.1.14. Definition. Let \( R \) be a ring and \( 0 \neq a \in R \). \( a \) is called a **left [right] zero divisor** if \( \exists b \in R \setminus \{0\}, ab = 0 \ [ba = 0] \), and it is called a **zero divisor** if it is a left or a right zero divisor. We say that \( R \) is **entire** if it possesses no zero divisors. A commutative entire ring is called an **integral domain**.

2.1.15. Examples. 1. \( \mathbb{Z} \) is an integral domain which is not a field.
2. Every field is an integral domain.
3. \( C([0, 1]) \) is not an integral domain.

2.1.16. Remark. Let \( R \) be a ring. Then we have
\( R \) is entire \( \iff \forall a, b \in R \ [ab = 0 \Rightarrow (a = 0 \text{ or } b = 0)] \iff (R \setminus \{0\}, \cdot) \) is a cancellative semigroup.

Theorem 1.3.14 implies an important result on finiteness of an integral domain.

2.1.17. Theorem. Every finite integral domain is a field.

**Proof.** Let \( D \) be a finite integral domain. Then \( (D \setminus \{0\}, \cdot) \) is a finite cancellative semigroup. By Theorem 1.3.14, \( (D \setminus \{0\}, \cdot) \) is a group. Hence, if \( a \in D \) and \( a \neq 0 \), then \( a \) has an inverse under \( \cdot \). Since \( D \) is commutative, \( D \) is a field. \( \square \)
2.1.2 Quaternions

In 1843, W. R. Hamilton constructed the first example of a division ring in which the commutative law of multiplication does not hold. This was an extension of the field of complex numbers, whose elements were quadruples of real numbers \((\alpha, \beta, \gamma, \delta)\) for which the usual addition and a multiplication were defined so that \(1 = (1, 0, 0, 0)\) is the unit and \(i = (0, 1, 0, 0), j = (0, 0, 1, 0), \) and \(k = (0, 0, 0, 1)\) satisfy \(i^2 = j^2 = k^2 = -1 = ijk.\) Hamilton called his quadruples, quaternions. Previously, he had defined complex numbers as pairs of real numbers \((\alpha, \beta)\) with the product \((\alpha x \beta) (\gamma x \delta) = (\alpha \gamma - \beta \delta, \alpha \delta + \beta \gamma)\). Hamilton's discovery of quaternions led to a good deal of experimentation with other such "hypercomplex" number systems and eventually to a structure theory whose goal was to classify such systems. A good deal of important algebra thus evolved from the discovery of quaternions.

We shall not follow Hamilton's way of introducing quaternions. Instead we shall define this system as a certain subring of the ring \(M_2(\mathbb{C})\) of \(2 \times 2\) matrices with complex number entries. This will have the advantage of reducing the calculations to a simple single verification.

We consider the subset \(\mathbb{H}\) of the ring \(M_2(\mathbb{C})\) of complex \(2 \times 2\) matrices that have the form

\[
\begin{bmatrix}
  a & b \\
  -\bar{b} & \bar{a}
\end{bmatrix} = \begin{bmatrix}
  a_0 + a_1 \sqrt{-1} & a_2 + a_3 \sqrt{-1} \\
  -a_2 + a_3 \sqrt{-1} & a_0 - a_1 \sqrt{-1}
\end{bmatrix}, \quad a_i \text{ real.}
\tag{2.1.1}
\]

We claim that \(\mathbb{H}\) is a subring of \(M_2(\mathbb{C})\). Since \((\bar{a_1} - a_2) = \bar{a_1} - a_2\) for complex numbers, it is clear that \(\mathbb{H}\) is closed under subtraction; hence \(\mathbb{H}\) is a subgroup of the additive group of \(M_2(\mathbb{C})\).

We obtain the unit matrix by taking \(a = 1, b = 0\) in (2.1.1). Hence, \(1 \in \mathbb{H}\). Since

\[
\begin{bmatrix}
  a & b \\
  -\bar{b} & \bar{a}
\end{bmatrix} \begin{bmatrix}
  c & d \\
  -\bar{d} & \bar{c}
\end{bmatrix} = \begin{bmatrix}
  ac - bd\bar{d} & ad + bc \\
  -b\bar{c} - \bar{a}d & b\bar{d} + \bar{a}c
\end{bmatrix}
\]

and \(\bar{a_1} \bar{a_2} = \bar{a_1} a_2\), the right-hand side has the form

\[
\begin{bmatrix}
  u & v \\
  \bar{v} & \bar{u}
\end{bmatrix}
\]

where \(u = ac - b\bar{d}, v = ad + bc\). Therefore, \(\mathbb{H}\) is closed under multiplication and so \(\mathbb{H}\) is a subring of \(M_2(\mathbb{C})\).

We now show that \(\mathbb{H}\) is a division ring. We note first that

\[
\Delta := \det \begin{bmatrix}
  a_0 + a_1 \sqrt{-1} & a_2 + a_3 \sqrt{-1} \\
  -a_2 + a_3 \sqrt{-1} & a_0 - a_1 \sqrt{-1}
\end{bmatrix} = \alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2.
\]

Since the \(\alpha_i\) are real numbers, this is real, and is 0 only if every \(\alpha_i = 0\), that is, if the matrix is 0. Hence, every non-zero element of \(\mathbb{H}\) has an inverse in \(M_2(\mathbb{C})\). Moreover, we have, by the definition of the adjoint, that

\[
\text{adj} \begin{bmatrix}
  a & b \\
  -\bar{b} & \bar{a}
\end{bmatrix} = \begin{bmatrix}
  \bar{a} & -b \\
  -\bar{b} & \bar{a}
\end{bmatrix}.
\]

Since \(\bar{a} = a\), this is obtained from the \(x\) in (2.1.1) by replacing \(a\) by \(\bar{a}\) and \(b\) by \(-b\), and so it is contained in \(\mathbb{H}\). Thus, if the matrix \(x\) is \(\neq 0\) then its inverse is

\[
\begin{bmatrix}
  a & b \\
  -\bar{b} & \bar{a}
\end{bmatrix}^{-1} = \begin{bmatrix}
  \bar{a} & -b \\
  -\bar{b} & \bar{a}
\end{bmatrix}
\]

and this is contained in \(\mathbb{H}\). Hence, \(\mathbb{H}\) is a division ring.

The ring \(\mathbb{H}\) contains in its center the field \(\mathbb{R}\) of real numbers identified with the set of diagonal matrices \(\begin{bmatrix}
  \alpha & 0 \\
  0 & \alpha
\end{bmatrix}, \alpha \in \mathbb{R}\). \(\mathbb{H}\) also contains the matrices

\[
i = \begin{bmatrix}
  \sqrt{-1} & 0 \\
  0 & -\sqrt{-1}
\end{bmatrix}, \quad j = \begin{bmatrix}
  0 & 1 \\
  -1 & 0
\end{bmatrix}, \quad k = \begin{bmatrix}
  0 & \sqrt{-1} \\
  \sqrt{-1} & 0
\end{bmatrix}.
\]
We verify that
\[ x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k \quad \text{(2.1.2)} \]
and if \( \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 k, \beta_i \in \mathbb{R}, \) then
\[
\begin{bmatrix}
\alpha_0 + \alpha_1 \sqrt{-1} & \alpha_2 + \alpha_3 \sqrt{-1} \\
-\alpha_2 + \alpha_3 \sqrt{-1} & \alpha_0 - \alpha_1 \sqrt{-1}
\end{bmatrix}
= \begin{bmatrix}
\beta_0 + \beta_1 \sqrt{-1} & \beta_2 + \beta_3 \sqrt{-1} \\
-\beta_2 + \beta_3 \sqrt{-1} & \beta_0 - \beta_1 \sqrt{-1}
\end{bmatrix}
\]
so \( \alpha_i = \beta_i, 0 \leq l \leq 3. \) Thus, any \( x \in \mathbb{H} \) can be written in one and only one way in the from (2.1.2). The product of two elements in \( \mathbb{H} \)
\[(\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k)(\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k)\]
is determined by the product and sum in \( \mathbb{R} \), the distributive laws and the multiplication table
\[ i^2 = j^2 = k^2 = -1 \]
\[ ij = -ji = k, jk = -kj = i, ki = -ik = j. \]
Incidentally, because these show that \( \mathbb{H} \) is not commutative we have constructed an infinite division ring that is not a field. The ring \( \mathbb{H} \) is called the division ring of real quaternions. Recall that \( Q_8 = \{ \pm 1, \pm i, \pm j, \pm k \} \) is called the quaternion group (see Project 1).

Algebra is not as rich in division ring as it is in fields. For example, there are no finite division rings. This is the content of a famous theorem of Wedderburn.

**2.1.18. Theorem.** [Wedderburn, 1909] A finite division ring is a field.

### 2.1.3 Characteristic

**2.1.19. Definition.** Let \( R \) be a ring. If there is a smallest positive integer \( n \) such that \( na = 0 \) for all \( a \in R \), then \( R \) is said to have characteristic \( n \). If no such \( n \) exists, \( R \) is said to have characteristic zero. We denote the characteristic of \( R \) by \( \text{char} R \).

The characteristic of a ring gives some information on its additive group structure.

**2.1.20. Remark.** It is easy to see that \( \text{char} R = n \) if and only if \( n \) is the smallest positive integer such that \( n1_R = 0 \).

**2.1.21. Example.** The rings \( \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C} \) and \( \mathbb{H} \) are of characteristic zero, \( \text{char} \mathbb{Z}_n = n \) and \( \text{char}(\mathbb{Z}_m \times \mathbb{Z}_n) = \text{lcm}(m, n) \).

**2.1.22. Theorem.** [Characteristic of an Integral Domain] If \( R \) is an integral domain, then \( R \) is of characteristic zero or a prime \( p \). In particular, every field is of characteristic zero or a prime \( p \).

**Proof.** Let \( R \) be an integral domain of characteristic \( n > 0 \). Assume that \( n = ab \) for some \( a, b \in \mathbb{N} \). It follows that \( 0 = n1_R = (ab)1_R = (a1_R)(b1_R) \). Since \( R \) has no zero divisor, \( a1_R = 0 \) or \( b1_R = 0 \). Then \( a = n \) or \( b = n \). Hence, \( n \) is a prime. \( \square \)

**2.1.23. Theorem.** Let \( R \) be a ring of characteristic a prime \( p \) and \( a, b \in R \). If \( a \) and \( b \) commute, then
\[ (a + b)^p = a^p + b^p \quad \text{and} \quad (a + b)^{pk} = a^{pk} + b^{pk} \quad \text{for all} \quad k \in \mathbb{N}. \]
Proof. Note that if 1 ≤ r ≤ p − 1, then the binomial coefficient \( \binom{p}{r} \) is a multiple of p, so it is 0 in R. Hence,
\[
(a + b)^p = a^p + \binom{p}{1}a^{p-1}b + \cdots + \binom{p}{p-1}ab^{p-1} + b^p = a^p + b^p.
\]
A simple induction on \( k \) gives the second equation. The inductive step is
\[
(a + b)^p = ((a + b)^{p-1})^p = (a^{p-1} - b^{p-1})^p = a^p + b^p
\]
and the proof is complete.

2.1.4 Ring Homomorphisms and Group Rings

Like in groups, a ring homomorphism is a function between two rings that preserves both addition and multiplication.

2.1.24. Definition. Let \( R \) and \( S \) be rings. A map \( \varphi : R \to S \) is called a homomorphism if
\[
\varphi(a + b) = \varphi(a) + \varphi(b) \quad \text{and} \quad \varphi(ab) = \varphi(a)\varphi(b)
\]
for all \( a, b \in R \). The definitions of monomorphisms, epimorphisms, endomorphisms, isomorphisms and automorphisms of rings are given as in groups.

2.1.25. Remarks. Let \( \varphi : R \to S \) be a ring homomorphism.
1. \( \varphi : (R, +) \to (S, +) \) is an additive group homomorphism.
2. \( \varphi(1_R) \) may not be the identity of \( S \).
3. If \( \varphi(1_R) = 0 \), then \( \varphi(x) = 0 \) for all \( x \in R \).
4. If \( R \) has an identity and \( \varphi \) is onto, then \( \varphi(1_R) \) is the identity of \( S \).

Proof. Let \( s \in S \). Since \( \varphi \) is onto, \( \exists x \in R, \varphi(x) = s \). Then \( s\varphi(1_R) = \varphi(x)\varphi(1_R) = \varphi(x1_R) = \varphi(x) = s \). Similarly, \( \varphi(1_R)s = s \). Hence, \( \varphi(1_R) \) is the identity of \( S \).

2.1.26. Examples. 1. If \( \varphi : \mathbb{Z} \to \mathbb{Z} \) is a ring homomorphism, then \( \varphi \) is the zero or the identity map.
2. If \( \varphi : \mathbb{Q} \to \mathbb{Q} \) is a ring homomorphism, then \( \varphi \) is the zero or the identity map.
3. If \( \varphi : \mathbb{R} \to \mathbb{R} \) is a ring homomorphism, then \( \varphi \) is the zero or the identity map.
4. If \( p \) is a prime and \( \varphi : \mathbb{Z}_p \to \mathbb{Z}_p \) is a ring homomorphism, then \( \varphi \) is the zero or the identity map.

Proof. (1) Observe that \( \varphi(1) = \varphi(1 \cdot 1) = \varphi(1)\varphi(1) \), so \( \varphi(1) = 0 \) or \( \varphi(1) = 1 \). Moreover, \( \varphi(n) = n\varphi(1) \) for all \( n \in \mathbb{Z} \). Thus, \( \varphi(n) = 0 \) for all \( n \in \mathbb{Z} \) or \( \varphi(n) = n \) for all \( n \in \mathbb{Z} \) as desired.
(2) Similar to \( \mathbb{Z} \), \( \varphi(1) = 0 \) or \( \varphi(1) = 1 \) and \( \varphi(n) = n\varphi(1) \) for all \( n \in \mathbb{N} \). For \( m \in \mathbb{Z} \) and \( n \in \mathbb{N} \), we have \( \varphi(m1/n) = m\varphi(1/n) \) and \( \varphi(1) = \varphi(1/n) = n\varphi(n) \). If \( \varphi(1) = 0 \), then \( \varphi(1/n) = 0 \) for all \( n \in \mathbb{N} \), so \( \varphi \) is the zero map. On the other hand, if \( \varphi(1) = 1 \), then \( \varphi(1/n) = 1/n \) for all \( n \in \mathbb{N} \) which implies \( \varphi(m/n) = m/n \) for all \( m \in \mathbb{Z} \) and \( n \in \mathbb{N} \).
(3) Assume that \( \varphi(x) \) is not the zero map. We can show that \( \varphi(x) = x \) for all \( x \in \mathbb{Q} \). Moreover, for \( x \in \mathbb{R}^+ \), \( \varphi(x) = \varphi(\sqrt{x})^2 = (\varphi(x))^2 > 0 \). This implies \( \forall a, b \in \mathbb{R}, a < b \implies \varphi(a) < \varphi(b) \). Now, let \( x \in \mathbb{R} \). Suppose that \( \varphi(x) \neq x \). Then \( \varphi(x) < x \) or \( x < \varphi(x) \). By the density theorem, \( \exists q_1, q_2 \in \mathbb{Q} \) such that \( \varphi(x) < q_1 < x \) or \( x < q_2 < \varphi(x) \). Thus, \( \varphi(x) < q_1 < \varphi(x) \) or \( \varphi(x) < q_2 < \varphi(x) \) yields a contradiction. Hence, \( \varphi(x) = x \) for all \( x \in \mathbb{R} \).
(4) is proved in the next section. 

\( \square \)
Let $G = \{g_i : i \in I\}$ be any multiplicative group, and let $R$ be any commutative ring. Let $RG$ be the set of all formal sums
\[ \sum_{i \in I} a_i g_i \]
for $a_i \in R$ and $g_i \in G$, where all but finite number of the $a_i$ are 0. Define the sum of two elements of $RG$ by
\[ \left( \sum_{i \in I} a_i g_i \right) + \left( \sum_{i \in I} b_i g_i \right) = \sum_{i \in I} (a_i + b_i) g_i. \]
Observe that $(a_i + b_i) = 0$ except for a finite number of indices $i$, so the above sum is again in $RG$. It is immediate that $(RG, +)$ is an abelian group with additive identity $\sum_{i \in I} 0 g_i$.

Multiplication of two elements of $RG$ is defined by the use of the multiplications in $G$ and $R$ as follows:
\[ \left( \sum_{i \in I} a_i g_i \right) \left( \sum_{i \in I} b_i g_i \right) = \sum_{i \in I} \left( \sum_{g_jg_k = g_i} a_j b_k \right) g_i. \]
Naively, we formally distribute the sum $\sum_{i \in I} a_i g_i$ over the sum $\sum_{i \in I} b_i g_i$ and rename a term $a_j g_j b_k g_k$ by $a_j b_k g_k$ where $g_j g_k = g_i$ in $G$. Since $a_i$ and $b_i$ are 0 for all but a finite number of $i$, the sum $\sum_{g_jg_k = g_i} a_j b_k$ contains only a finite number of nonzero summands $a_j b_k \in R$ and may thus be viewed as an element of $R$. Again at most a finite number of such sums $\sum_{g_jg_k = g_i} a_j b_k$ are nonzero. Thus, multiplication is closed on $RG$. We can proceed to show that

2.1.27. Theorem. [Group Ring] If $G$ is a multiplicative group and $R$ is a commutative ring, then $(RG, +, \cdot)$ is a ring with unity $1_R e$.

If we rename the element $\sum_{i \in I} a_i g_i$ of $RG$, where $a_i = 0$ for $i \neq j$ and $a_j = 1$, by $g_j$, we see that $(RG, \cdot)$ can be considered to contain $G$ naturally. Thus, if $G$ is not abelian, $RG$ is not commutative. Clearly, $\text{char } RG = \text{char } R$, for any group $G$. The ring $RG$ defined above is the group ring of $G$ over $R$. If $F$ is a field, then $FG$ is the group algebra of $G$ over $F$.

2.1. Exercises. 1. Define an addition and a multiplication on $\mathbb{Z}$ by
\[ a \oplus b = a + b - 1 \quad \text{and} \quad a \odot b = ab - (a + b) + 2 \quad \text{for all } a, b \in \mathbb{Z}. \]
Prove that $(\mathbb{Z}, \oplus, \odot)$ is an integral domain.
2. Let $S$ be the set of complex numbers of the form $m + n\sqrt{-3}$ where either $m, n \in \mathbb{Z}$ or both $m$ and $n$ are halves of odd integers. Show that $S$ is a subring of $\mathbb{C}$.
3. Show that if $1 - ab$ is invertible in a ring, then so is $1 - ba$.
4. Let $a$ and $b$ be elements of a ring such that $a, b$ and $ab - 1$ are units. Show that $a - b^{-1}$ and $(a - b^{-1})^{-1} - a^{-1}$ are units and the following identity holds:
\[ ((a - b^{-1})^{-1} - a^{-1})^{-1} = ab - a. \]
5. A ring $R$ is called a Boolean ring if $x^2 = x$ for all $x \in R$. Prove that every Boolean ring is commutative.
6. (a) Show that $\varphi : \mathbb{Z}_{12} \to \mathbb{Z}_{30}$ given by $\varphi([a]_{12}) = [10a]_{30}$ is a ring homomorphism.
   (b) Show that $\varphi : \mathbb{Z}_{12} \to \mathbb{Z}_{30}$ given by $\varphi([a]_{12}) = [5a]_{30}$ is an additive group homomorphism. Is it a ring homomorphism?
7. Consider $(S, +, \cdot)$, where $S$ is a set and $+$ and $\cdot$ are binary operations on $S$ which satisfy the distributive laws such that $(S, +)$ and $(S \setminus \{0\}, \cdot)$ are groups. Show that $(S, +, \cdot)$ is a division ring.
8. Let $R$ be a ring. Define $C(R) = \{x \in R : \forall y \in R, xy = yx\}$, called the center of $R$.
   (a) Prove that $C(R)$ is a commutative subring of $R$.
   (b) Determine the centers of $\mathbb{H}$ and $M_n(F)$ where $F$ is a field.
   (c) If $R$ is a division ring, show that $C(R)$ is a field.
9. If \( p \) is a prime and \( \varphi : \mathbb{Z}_p \to \mathbb{Z}_p \) is a ring homomorphism, show that \( \varphi \) is the zero or the identity map.

10. Show that if \( F \) is a field, \( A \in M_n(F) \) is a zero divisor in this ring if and only if \( A \) is not invertible. Does this hold for arbitrary commutative ring \( R \)? Explain.

11. Let \( m \) and \( n \) be non-zero integers and let \( R \) be the subset of \( M_2(\mathbb{C}) \) consisting of the matrices of the form

\[
\begin{bmatrix}
a + b\sqrt{m} & c + d\sqrt{m} \\
(n-c-d\sqrt{m}) & a - b\sqrt{m}
\end{bmatrix}
\]

where \( a, b, c, d \in \mathbb{Q} \). Show that \( R \) is a subring of \( M_2(\mathbb{C}) \) and that \( R \) is a division ring if and only if the only rational numbers \( x, y, z, t \) satisfying the equation \( x^2 - my^2 - nz^2 + mnt^2 = 0 \) are \( x = y = z = t = 0 \). Give a choice of \( m, n \) that \( R \) is a division ring and a choice of \( m, n \) that \( R \) is not a division ring.

12. Let \( R \) be a ring which may not contain the unity 1. Define two binary operations on \( R \times \mathbb{Z} \) by

\[
(r, k) + (s, m) = (r + s, k + m) \quad \text{and} \quad (r, k) \cdot (s, m) = (rs + ks + mr, km).
\]

Prove that \( (R \times \mathbb{Z}, +, \cdot) \) is a ring with unity \((0, 1)\) and of characteristic zero. Ditto the set \( R \times \mathbb{Z}_n \) and prove that it is a ring of characteristic \( n \).

13. A ring \( R \) is simple if \( R \) and \( \{0\} \) are the only ideals in \( R \). Show that the characteristic of a simple ring is either 0 or a prime \( p \).

14. If \( R \) is a finite integral domain, prove that \(|R|\) is a prime power.

2.2 Ideals, Quotient Rings and the Field of Fractions

Ideals play an important role in ring theory. They are used to construct quotient rings like normal subgroups.

**2.2.1. Definition.** Let \( R \) be a ring. A subset \( I \) of \( R \) is called a left [right] ideal of \( R \) if

1. \( I \) is a subgroup of \((R, +)\) and
2. \( \forall r \in R \forall a \in I, ra \in I \) \( [ar \in I] \).

It is called a two-sided ideal or an ideal of \( R \) if \( I \) is both a left and a right ideal.

E.g., \( \{0\} \) and \( R \) are two-sided ideals of \( R \).

**2.2.2. Theorem.** Let \( \varphi : R \to S \) be a homomorphism of rings. Then the kernel of \( \varphi \) given by

\[
\ker \varphi = \{ x \in R : \varphi(x) = 0_S \}
\]

is an ideal of \( R \).

**Proof.** It is immediate that \( \ker \varphi \) is a subgroup of \((R, +)\). If \( a \in R \) and \( x \in \ker \varphi \), then \( \varphi(ax) = \varphi(a)\varphi(x) = \varphi(a)0 = 0 \) and \( \varphi(xa) = \varphi(x)\varphi(a) = 0\varphi(a) = 0 \). Hence, \( ax \) and \( xa \) are in \( \ker \varphi \).

**2.2.3. Remark.** Similar to a group homomorphism, for a ring homomorphism, we have \( \varphi \) is one-to-one if and only if \( \ker \varphi = \{0\} \).

For subsets \( X \) and \( Y \) of \( R \), let \( XY \) denote the set of all finite sums in the form

\[
\sum_{i=1}^{n} x_i y_i, \quad \text{where} \quad x_i \in X, y_i \in Y \quad \text{and} \quad n \in \mathbb{N}.
\]

For \( a \in R \), we have \( Ra = \{ ra : r \in R \} \) and \( aR = \{ ar : r \in R \} \).

**2.2.4. Examples.**

1. All distinct ideals of \( \mathbb{Z}_n \) are \( d\mathbb{Z}_n \), where \( d = 0 \) or \((d \in \mathbb{N} \text{ and } d \mid n)\).
2. All distinct ideals of \( \mathbb{Z} \) are \( m\mathbb{Z} \), where \( m \in \mathbb{N} \cup \{0\} \).
2.2.5. Remarks. Let \( R \) be a ring.
1. If \( I \) is an ideal of \( R \), then \( IR = I = RI \).
2. If a left [right, two-sided] ideal \( I \) of \( R \) contains a unit, then \( I = R \).
3. If \( R \) is a division ring, then \( \{0\} \) and \( R \) are the only left [right, two-sided] ideals of \( R \).
4. An arbitrary intersection of left [right, two-sided] ideals of \( R \) is a left [right, two-sided] ideal of \( R \).
5. If \( S \) is a subring of \( R \) and \( I \) is an ideal of \( R \), then \( S + I \) is a subring of \( R \), \( I \) is an ideal of \( S + I \) and \( S \cap I \) is an ideal of \( S \).
6. If \( I \) and \( J \) are left [right, two-sided] ideals of \( R \), then \( I + J \) and \( IJ \) are left [right, two-sided] ideals of \( R \).

2.2.6. Example. Let \( R \) be a ring and \( S = \left\{ \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} : a, b, c \in R \right\} \) a subring of \( M_2(R) \).

Then \( \left\{ \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} : a \in R \right\} \) is a left ideal of \( S \) and \( \left\{ \begin{bmatrix} 0 & 0 \\ 0 & a \end{bmatrix} : a \in R \right\} \) is a right ideal of \( S \).

They are not ideals of \( S \).

2.2.7. Definition. Let \( R \) be a ring. If \( X \subseteq R \), then the ideal of \( R \) generated by \( X \) is the intersection of all ideals containing \( X \) and it is denoted by \( (X) \), so \( (X) \) is the smallest ideal of \( R \) containing \( X \). For \( a_1, \ldots, a_n \in R \), let \( (a_1, \ldots, a_n) \) denote \( \{\{a_1, \ldots, a_n\}\} \). An ideal \( I \) of \( R \) is called a principal ideal if \( I = (a) \) for some \( a \in R \). Observe that for \( a \in R \),

\[
(a) = \left\{ \sum_{i=1}^{m} r_i a s_i : r_i, s_i \in R \text{ and } m \in \mathbb{N} \right\}.
\]

If \( R \) is a commutative ring, then \( \forall a \in R, (a) = aR = Ra \).

2.2.8. Definition. A ring \( R \) is a principal ideal ring if every ideal of \( R \) is principal. A principal ideal domain (PID) is a principal ideal ring which is an integral domain. Hence, an integral domain \( R \) is a PID if \( \{Ra : a \in R\} \) is the set of all ideals of \( R \).

2.2.9. Examples. 1. \( \mathbb{Z}_n \) is a principal ideal ring.

2. \( \mathbb{Z} \) is a PID.

3. Every field has only two ideals, namely \( \{0\} \) and \( \{1\} = F \), so it is a PID.

2.2.10. Remark. Let \( F \) be a field, \( R \) a ring and \( \varphi : F \to R \) a ring homomorphism. Then \( \ker \varphi \) is either \( \{0\} \) or \( F \) which implies \( \varphi \) is 1-1 or is the zero map, respectively. Hence, every nonzero ring homomorphism of fields must be 1-1. In particular, one can readily verify that the only ring endomorphisms of \( \mathbb{Z}_p \) are the zero map and the identity map. This finishes the proof of Example 2.1.26.

2.2.11. Theorem. Let \( R \) be a commutative ring whose only ideals are \( \{0\} \) and \( R \) itself. Then \( R \) is a field.

Proof. Let \( a \in R \setminus \{0\} \). Then \( (a) = R \), so \( 1 \in (a) \). Since \( R \) is commutative, there is a \( b \in R \) such that \( ab = 1 = ba \).

Suppose \( I \) is an ideal of \( R \). Then \( I \) is a subgroup of \( R \), considered as an abelian group, and so we can form the abelian group \( R/I \). The elements of \( R/I \) are cosets

\[
r + I = \{ r + a : a \in I \}.
\]

The addition in \( R/I \) is given by \( (r + I) + (s + I) = (r + s) + I \).
Now let us define a multiplication on \( R/I \), namely
\[
(r + I)(s + I) = rs + I.
\]
Note that if \( r + I = r' + I \) and \( s + I = s' + I \), then \( r - r' \) and \( s - s' \) are in \( I \), so
\[
rs - r's' = (r - r')s + r'(s - s') \in I.
\]
Thus, the above multiplication is well defined, it is easy to see that \( R/I \) is a ring. Hence, we have the next theorem.

**2.2.12. Theorem.** [Quotient Ring] Let \( R \) be a ring and \( I \) an ideal of \( R \). Then the operators
\[
(r + I) + (s + I) = (r + s) + I \quad \text{and} \quad (r + I)(s + I) = rs + I
\]
make \( R/I \) into a ring with unity \( 1 + I \), called the factor or quotient ring of \( R \) by \( I \). The map \( \varphi : R \to R/I \) defined by \( \varphi(r) = r + I \) is an onto ring homomorphism which has kernel \( I \). It is called the canonical projection of \( R \) onto \( R/I \).

There also are three isomorphism theorems for rings. Their proofs are similar to isomorphism theorems for groups. Hence, we shall just sketch them.

**2.2.13. Theorem.** [First Isomorphism Theorem] If \( \varphi : R \to S \) is an onto ring homomorphism, then
\[
R/\ker \varphi \cong \operatorname{im} S.
\]

**Proof.** Define \( \bar{\varphi} : R/\ker \varphi \to S \) by \( \bar{\varphi}(r + \ker \varphi)\varphi(r) \) for all \( r \in R \). Clearly, \( \bar{\varphi} \) is onto and it is easy to check that \( \bar{\varphi} \) is a ring homomorphism. Moreover, for \( r, s \in R \), we have
\[
\varphi(r) = \varphi(s) \iff \varphi(r - s) = 0 \iff r - s \in \ker \varphi \iff r + \ker \varphi = s + \ker \varphi.
\]
Hence, \( \varphi \) is an isomorphism. \( \square \)

**2.2.14. Theorem.** [Second Isomorphism Theorem] If \( S \) is a subring and \( I \) is an ideal of \( R \), then
\[
S/(S \cap I) \cong (S + I)/I.
\]

**Proof.** Define \( \varphi : S \to (S + I)/I \) by \( \varphi(s) = s + I \) for all \( s \in S \). It is easy to verify that \( \varphi \) is a ring homomorphism with kernel \( S \cap I \) and the theorem follows from the first isomorphism theorem. \( \square \)

**2.2.15. Theorem.** [Third Isomorphism Theorem] If \( I \) and \( J \) are ideals of a ring \( R \) such that \( I \subseteq J \), then \( J/I \) is an ideal of \( R/I \) and
\[
(R/I)/(J/I) \cong R/J.
\]

**Proof.** Define \( \varphi : R/I \to R/J \) by \( \varphi(r + I) = r + J \) for all \( r \in R \). It can be verified that \( \varphi \) is a ring homomorphism with kernel \( J/I \) and the theorem follows from the first isomorphism theorem. \( \square \)

**2.2.16. Remark.** As for groups, the third isomorphism theorem gives a 1-1 correspondence between the set of ideals of \( R \) containing \( I \) and the set of ideals of \( R/I \).
We end this section by embedding an integral domain into a field. We say that a ring \( R \) can be embedded in a ring \( R' \) if there exists a monomorphism (i.e., an injective homomorphism) of \( R \) into \( R' \).

2.2.17. Example. A ring \( R \) can be embedded in the ring \( M_n(R) \) by the diagonal map \( a \mapsto aI_n \).

2.2.18. Theorem. Any ring \( R \) without identity can be embedded in a ring \( R' \) with identity. Moreover, \( R' \) can be chosen to be either of characteristic zero or of same characteristic as \( R \).

Proof. Consider the rings \( R \times \mathbb{Z} \) and \( R \times \mathbb{Z}_n \) defined in Exercises 2.1. They are rings with unity \((0, 1)\) and \((0, 1)\), and of characteristic 0 and \( n \), respectively. If \( \text{char} \ R = 0 \), we define \( \varphi : R \to R \times \mathbb{Z} \) by \( \varphi(x) = (x, 0) \) and if \( \text{char} \ R = n \), we define \( \varphi : R \to R \times \mathbb{Z}_n \) by \( \varphi(x) = (x, 0) \). It is easy to show that both functions are monomorphisms. This finishes the proof.

We now wish to show that every integral domain can be embedded in a field, called its field of fractions such that every element of the field is a fraction \( a/b \) where \( a \) and \( b \) lie in the integral domain and \( b \neq 0 \). There is only one problem to overcome: we might wish to define the field to be the set of all “fraction” \( a/b \), with \( b \neq 0 \). But this is not quite right because two different fractions may be the same number. E.g., \( 1/2 = 2/4 = 3/6 \). We overcome this problem by defining an equivalence relation on certain pairs of elements in the integral domain. The results are presented in the next theorem. Its proof is routine and omitted.

2.2.19. Theorem. [Field of Fractions] Suppose \( D \) is an integral domain, and let \( S \) be the set of pairs
\[
\{(r, s) : r, s \in D \text{ and } s \neq 0\}.
\]
1. \( (r, s) \sim (r', s') \iff rs' = r's \) defines an equivalence relation on \( S \).
2. Let \( [r, s] \) denote the equivalence class of \( (r, s) \) and let \( Q(D) \) denote the set of all equivalence classes. Then
\[
[r, s] + [r', s'] = [rs' + r's, ss'] \quad \text{and} \quad [r, s][r', s'] = [rr', ss']
\]
are well defined binary operations on \( Q(D) \).
3. The set \( Q(D) \) is a field with these operations and \( D \) is embedded in \( Q(D) \) by the monomorphism \( r \mapsto [r, 1] \). The field \( Q(D) \) is called the field of fractions or quotient field of \( D \). The equivalence class \([r, s]\) is denoted by \( r/s \).

2.2.20. Remark. If \( R \) is an entire ring which is not commutative, the construction \( Q(R) \) above does not exist in general.

2.2.21. Example. Let \( D \) be an integral domain and \( a, b \in D \). If \( a^m = b^m \) and \( a^n = b^n \), for \( m \) and \( n \) relatively prime positive integers, prove that \( a = b \).

Proof. If \( a = 0 \), then \( b = 0 \) since \( D \) has no zero divisor. Assume that \( a \neq 0 \). Then \( b \neq 0 \). Let \( F \) be the field of fractions of \( D \). Since \( (m, n) = 1 \), \( \exists \, x, y \in \mathbb{Z} \), \( mx + ny = 1 \). Thus, in \( F \), we have
\[
a = a^1 = a^{mx+ny} = (a^m)^x(a^n)^y = (b^m)^x(b^n)^y = b^{mx+ny} = b^1 = b,
\]
so \( a = b \) in \( D \).
2.2. Exercises. 1. An element \( a \) of a ring \( R \) is nilpotent if \( a^n = 0 \) for some \( n \in \mathbb{N} \). Show that the set of all nilpotent elements \( N \) in a commutative ring \( R \) is an ideal, called the nilradical of \( R \). Moreover, prove that \( R/N \) has no nonzero nilpotent.

2. Show that a ring \( R \) has no nonzero nilpotent element if and only if 0 is the only solution of \( x^2 = 0 \) in \( R \).

3. Let \( \varphi : R \to S \) be a homomorphism of rings. Prove the following statements.
   (a) If \( I \) is an ideal of \( R \) and \( \varphi \) is onto, then \( \varphi(I) \) is an ideal of \( S \).
   (b) If \( J \) is an ideal of \( S \), then \( \varphi^{-1}(J) \) is an ideal of \( R \) containing \( \ker \varphi \).

4. Let \( R \) be a commutative ring and \( I \) an ideal of \( R \). Show that
   \[ \sqrt{I} = \{ x \in R : \exists n \in \mathbb{N}, x^n \in I \} \]
   is an ideal of \( R \) which contains \( I \), called the radical of \( I \). In addition, prove that
   (a) \( \sqrt{\sqrt{I}} = \sqrt{I} \)
   (b) if \( \sqrt{I} = R \), then \( I = R \).

5. Let \( R \) and \( S \) be rings and \( \varphi : R \to S \) be such that
   (i) \( \forall t, r \in R, \varphi(r + s) = \varphi(r) + \varphi(s) \)
   (ii) \( \forall r, s \in R, \varphi(rs) = \varphi(r)\varphi(s) \lor \varphi(rs) = \varphi(s)\varphi(r) \).
   Prove that \( \forall r, s \in R, \varphi(rs) = \varphi(r)\varphi(s) \lor \forall r, s \in R, \varphi(rs) = \varphi(s)\varphi(r) \).

6. [Chinese Remainder Theorem] If \( I \) and \( J \) are ideals of a ring \( R \) such that \( I + J = R \), prove that \( R/(I \cap J) \cong R/I \times R/J \).

7. Let \( R \) be a division ring. Prove that any nonzero ring homomorphism \( \varphi : R \to R \) is 1-1.

8. Let \( I \) be an ideal of a ring \( R \) and let \( M_n(I) \) be the set of \( n \times n \) matrices with entries in \( I \). Prove that
   (a) \( M_n(I) \) is an ideal of \( M_n(R) \) and \( M_n(R)/M_n(I) \cong M_n(R/I) \), and
   (b) every ideal of \( M_n(R) \) is of the form \( M_n(I) \) for some ideal \( I \) of \( R \). In particular, if \( R \) is a division ring, then the ring \( M_n(R) \) has only two ideals.

10. Project. Let \( n \in \mathbb{N} \) and \( n \geq 2 \). Define \( Z_n[i] = \{ a + ib : a, b \in \mathbb{Z}_n \} \) where \( i^2 \equiv -1 \) (mod \( n \)).
   (a) Prove that \( Z_n[i] \) is a ring containing \( Z_n \) as a subring.
   (b) Determine all units, zero divisors and nilpotent elements in \( Z_n \).
   (c) Determine all units, zero divisors and nilpotent elements in \( Z_n[i] \).

2.3 Maximal Ideals and Prime Ideals

We have learned that a ring \( R \) has two trivial ideals, namely \( \{0\} \) and \( R \) itself. In this section, we shall discover properties of maximal ideals and prime ideals. These are two kinds of important ideals in commutative algebra and algebraic geometry.

2.3.1. Definition. An ideal \( M \) of \( R \) is maximal if \( M \neq R \) and for every ideal \( J \) of \( R \),
\[ M \subsetneq J \subsetneq R \Rightarrow J = M \text{ or } J = R. \]

2.3.2. Example. In the ring \( \mathbb{Z}_n \), for \( n \in \mathbb{N} \), \( n\mathbb{Z} \) is maximal if and only if \( n \) is a prime.

Proof. Let \( n \) be a prime and let \( J \) be an ideal of \( \mathbb{Z} \) such that \( n\mathbb{Z} \subseteq J \subseteq \mathbb{Z} \). Then \( J = d\mathbb{Z} \) for some \( d \in \mathbb{N} \) and \( d \mid n \), so \( d = 1 \) or \( d = n \). Hence, \( J = n\mathbb{Z} \) or \( J = \mathbb{Z} \). On the other hand, assume that \( n = ab \) for some \( 1 < a, b < n \). Then \( n\mathbb{Z} \subseteq a\mathbb{Z} \subseteq \mathbb{Z}, a\mathbb{Z} \neq n\mathbb{Z} \) and \( a\mathbb{Z} \neq \mathbb{Z} \), so \( n\mathbb{Z} \) is not maximal. \( \square \)

2.3.3. Remarks. 1. Every ideal \( I \neq R \) is contained in some maximal ideal \( M \).

Proof. Let \( \mathcal{I} = \{ J : J \neq R \text{ and } J \text{ is an ideal of } R \text{ containing } I \} \). Let \( \mathcal{C} = \{ J_\alpha \}_{\alpha \in A} \) be a chain in \( \mathcal{I} \). Then \( \bigcup \mathcal{C} \) is an ideal of \( R \). If \( \bigcup \mathcal{C} = R \), then \( 1 \in J_\alpha \) for some \( \alpha \in A \), so \( J_\alpha = R \), a contradiction. Hence, \( \bigcup \mathcal{C} \) is an upper bound of \( \mathcal{C} \) in \( \mathcal{I} \). By Zorn's lemma, we have \( \mathcal{I} \) has a maximal element which turns out to be our desired maximal ideal containing \( I \). \( \square \)
2. If $M$ is a maximal ideal and $I$ is an ideal of $R$ such that $I \nsubseteq M$, then $M + I = R$.

Proof. Let $x \in I, \notin M$. Consider the ideal $J = M + Rx$ which is larger than $M$. Since $M$ is maximal, $J = R$. Thus, $R = M + Rx \subseteq M + I$. \qed

3. If $M_1$ and $M_2$ are distinct maximal ideals, then $M_1 + M_2 = R$. In addition, if $R$ is commutative, then $M_1M_2 = M_1 \cap M_2$.

4. If $R$ is commutative, then $Ra = R$ if and only if $a$ is a unit.

2.3.4. Theorem. Let $R$ be a commutative ring and $M$ an ideal of $R$. Then $M$ is a maximal ideal of $R$ if and only if $R/M$ is a field.

Proof. Clearly, $R/M$ is a commutative ring with unity $1 + M$. Assume that $M$ is a maximal ideal. Let $a \notin M$. Then $M + Ra = R$, so $\exists b \in R, 1 = m + ba$. Thus, $1 + M = ba + M = (b + M)(a + M)$, and hence $R/M$ is a field. Conversely, suppose that $R/M$ is a field. Let $M \subseteq J \subseteq R$ and $J \neq M$. Then $\exists a \in J \setminus M$. Since $R/M$ is a field and $a \notin M$, $\exists b \in R, 1 + M = (a + M)(b + M) = ab + M$, so $1 - ab \in M \subseteq J$. Since $a \in J, ab \in J$ which implies $1 \in J$. Hence, $J = R$. \qed

2.3.5. Definition. An ideal $P$ of $R$ is prime if $P \neq R$ and for any ideals $A, B$ of $R$, $AB \subseteq P \Rightarrow A \subseteq P$ or $B \subseteq P$.

2.3.6. Theorem. Let $P$ be an ideal of $R$ such that $P \neq R$.

1. If $\forall a, b \in R, ab \in P \Rightarrow a \in P$ or $b \in P$, then $P$ is prime.

2. If $R$ is commutative and $P$ is prime, then $\forall a, b \in R, ab \in P \Rightarrow a \in P$ or $b \in P$.

Proof. (1) Assume that $AB \subseteq P$ and $A \nsubseteq P$. Choose $a \in A, \notin P$. Let $b \in B$. Thus, $ab \in AB \subseteq P$, so $a \in P$ or $b \in P$. But $a \notin P$, hence $B \subseteq P$.

(2) Let $a, b \in R$ be such that $ab \in P$. Since $R$ is commutative, $Rab = RaRb \subseteq P$, so $Ra \subseteq P$ or $Rb \subseteq P$. Hence, $a \in P$ or $b \in P$. \qed

2.3.7. Theorem. Let $R$ be a commutative ring and $P$ an ideal of $R$. Then $P$ is a prime ideal of $R$ if and only if $R/P$ is an integral domain.

Proof. This follows from Theorem 2.3.6 as follows. For an ideal $P$,

$P$ is prime $\iff \forall a, b \in R, ab \in P \Rightarrow a \in P$ or $b \in P$

$\iff \forall a, b \in R, (a + P)(b + P) = 0 + P \Rightarrow a + P = 0 + P$ or $b + P = 0 + P$

$\iff R/P$ is an integral domain

as desired. \qed

Theorems 2.3.4 and 2.3.7 are the most useful for characterizing maximal ideals and prime ideals in commutative rings.

2.3.8. Corollary. Let $R$ be a commutative ring.

1. Every maximal ideal of $R$ is prime.

2. If $R$ is finite, then every prime ideal of $R$ is maximal.

2.3.9. Example. In $\mathbb{Z}$, $n\mathbb{Z}$ is prime if and only if $n = 0$ or $n$ is a prime.

2.3.10. Remark. In the ring $\mathbb{Z}$, $\{0\}$ is a prime ideal which is not maximal.
2.3.11. Definition. The set of all prime ideals of a commutative ring \( R \) is denoted by \( \text{Spec } R \), called the spectrum of \( R \). E.g., \( \text{Spec } \mathbb{Z} = \{ p\mathbb{Z} : p \text{ is a prime} \} \cup \{ \{0\} \} \). A local ring is a commutative ring which has a unique maximal ideal.

2.3.12. Examples.  
1. \( \mathbb{Z} \) has infinitely many maximal ideals of the form \( p\mathbb{Z} \) where \( p \) is a prime, so it is not a local ring.
2. Every field is a local ring with maximal ideal \( \{0\} \).
3. \( \mathbb{Z}_{p^n} \) is a local ring with the maximal ideal \( p\mathbb{Z}_{p^n} \) for all primes \( p \) and \( n \in \mathbb{N} \).

2.3.13. Theorem. Let \( R \) be a commutative ring. Then \( R \) is a local ring if and only if the nonunits of \( R \) form an ideal.

Proof. Assume \( R \) is a local ring with the maximal ideal \( M \). Let \( a \in R \setminus M \). If \( aR \neq R \), then \( aR \) is contained in some maximal ideal, so \( aR \subseteq M \) which yields a contradiction. Thus, \( aR = R \), so \( a \) is a unit. Hence, \( M \) is the set of nonunits of \( R \). Conversely, suppose that the nonunits of \( R \) form an ideal \( M \) of \( R \). Clearly, \( M \) is maximal. Let \( M' \) be another maximal ideal of \( R \). If \( \exists a \in M' \setminus M \), then \( a \) is a unit, so \( M' = R \), a contradiction. Thus, \( M' \subseteq M \). Since \( M' \) is maximal, \( M' = M \). \( \square \)

2.3.14. Corollary. In a finite local ring \( R \), every element is either a unit or a nilpotent element.

2.3.15. Example. Fix a prime \( p \) and let \( \mathbb{Z}_p = \{ m/n : m, n \in \mathbb{Z} \text{ and } p \text{ does not divide } n \} \).

Then \( \mathbb{Z}_p \) is a subring of \( \mathbb{Q} \) and is local. Its unique maximal ideal is \( \{ pk/n : k, n \in \mathbb{Z} \text{ and } p \nmid n \} \).

2.3 Exercises.  
1. Let \( R \) be a ring and \( I \) an ideal of \( R \). Prove that the map \( J \mapsto J/I \) gives a 1-1 correspondence \( \{ \text{ideals of } R \text{ containing } I \} \leftrightarrow \{ \text{ideals of } R/I \} \).

Moreover, this correspondence carries maximal ideals to maximal ideals.

2. Prove Corollary 2.3.14 and Example 2.3.15.

3. Find all ideals, all prime ideals and all maximal ideals of  
   (a) \( \mathbb{Z}_{12} \)  
   (b) \( \mathbb{Z}_2 \times \mathbb{Z}_4 \)  
   (c) \( \mathbb{Q} \times \mathbb{Q} \)  
   (d) \( \mathbb{Q} \times \mathbb{Z} \)  
   (e) \( \mathbb{Z} \times \mathbb{Z}_4 \times \mathbb{Z}_5 \).

4. Let \( R \) be a commutative ring. If every ideal proper of \( R \) is prime, show that \( R \) is a field.

5. Show that in a Boolean ring \( R \), every prime ideal \( P \neq R \) is maximal.

6. Let \( R \) be a commutative ring and \( b \in R \) a nilpotent element. Prove that \( u + b \) is a unit for all units \( u \) in \( R \).

11. Project. (Chain ring) A ring is called a chain ring if all its ideals form a chain under inclusion. For example, \( \mathbb{Z}_{p^n}, p \text{ a prime and } n \in \mathbb{N} \), is a chain ring. Also, every field is a chain ring. Let \( R \) be a finite commutative ring. Prove that \( R \) is a chain ring if and only if \( R \) is a local ring whose maximal ideal is principal.

A finite chain ring arises in algebraic number theory as quotient rings of rings of integers in number fields. It has many applications in coding theory because of the similarity with finite fields. Galois rings in Project 16 are examples for this situation.

2.4 Factorizations

From elementary number theory, we know that every positive integer can be decomposed uniquely into a product of prime numbers (Theorem 1.2.9). It is the unique factorization property of the ring \( \mathbb{Z} \). In this section, we shall learn about factorizations in any other integral domains.
2.4.1 Irreducible Elements and Prime Elements

2.4.1. Definition. Let \( R \) be a commutative ring and suppose that \( a, b \in R \). We say that \( a \) divides \( b \) and write \( a \mid b \), if there is an \( r \in R \) such that \( ra = b \).

This definition coincides the divisibility discussed previously in Section 1.2.

2.4.2. Remarks. Let \( R \) be a commutative ring and \( a, b \in R \).

1. \( a \) divides \( b \iff b \in Ra \iff Rb \subseteq Ra \).
2. \( a \) divides 0 (\( R0 \subseteq Ra \)).
3. \( a \in R \), 1 divides \( a \) (\( Ra \subseteq R \cdot 1 = R \)).
4. \( a \) divides 1 \( \iff R = Ra \iff a \) is a unit.
5. 0 divides \( a \) \( \iff Ra \subseteq R0 \iff a = 0 \).

2.4.3. Definition. Let \( R \) be an integral domain and suppose \( a, b \in R \). We say that \( a \) and \( b \) are associates if \( a \mid b \) and \( b \mid a \).

2.4.4. Theorem. Let \( R \) be an integral domain, \( a, b \in R \). The following statements are equivalent.

(i) \( a \) and \( b \) are associates. (ii) \( Ra = Rb \). (iii) \( a = ub \) for some unit \( u \in R \).

Proof. (i) \( \Rightarrow \) (iii) If \( a = 0 \), then \( b = 0 \) and (3) is clear. Suppose then that \( a \neq 0 \). Since \( a \mid b \) and \( b \mid a \), we can write \( a = ub \) and \( b = va \). Thus, \( a = ub = uva \), so \( (uv - 1)a = 0 \), so \( uv = 1 \). Hence, \( a = ub \) and \( u \) is a unit of \( R \).

(iii) \( \Rightarrow \) (ii) If \( a = ub \) where \( u \) is a unit, then \( Ra = Rub = (Ru)b = Rb \).

(ii) \( \Rightarrow \) (i) If \( Ra = Rb \), then \( a = rb, b = sa \), so \( b \mid a \) and \( a \mid b \). Hence, \( a \) and \( b \) are associates. \( \square \)

2.4.5. Definition. Let \( R \) be an integral domain. We say that a nonzero nonunit element \( a \) in \( R \) is an irreducible element or atom if \( a \) cannot be expressed as a product \( a = bc \) where \( b \) and \( c \) are nonunits.

For example, in \( \mathbb{Z} \), \( p \) and \( -p, p \) a prime number, are irreducible elements.

2.4.6. Theorem. Let \( R \) be an integral domain and \( a \) a nonzero nonunit in \( R \).

1. \( a \) is irreducible \( \iff (\forall b, c \in R, a = bc \Rightarrow b \text{ or } c \text{ is a unit}) \).
2. If \( Ra \) is maximal, then \( a \) is irreducible. The converse holds if \( R \) is a PID.

Proof. (1) It follows directly from the definition.

(2) Assume that \( Ra \) is maximal. Let \( b, c \in R \) be such that \( a = bc \). Then \( Ra \subseteq Rb \subseteq R \). Since \( Ra \) is maximal, \( Ra = Rb \) or \( Rb = R \). If \( Rb = R \), then \( b \) is a unit. Let \( Ra = Rb \). Then \( a = bu \) for some unit \( u \in R \), so \( bc = bu \) which implies \( c = u \) is a unit since \( R \) has no zero divisor.

Finally, we assume that \( R \) is a PID and \( a \in R \) is irreducible. Let \( J \) be an ideal of \( R \) such that \( Ra \subseteq J \subseteq R \). Since \( R \) is a PID, \( J = Rb \) for some \( b \) in \( R \), and so \( a \in Rb \). Thus, \( a = cb \) for some \( c \in R \), so \( b \) or \( c \) is a unit. Hence, \( Rb = R \) or \( Ra = Rb \). \( \square \)

2.4.7. Definition. A nonzero nonunit element \( p \) in \( R \) is a prime element if

\[ \forall a, b \in R, p \mid ab \Rightarrow p \mid a \text{ or } p \mid b. \]

Note that a prime number is a prime element in \( \mathbb{Z} \) by Corollary 1.2.8 (2).
2.4.8. Theorem. Let $R$ be an integral domain.

(1) For a nonzero nonunit $p$ in $R$, $p$ is prime $\iff R_p$ is a prime ideal.

(2) Every prime element is irreducible. The converse holds if $R$ is a PID.

Proof. (1) It follows directly from the definition and Theorem 2.3.6.

(2) Let $p$ be a prime element. Assume that $p = ab$ for some $a, b \in R$. Then $Rab = Rp$, so $Ra \subseteq Rp$ or $Rb \subseteq Rp$. Since $Rp = Rab \subseteq (Ra \cap Rb)$, $Ra = Rp$ or $Rb = Rp$, so $au = p$ or $bv = p$ for some units $u$ and $v$ in $R$. Hence, $b = u$ or $a = v$ is a unit in $R$. Finally, suppose that $R$ is a PID and $p$ is irreducible. Then $Rp$ is maximal, so it is a prime ideal. Hence, $p$ is prime.

\[\Box\]

2.4.2 Unique Factorization Domains

2.4.9. Definition. A unique factorization domain (UFD) is an integral domain $R$ which satisfies:

1. Every nonzero nonunit of $R$ is a product of atoms.
2. If $a$ is a nonzero nonunit of $R$, then the expression of $a$ as a product of atoms is unique in the following sense: "If $a = a_1 \ldots a_r = b_1 \ldots b_s$ where $a_1, \ldots, a_r, b_1, \ldots, b_s$ are atoms, then $r = s$ and there is a reordering $b_{i_1}, \ldots, b_{i_r}$ of $b_1, \ldots, b_s$ such that $a_1$ and $b_{i_1}$ are associates, $a_2$ and $b_{i_2}$ are associates, $\ldots$, $a_r$ and $b_{i_r}$ are associates".

2.4.10. Examples. 1. The ring of rational integers $\mathbb{Z}$ is a UFD by the fundamental theorem of arithmetic (Theorem 1.2.9). Since $\mathcal{U}(\mathbb{Z}) = \{\pm 1\}$, the atoms of $\mathbb{Z}$ are $\pm p$ where $p$ is a prime.

Note that $p$ and $-p$ are associates (e.g., $12 = 2 \cdot 2 \cdot 3 = (-2)(-3) \cdot 2$).

2. Let $F$ be a field. Every element of $F$ except 0 is a unit. Hence, every nonzero nonunit of $F$ is uniquely a product of atoms (vacuously!). That is, $F$ has no nonzero nonunits.

2.4.11. Theorem. Let $R$ be an integral domain. Then $R$ is a UFD if and only if

1. every nonzero nonunit of $R$ is a product of atoms and
2. every irreducible element is prime.

Proof. Suppose $R$ is a UFD. Then (1) holds, by the definition of a UFD. It remains to show that if $x$ is irreducible, then $x$ is prime. Suppose $x \mid bc$, and let $ax = bc$. If $b$ and $c$ are nonunit, write $b$ and $c$ as products of atoms, so that

$$ax = b_1 \ldots b_{i_1} c_1 \ldots c_{i_r}.$$ 

Since $x$ is an atom, $x$ must be an associate of some $b_i$ or some $c_j$. Hence, $x \mid b$ or $x \mid c$. Thus, $x$ is prime.

Conversely, suppose (1) and (2) are given. Then to show $R$ is a UFD, it suffices to show that if

$$a_1 \ldots a_r = b_1 \ldots b_s$$

where the $a_i$ and $b_i$ are atoms, then $r = s$ and the $b_i$ may be arranged so the $a_i$ and $b_i$ are associates for $i = 1, \ldots, r$. The proof proceeds by induction on $r$.

When $r = 1$, $a_1 = b_1 \ldots b_s$. Since $a_1$ is prime, $a_1$ divides $b_i$ for some $i$. Assume that $a_1 \mid b_1$, and let $b_1 = u a_1$. Since $b_1$ is an atom, $u$ must be a unit, so $a_1$ and $b_1$ are associates. Furthermore, $a_1 = b_1 \ldots b_s = u a_1 b_2 \ldots b_s$, so $1 = u b_2 \ldots b_s$. That is, $s = 1$ and $a_1 = b_1$. For the inductive step, write $a_1 \ldots a_r = b_1 \ldots b_s$. Since $a_1$ is prime, $a_1$ divides $b_i$ for some $i$. As above, let $b_1 = u a_1$ where $u$ is a unit and $a_1$ and $b_1$ are associates. Then $a_1 \ldots a_r = b_1 \ldots b_s = u a_1 b_2 \ldots b_s$, so $a_2 \ldots a_r = u b_2 \ldots b_s$. Now the inductive hypothesis applies since we have $r - 1$ factors on the left. It follows that $r = s$ and after reordering the $b_i$, $a_i$ and $b_i$ are associates for $i = 2, \ldots, r$. This completes the induction.

\[\Box\]
To obtain more examples of a UFD and an integral domain which is not a UFD, we introduce:

2.4.12. Definition. Let $d$ be a square free integer. The set

$$\mathbb{Z}[\sqrt{d}] = \{x + y\sqrt{d} : x, y \in \mathbb{Z}\}$$

is a subring of $\mathbb{C}$. It is called the **ring of quadratic integers**. Note that if $x_1, x_2, y_1, y_2 \in \mathbb{Z}$ are such that $x_1 + y_1\sqrt{d} = x_2 + y_2\sqrt{d}$, then $x_1 = x_2$ and $y_1 = y_2$ because $d$ is non-square. Define a function $N : \mathbb{Z}[\sqrt{d}] \to \mathbb{Z}$ by

$$N(x + y\sqrt{d}) = x^2 - dy^2 \quad \text{for all } x, y \in \mathbb{Z}.$$ 

It is called the **norm map** on $\mathbb{Z}[\sqrt{d}]$.

2.4.13. Theorem. 1. If $\alpha \in \mathbb{Z}[\sqrt{d}]$ and $N(\alpha) = 0$, then $\alpha = 0 = 0 + 0\sqrt{d}$.
2. $\forall \alpha, \beta \in \mathbb{Z}[\sqrt{d}], N(\alpha\beta) = N(\alpha)N(\beta)$ and $(\alpha | \beta \iff N(\alpha) | N(\beta))$.
3. $\forall \alpha \in \mathbb{Z}[\sqrt{d}], \alpha$ is a unit $\iff N(\alpha) = \pm 1$.
4. If $\alpha \in \mathbb{Z}[\sqrt{d}]$ and $N(\alpha) = p$ is a prime number, then $\alpha$ is irreducible in $\mathbb{Z}[\sqrt{d}]$.

**Proof.** Let $x, y \in \mathbb{Z}$ be such that $x^2 - dy^2 = 0$. Then $x^2 = dy^2$. If $y \neq 0$, then $d = x^2/y^2$, so $\sqrt{d} = |x/y| \in \mathbb{Q}$, which is a contradiction. Thus, we must have $y = 0$ which also forces $x = 0$. This proves (1). A direct calculation gives (2). For (3), let $\alpha \in \mathbb{Z}[\sqrt{d}]$. Suppose that $\alpha$ is a unit. Then $\alpha\beta = 1$ for some $\beta \in \mathbb{Z}[\sqrt{d}]$. Thus, $1 = N(1) = N(\alpha\beta) = N(\alpha)N(\beta)$, so $N(\alpha)$ divides 1 in $\mathbb{Z}$. This gives $N(1) = \pm 1$. Conversely, assume that $N(\alpha) = \pm 1$. Write $\alpha = x + y\sqrt{d}$. Then $\pm 1 = N(x + y\sqrt{d}) = x^2 - y^2d = (x + y\sqrt{d})(x - y\sqrt{d})$ which implies that $x + y\sqrt{d}$ is a unit. Finally, (4) follows from (3).

2.4.14. Example. The unit group of the ring $\mathbb{Z}[i]$ is $\{1, -1, i, -i\}$ where $i$ denotes $\sqrt{-1}$.

2.4.15. Remark. The equation $x^2 - dy^2 = 1$ is called the **Pell’s equation**. Every unit in $\mathbb{Z}[\sqrt{d}]$ is a solution of Pell’s equation, or else of $x^2 - dy^2 = -1$, the **negative Pell’s equation**. If $d < 0$, then $x^2 - dy^2 \geq 0$. In this case the negative Pell’s equation has no solutions. In fact, Pell’s equation only has very few solutions in this case, namely two, unless $d = -1$ when there are four solutions. If $d > 0$, there are infinitely many solutions to Pell’s equation. The negative Pell’s equation may or may not have solutions.

2.4.16. Example. Consider the ring $\mathbb{Z}[\sqrt{-5}]$.

1. $\mathbb{Z}[\sqrt{-5}]$ is not a UFD by Theorem 2.4.11.
2. $1 + \sqrt{-5}$ and $2$ are not prime elements. Hence, $\mathbb{Z}[\sqrt{-5}]$ is not a UFD by Theorem 2.4.11.

**Solution.** (1) Assume that $1 + \sqrt{-5} = (a + b\sqrt{-5})(c + d\sqrt{-5})$ for some $a, b, c, d \in \mathbb{Z}$. By taking norms, we have

$$6 = (a^2 + 5b^2)(c^2 + 5d^2),$$

which implies that $a^2 + 5b^2 = 1, 2, 3$ or 6. Observe that $b = 0$ implies $a^2 = 1$, so $a + b\sqrt{-5}$ is a unit. If $b \neq 0$, then $a^2 + 5b^2 \geq 5$, so $a^2 + 5b^2 = 6$. This forces that $c^2 + 5d^2 = 1$ and thus $c + d\sqrt{-5}$ is a unit. Hence, $1 + \sqrt{-5}$ is irreducible. Next, assume that $2 = (a + b\sqrt{-5})(c + d\sqrt{-5})$ for some $a, b, c, d \in \mathbb{Z}$. By taking norms, we have

$$4 = (a^2 + 5b^2)(c^2 + 5d^2),$$

which implies that $a^2 + 5b^2 = 1$ or $a^2 + 5b^2 = 4$. If $a^2 + 5b^2 = 2$, then 2 is a square modulo 5 which is a contradiction. Thus, $a^2 + 5b^2 = 1$ or $c^2 + 5d^2 = 1$. Hence, $a + b\sqrt{-5}$ or $c + d\sqrt{-5}$ is a unit and so
2 is irreducible. Similarly, 1 + \sqrt{-5} and 3 are irreducible.

(2) Note that

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

Then \(1 + \sqrt{-5} \mid 2 \cdot 3\). But, if \((1 + \sqrt{-5}) \mid 2\) or \((1 + \sqrt{-5}) \mid 3\), then 6 \mid 4 or 6 \mid 9, which are absurd. Thus, 1 + \sqrt{-5} is not a prime element. Similarly, 2 \mid (1 + \sqrt{-5})(1 - \sqrt{-5})\). If 2 \mid (1 + \sqrt{-5}) or 2 \mid (1 - \sqrt{-5}), then 4 \mid 6, a contradiction. Hence, 2 is not a prime element.

Next, we talk about common factors, gcd and lcm of elements in an integral domain (cf. Section 1.2).

**2.4.17. Definition.** Let \(R\) be an integral domain and suppose \(a, b \in R\). A greatest common divisor of \(a\) and \(b\), gcd\((a, b)\), is an element \(d \in R\) which satisfies

1. \(d \mid a\) and \(d \mid b\) and
2. \(\forall c \in R, (c \mid a \land c \mid b) \Rightarrow c \mid d\).

A least common multiple of \(a\) and \(b\), lcm\((a, b)\), is an element \(m \in R\) which satisfies

1. \(a \mid m\) and \(b \mid m\) and
2. \(\forall c \in R, (a \mid c \land b \mid c) \Rightarrow m \mid c\).

**2.4.18. Remark.** +3 and −3 are greatest common divisors of 12 and 15. 60 and −60 are least common multiples of 12 and 15. Thus, the gcd or lcm of two elements is not unique, (however we adopt the above notation anyway, e.g., gcd(12, 15) = 3 and gcd(12, 15) = −3 are both correct!). By their definitions, they are unique up to associates as recorded in the next theorem.

**2.4.19. Theorem.** Let \(R\) be an integral domain and let \(a, b \in R\).

1. If \(d\) and \(d'\) are gcd's of \(a\) and \(b\), then \(d\) and \(d'\) are associates.
2. If \(m\) and \(m'\) are lcm's of \(a\) and \(b\), then \(m\) and \(m'\) are associates.

Let \(R\) be an integral domain and let \(\mathbb{Q}(R)\) be the set of atoms of \(R\). Define an equivalence relation on \(\mathbb{Q}(R)\) by \(a \sim b\) if \(a\) and \(b\) are associates. Then a set of representative atoms for \(R\) is a set \(\mathcal{P} = \mathcal{P}(R)\) which contains exactly one atom from each equivalence class.

**2.4.20. Example.** \(\mathbb{Q}(\mathbb{Z}) = \{\pm p \mid p\text{ is a prime}\}\) is the set of all atoms in \(\mathbb{Z}\).

\(\mathcal{P}(\mathbb{Z}) = \{p \mid p\text{ is a positive prime}\}\) is a set of representative atoms.

\(\mathcal{P}(\mathbb{Z}) = \{+2, -3, +5, -7, \ldots\}\) is another set of representative atoms.

We obtain the next theorem directly from the definition of a UFD.

**2.4.21. Theorem.** Let \(R\) be an integral domain and let \(\mathcal{P}\) be a set of representative atoms for \(R\). Then the following statements are equivalent.

(i) \(R\) is a UFD.

(ii) Every nonzero element of \(R\) can be expressed uniquely (up to order of factors) as \(a = u b_1^{i_1} \cdots b_k^{i_k}\), where \(u\) is a unit of \(R\), \(k \geq 0\), \(i_1, \ldots, i_k > 0\) and \(b_1, \ldots, b_k\) are distinct elements of \(\mathcal{P}\).

Another important result from \(R\) being a UFD is the existence of gcd and lcm for any pair of nonzero elements. We also have the same relation for gcd and lcm as in elementary number theory.

**2.4.22. Theorem.** Let \(R\) be a UFD and suppose \(a, b \in R \setminus \{0\}\).

1. \(a\) and \(b\) have a gcd and an lcm.
2. Let \(\mathcal{P}\) be a set of representative atoms for \(R\). Then among the gcd's of \(a\) and \(b\) there is exactly one which is a product of elements of \(\mathcal{P}\). The same is true for the lcm's of \(a\) and \(b\).
3. If \(a\) and \(b\) are nonzero, \(\text{gcd}(a, b) = r\), and \(\text{lcm}(a, b) = s\), then \(ab\) and \(rs\) are associates. In other words,

\[
\text{lcm}(a, b) = \frac{ab}{\text{gcd}(a, b)}.
\]
2.4. Factorizations

Proof. Let \( P \) be a set of representative atoms of \( R \), and let \( b_1, \ldots, b_k \in P \) be all the atoms of \( P \) which occur in either \( a \) or \( b \) when they are factored as in Theorem 2.4.21. Write

\[
a = ub_1^{i_1} \cdots b_k^{i_k} \quad \text{and} \quad b = vb_1^{j_1} \cdots b_k^{j_k}
\]

where \( u \) and \( v \) are units and \( i_s, j_s \geq 0 \). Then we derive:

(a) \( r = b_1^{\min(i_1, j_1)} \cdots b_r^{\min(i_r, j_r)} \) is a gcd for \( a \) and \( b \).
(b) \( s = b_1^{\max(i_1, j_1)} \cdots b_r^{\max(i_r, j_r)} \) is a lcm for \( a \) and \( b \).
(c) \( r \) is the only gcd of \( a \) and \( b \) which is a product of elements of \( P \), and \( s \) is the only lcm of \( a \) and \( b \) which is a product of elements of \( P \).
(d) Since \( i + j = \min(i, j) + \max(i, j) \) for any integers \( i \) and \( j \),

\[
ab = uvb_1^{i_1+j_1} \cdots b_k^{i_k+j_k} = uvsr.
\]

Hence, \( ab \) and \( rs \) are associates. \( \square \)

2.4.23. Remark. Suppose \( R \) is an integral domain and \( Ra + Rb = Rc \). Then \( c = \gcd(a, b) \). The converse does not hold. E.g., \( \mathbb{Q}[s, t] \), where \( s \) and \( t \) are indeterminates. Then \( \gcd(s, t) = 1 \) and \( \mathbb{Q}[s, t] = \mathbb{Q}[s, t]s + \mathbb{Q}[s, t]t \).

Proof. Since \( Rc \supseteq Ra \) and \( Rc \supseteq Rb \), \( c \mid a \) and \( c \mid b \). Suppose \( d \mid a \) and \( d \mid b \). Then \( Rd \supseteq Ra + Rb = Rc \), so \( d \mid c \). Hence, \( c = \gcd(a, b) \). \( \square \)

Now, we shall prove that a PID is also a UFD. The following lemma is a key for \( R \) being a PID.

2.4.24. Lemma. [Ascending Chain Condition (ACC) for a PIR] Let \( R \) be a principal ideal ring. If \( I_1 \subseteq I_2 \subseteq \cdots \) is a chain of ideals in \( R \), then \( \exists m \in \mathbb{N}, I_n = I_m \) for all \( n \geq m \).

Proof. Let \( I = \bigcup_{n=1}^{\infty} I_n \). Then \( I \) is an ideal of \( R \). Since \( R \) is a PIR, \( \exists a \in R, (a) = I \). Then \( a \in \bigcup_{n=1}^{\infty} I_n \), so \( \exists m \in \mathbb{N}, a \in I_m \). Thus, \( I = (a) \subseteq I_m \subseteq I \) which implies that \( I_m = I \). Hence, \( \forall n \geq m, I_n = I_m \). \( \square \)

2.4.25. Lemma. If \( R \) is a PID and \( a \) is a nonzero nonunit element in \( R \), then there exists an atom \( p \in R \) such that \( p \mid a \).

Proof. Since \( a \) is nonunit, \( Ra \not\subseteq R \). Then there exists a maximal ideal \( M \) of \( R \) such that \( Ra \subseteq M \). Since \( R \) is a PID, \( M = Rp \) for some atom \( p \) by Theorem 2.4.6. Since \( Ra \subseteq Rp, p \mid a \). \( \square \)

2.4.26. Theorem. Every PID is a UFD.

Proof. Let \( R \) be a PID. By Theorems 2.4.8 and 2.4.11, it suffices to show that every nonzero nonunit of \( R \) is a product of atoms. Let \( a \in R \) be nonzero nonunit. By Lemma 2.4.25, there exists an atom \( p_1 \) dividing \( a \). Write \( a = p_1b_1 \) for some \( b_1 \in R \). If \( b_1 \) is a unit, then \( a \) is an atom. If \( b_1 \) is nonunit, then there exists an atom \( p_2 \) dividing \( b_1 \), so we write \( a = p_1b_1 = p_1p_2b_2 \). Continuing, we get a strictly ascending chain of ideals

\[
(a) \subset (b_1) \subset (b_2) \subset \cdots
\]

Since \( R \) is a PID, this chain must terminate, by the ACC in Lemma 2.4.24, with some \( b_r = p_rp_r \), where \( u_r \) is a unit and \( p_r \) is an atom. Hence, \( a = p_1p_2 \cdots p_rp_r \), and so \( R \) is a UFD as desired. \( \square \)

Finally, we study a generalization of the division algorithm which leads to a special kind of integral domains.
2.4.27. Definition. An integral domain $D$ is called a Euclidean domain if there exists a map

$$d : D \setminus \{0\} \to \mathbb{N} \cup \{0\},$$

called a valuation map, such that

1. $\forall a, b \in D \setminus \{0\}, d(a) \leq d(ab)$ and
2. $\forall a \in D, b \in D \setminus \{0\}, \exists q, r \in D, a = bq + r$ with $r = 0$ or $d(r) < d(b)$.

2.4.28. Examples.

1. Any field $F$ is a Euclidean domain with valuation $d(a) = 1$ for all $a \neq 0$.
2. From the division algorithm for $\mathbb{Z}$ (Theorem 1.2.1), we have $\mathbb{Z}$ is a Euclidean domain if we define $d(a) = |a|$ for all $a \neq 0$.
3. The ring $\mathbb{Z}[i] = \{m + ni : m, n \in \mathbb{Z}\}$ is called the ring of Gaussian integer. This is a subring of $\mathbb{C}$, hence an integral domain. Its elements can be identified with the set of “lattice points”, that is, points with integral coordinates in the complex plane. If $a = m + ni$, we put $d(a) = a\bar{a} = |a|^2 = m^2 + n^2$, the norm map. Then $d(a) \in \mathbb{N}$ and $d(ab) = d(a)d(b) \geq d(a)$ for all $a, b \in \mathbb{Z}[i] \setminus \{0\}$. To prove that $d$ satisfies the condition of the definition of a Euclidean domain, we note that if $b \neq 0$, then $ab^{-1} = \mu + \nu i$, where $\mu$ and $\nu$ are rational numbers. Now we can find integers $u$ and $v$ such that $|u - \mu| \leq 1/2, |v - \nu| \leq 1/2$. Set $\varepsilon = \mu - u, \eta = \nu - v$, so that $|\varepsilon| \leq 1/2$ and $|\eta| \leq 1/2$. Then

$$a = b[(u + \varepsilon) + (v + \eta)i] = bq + r$$

where $q = u + vi$ is in $\mathbb{Z}[i]$. Since $r = a - bq, r \in \mathbb{Z}[i]$. Moreover if $r \neq 0$, then

$$d(r) = |r|^2 = |b|^2(|\varepsilon|^2 + |\eta|^2) \leq |b|^2(1/4 + 1/4) = d(b)/2.$$

Thus, $d(r) < d(b)$. Hence, $\mathbb{Z}[i]$ is a Euclidean domain.

2.4.29. Theorem. A Euclidean domain is a PID, and hence is a UFD.

Proof. Let $I$ be an ideal in a Euclidean domain $D$. If $I = \{0\}$, we have $I = (0)$. Otherwise, let $b \neq 0$ be an element of $I$ for which $d(b)$ is minimal for the nonzero elements of $I$. Let $a$ be any element of $I$. Then $a = bq + r$ for some $q, r \in D$ with $r = 0$ or $d(r) < d(b)$. Since $r = a - bq \in I$ and $d(r) < d(b)$, we must have $r = 0$ by the choice of $b$ in $I$. Hence, $a = bq$, so $I = (b)$.

2.4.30. Example. Let $\theta = \frac{1}{2}(1 + \sqrt{-19})$ and $\mathbb{Z}[\theta] = \{a + b\theta : a, b \in \mathbb{Z}\}$. Assume that $u = a + b\theta$ is a unit in $\mathbb{Z}[\theta]$. Then $(a + b\theta)(c + d\theta) = 1$ for some $c, d \in \mathbb{Z}$. The squares of absolute value on both sides give

$$(2a + b)^2 + 19b^2)((2c + d)^2 + 19d^2) = 16$$

which implies $b = d = 0$ and so $ac = 1$. Hence, the unit group of $\mathbb{Z}[\theta] = \{\pm 1\}$. By a similar technique, we can show that 2 and 3 are irreducible in $\mathbb{Z}[\theta]$. Now, suppose that $d$ is a valuation map on $\mathbb{Z}[\theta]$. Choose $m \in \mathbb{Z}[\theta]$ which is nonzero nonunit such that $d(m)$ is minimal. First, we divide 2 by $m$ and get $q, r \in \mathbb{Z}[\theta]$ and

$$2 = mq + r \quad \text{with} \quad d(r) < d(m) \quad \text{or} \quad r = 0.$$

This means $r = 0, 1$ or $-1$. If $r = 0$, then $m | 2$ which forces $m = \pm 2$ since 2 is irreducible and $m$ is not a unit. Similarly, if $r = -1$, then $m = \pm 3$. The case $r = 1$ cannot happen, for if it did, then $m | 1$, so $m$ is a unit. Next, we divide $\theta$ by $m$ in the same way, we obtain $q', r' \in \mathbb{Z}[\theta]$ and

$$\theta = m q' + r' \quad \text{with} \quad d(r') < d(m) \quad \text{or} \quad r' = 0.$$

Again, we have $r' = 0, 1$ or $-1$. Thus, one of $\theta, \theta + 1$ or $\theta - 1$ is divisible by $m$. But $m = \pm 2$ or $\pm 3$ and it is easy to see that none of these quotients is in $\mathbb{Z}[\theta]$. This contradiction tell us that $\mathbb{Z}[\theta]$ is not a Euclidean domain.
2.4. Factorizations

Next, we shall show that \( \mathbb{Z}[\theta] \) is a PID. Let \( I \) be a nonzero ideal of \( \mathbb{Z}[\theta] \). Choose \( b \in I \) so that \( |b| \) is as small as possible. We aim to show that \( I = \mathbb{Z}[\theta]b \). Suppose not. Then there is an element \( a \in I \setminus \mathbb{Z}[\theta]b \). Note that \( ap - bq \in I \) for all \( p, q \in \mathbb{Z}[\theta] \), so if we can find \( p, q \) with \( |ap - bq| < |b| \) (or equivalently \( |(a/b)p - q| < 1 \)), then we shall be done. Since we may replace \( a \) by any element \( a' = a - bq \), we can subtract any desired element of \( R \) from \( a/b \). In particular, we can assume that the imaginary part \( y \) of \( a/b = x + iy \) lies between \( \pm \sqrt{19}/4 \). Now, if the imaginary part of \( a/b \) lies strictly between \( \pm \sqrt{3}/2 \), then \( a/b \) lies at a distance less than \( 1 \) from some rational integer and we are done. Thus, we may assume the imaginary part of \( a/b \) lies between \( \sqrt{3}/2 \) and \( \sqrt{19}/4 \) (or the negative of this, where the argument is similar). Hence, the imaginary part of \( 2(a/b) - (1 + \sqrt{-19})/2 \) lies between \( \sqrt{3} - \sqrt{19}/2 \) and \( 0 \). But \( \sqrt{19} < \sqrt{27} = 3\sqrt{3} \), so \( \sqrt{3}/2 > \sqrt{19}/2 - \sqrt{3} > 0 \). Therefore, the imaginary part of \( 2(a/b) - (1 + \sqrt{-19})/2 \) is sufficient small that the complex number lies at a distance less than \( 1 \) for some rational integer. In both cases, we have found elements \( p, q \in \mathbb{Z}[\theta] \) such that \( |ap - bq| < |b| \) which is a contradiction. Hence, \( \mathbb{Z}[\theta] \) is a PID.

2.4.31. Remark. In conclusion, recall that \( \mathbb{Z} \) is an integral domain which is not a field and \( \mathbb{Z}[\sqrt{-5}] \) is not a UFD. Besides, \( \mathbb{Z}[\theta] = \{a + b\theta : a, b \in \mathbb{Z}\} \), where \( \theta = (1 + \sqrt{-19})/2 \), is a PID which is not a Euclidean domain as shown above. Finally, \( \mathbb{Z}[x] \) (in the next section) is a UFD which is not a PID.

2.4. Exercises.

1. If \( p \) and \( q \) are prime elements in an integral domain \( R \) such that \( p \mid q \), prove that \( p \) and \( q \) are associates.

2. Let \( R \) be a UFD and \( c \) a nonzero element in \( R \). Prove that \( R/RC \) contains a nonzero nilpotent element if and only if there is a prime element \( p \in R \) with \( p^2 \mid c \).

3. Let \( R \) be a UFD. If \( a \in R \) is a nonzero nonunit element, prove that \( Ra \) is the product of a finite number of prime ideals.

4. If \( R \) is a PID and \( \gcd(a, b) = 1 \), show that \( Ra + Rb = R \), so \( 1 = ax + by \) for some \( x, y \in R \).

5. Let \( R \) be a PID and suppose that \( a, b \) and \( c \) are nonzero elements of \( R \) such that \( Ra + Rb = Rc \). Show that there exist \( u, v \in R \) such that \( uax + vb = c \) and \( Rau + Rbv = R \).

6. Prove that \( 4 + \sqrt{10} \) is irreducible but not prime in the ring \( \{a + b\sqrt{10} : a, b \in \mathbb{Z}\} \). Deduce that \( \mathbb{Z}[\sqrt{10}] \) is not a UFD.

7. Show that the ring \( \mathbb{Z}[\sqrt{2}] \) has infinitely many units. (Hint. If \( u \) is a unit, so is \( u^n \) for all \( n \in \mathbb{Z} \).)

8. (a) Let \( D \) be a Euclidean domain. Prove that \( u \) is a unit in \( D \) if and only if \( d(u) = d(1) \).

(b) Show that \( \pm 1 \) and \( \pm i \) are units in \( \mathbb{Z}[i] \) and prove that \( a + bi \) is not a unit in \( \mathbb{Z}[i] \), then \( a^2 + b^2 > 1 \).

9. Let \( R \) be a Euclidean ring and \( a, b \in R \), \( b \neq 0 \). Prove that there exist \( q_0, q_1, \ldots, q_n \) and \( r_1, \ldots, r_n \) in \( R \) such that

\[
\begin{align*}
a &= q_0b + r_1, \\
b &= q_1r_1 + r_2, \\
r_1 &= q_2r_2 + r_3, \\
\cdots & \cdots \\
r_n-2 &= q_{n-1}r_{n-1} + r_n, \\
r_n &\end{align*}
\]

and if \( a \) and \( b \) satisfy the above conditions, then \( r_n \) is a \( \gcd \) of \( a \) and \( b \). This algorithm is called the Euclidean algorithm. Find a \( \gcd \) of \( 8 + 6i \) and \( 5 - 15i \) in \( \mathbb{Z}[i] \) by using the Euclidean algorithm.

10. Let \( D \) be a UFD with field of fractions \( F \) and suppose \( \alpha \in F \). Show that it is possible to write \( \alpha = a/b \) with \( a, b \in D \) and \( \gcd(a, b) = 1 \).

11. Let \( R \) be a PID with field of fractions \( F \), and let \( S \) be a ring with \( R \subseteq S \subseteq F \).

(a) If \( \alpha \in S \), show that \( \alpha = a/b \) with \( a, b \in R \) and \( 1/b \in S \). (b) Prove that \( S \) is a PID.

12. Let \( R = \{m/2^n : m, n \in \mathbb{Z} \text{ and } n \geq 0\} \).

(a) Prove that \( R \) is a subring of \( \mathbb{Q} \) and determine all units of \( R \).

(b) Show that 3 is an irreducible element in \( R \).

(c) Prove that \( R \) is a PID.
12. Project. (Prime elements in the ring of Gaussian integers) We have learned that all units in \( \mathbb{Z}[i] \) are \( \pm 1, \pm i \). In this project, we shall determine all prime elements in \( \mathbb{Z}[i] \). Use the norm map, show that up to multiplication by units, the prime elements in \( \mathbb{Z}[i] \) are of three types:

(a) \( p \), where \( p \) is a prime in \( \mathbb{Z} \) satisfying \( p \equiv 3 \pmod{4} \),
(b) \( \pi \) or \( \bar{\pi} \), where \( q = \pi \bar{\pi} \) is a prime in \( \mathbb{Z} \) satisfying \( q \equiv 1 \pmod{4} \),
(c) \( \alpha = 1 + i \).

13. Project. (Quadratic norm Euclidean domains) Find all square free integers \( d \equiv 2, 3 \pmod{4} \) such that the norm map on \( \mathbb{Z}[\sqrt{d}] \) satisfies the axiom of a Euclidean function. [Answer. They are \(-2, -1, 2, 3, 6, 7, 11, 19, 33\).]

Moreover, for a square free integer \( d \equiv 1 \pmod{4} \), let
\[
\mathbb{Z}
\left\lfloor \frac{1 + \sqrt{d}}{2} \right\rfloor = \left\{ \frac{x + y\sqrt{d}}{2} : x, y \in \mathbb{Z}, a \equiv b \pmod{2} \right\}.
\]

Define the norm map on \( \mathbb{Z}
\left\lfloor \frac{1 + \sqrt{d}}{2} \right\rfloor \) by
\[
N\left( \frac{x + y\sqrt{d}}{2} \right) = \frac{x^2 - dy^2}{4}.
\]

Find all square free integers \( d \equiv 1 \pmod{4} \) such that the norm map on \( \mathbb{Z}
\left\lfloor \frac{1 + \sqrt{d}}{2} \right\rfloor \) satisfies the axiom of a Euclidean function. [Answer. They are \(-11, -7, -3, 5, 13, 17, 21, 29, 37, 41, 57, 73\). Note that \( \mathbb{Z}
\left\lfloor \frac{1 + \sqrt{d}}{2} \right\rfloor \) is a Euclidean domain but not for norm.]

2.5 Polynomial Rings

One of the familiar topics in elementary algebra is “polynomials”. Algebraic equations usually involve factorization of polynomials. Here, we treat them in a more abstract way with the things that we have studied from the previous section. There will be many important results in this section.

2.5.1 Polynomials and Their Roots

**2.5.1. Definition.** Let \( R \) be a ring with identity 1 and let \( x \) be a symbol called an indeterminate, not representing any element in \( R \). Let \( R[x] \) denote the set of all polynomials
\[a_0 + a_1 x + \cdots + a_n x^n \text{ where } n \in \mathbb{N} \cup \{0\}, a_i \in R, x^0 = 1, x^1 = x.\]

For \( i \in \mathbb{N} \), let \( x^i \) denote \( 1 \cdot x^i \). In the symbol \( a_0 + a_1 x + \cdots + a_n x^n \), we may drop \( a_i x^i \) if \( a_i = 0 \). Each element \( a_0 + a_1 x + \cdots + a_n x^n \) is called a polynomial and \( a_i \) is called the coefficient of \( x^i \) for \( i \in \{1, \ldots, n\} \) and \( a_0 \) is called the constant term.

For \( p(x) = a_0 + a_1 x + \cdots + a_n x^n \) and \( q(x) = b_0 + b_1 x + \cdots + b_m x^m \) in \( R[x] \), we can write \( p(x) q(x) = c_0 + c_1 x + \cdots + c_{n+m} x^{n+m} \) where \( k \geq \max\{m, n\}, a_i = 0 \) if \( i > n \) and \( b_j = 0 \) if \( j > m \) and we define

1. \( p(x) = q(x) \iff a_i = b_i \) for all \( i \in \{0, 1, \ldots, k\} \)
2. \( p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_k + b_k)x^k \) and
3. \( p(x)q(x) = c_0 + c_1 x + \cdots + c_{n+m} x^{n+m} \), where \( c_l = \sum_{i=0}^{l} a_ib_{l-i} \) for all \( l \in \{0, 1, \ldots, n \} \).

Hence, under the operation defined above \( R[x] \) is a ring which has 1 as its identity and contains \( R \) as a subring (considered elements as constant polynomials).
2.5.2. **Definition.** The ring $R[x]$ is called the ring of polynomials over $R$. If $R$ is commutative, so is $R[x]$. Set $R[x_1, x_2] = R[x_1][x_2]$ and $R[x_1, \ldots, x_n] = R[x_1, \ldots, x_{n-1}][x_n]$ if $n > 2$.

2.5.3. **Definition.** If $p(x) = a_0 + a_1 x + \cdots + a_n x^n \in R[x]$ and $a_n \neq 0$, then the degree of $p(x)$, denoted by $\deg p(x)$, is defined to be $n$. For technical reasons, we define the degree of the zero polynomial to be $-\infty$ and adopt the following conventions: $(-\infty) < n$ and $(-\infty) + n = -\infty = n + (-\infty)$ for every integer $n$; $(-\infty) + (-\infty) = -\infty$.

2.5.4. **Definition.** Let $f(x) = a_0 + a_1 x + \cdots + a_n x^n$. If $a_n \neq 0$, $a_n$ is called the leading coefficient of $f(x)$ and $f(x)$ is a monic polynomial if $a_n = 1$. If $R$ is commutative and $c \in R$, then

$$f(x) \mapsto f(c) := a_0 + a_1 c + \cdots + a_n c^n$$

gives a homomorphism from $R[x]$ to $R$, called the evaluation at $c$. In addition, if $f(c) = 0$, then $c$ is called a root of $f(x)$.

The following statements are clearly true.
1. Every unit in $R$ is a unit in $R[x]$.
2. If $f(x) = a_0 + a_1 x + \cdots + a_m x^m$, $g(x) = b_0 + b_1 x + \cdots + b_n x^n \in R[x]$ and $a_m b_n \neq 0$, then $\deg(f(x) g(x)) = m + n$.

In particular, if $R$ is an integral domain, we have:

2.5.5. **Theorem.** Let $R$ be an integral domain.
1. $R[x]$ is an integral domain.
2. $\forall f(x), g(x) \in R[x] \setminus \{0\}$, $\deg f(x) g(x) = \deg f(x) + \deg g(x)$.
3. The set of all units of $R[x]$ is the set of all units of $R$. In particular, $\mathcal{U}({\mathbb{Z}[x]}) = \{\pm 1\}$ and $\mathcal{U}({\mathbb{F}[x]}) = \mathbb{F} \setminus \{0\}$, where $\mathbb{F}$ is a field.
4. $\forall a \in R$, $a$ is irreducible in $R \iff a$ is irreducible in $R[x]$.
5. $\forall a, b \in R$, $a$ is a unit $\Rightarrow a + bx$ is irreducible in $R[x]$.

**Proof.** (1) and (2) are clear from the above observation. Note that if $f(x)$ is a unit, then $f(x)g(x) = 1$ for some $g(x) \in R[x]$, so $\deg f(x) + \deg g(x) = \deg 1 = 0$ by (2). This forces that $\deg f(x) = 0 = \deg g(x)$ which implies that $f(x)$ lies in $R$ and (3) follows. Next, let $a \in R$. If $a$ is irreducible in $R[x]$, then $a$ is clearly irreducible in $R$. On the other hand, if $a = f(x)g(x)$ for some nonzero nonunits $f(x)$ and $g(x)$ in $R[x]$, we have $0 = \deg a = \deg f(x) + \deg g(x)$, so this again gives $\deg f(x) = \deg g(x) = 0$. This means that $f(x)$ and $g(x)$ indeed lie in $R$, and thus $a$ is reducible in $R$. Finally, let $a, b \in R$ with $b$ a unit. Then $\deg(a + bx) = 1$. If $a + bx = f(x)g(x)$ for some $f(x), g(x) \in R[x]$, then $1 = \deg f(x) + \deg g(x)$, so $f(x)$ or $g(x)$ must lie in $R$, say $f(x) = c$ a constant in $R$ and $g(x) = u + vx$. Since $b = cv$ is a unit, $c$ is a unit. Hence, $a + bx$ is irreducible.

2.5.6. **Theorem.** [Division Algorithm] Let $R$ be a ring, $f(x), g(x) \in R[x]$ and $g(x) \neq 0$. Assume that the leading coefficient of $g(x)$ is a unit in $R$. Then $\exists$ unique $q(x), r(x) \in R[x]$ such that $f(x) = q(x)g(x) + r(x)$ where $r(x) = 0$ or $\deg r(x) < \deg g(x)$.

**Proof.** If there exists an $h(x) \in R[x]$ such that $f(x) = h(x)g(x)$, let $q(x) = h(x)$ and $r(x) = 0$. Assume that $f(x) \neq h(x)g(x)$ for all $h(x) \in R[x]$. Let

$$S = \{\deg(f(x) - h(x)g(x)) : h(x) \in R[x]\} \subset \mathbb{N} \cup \{0\}.$$

Then $S \neq \emptyset$. By the Well-Ordering Principle, there exists a polynomial $q(x)$ in $R[x]$ such that $\deg(f(x) - q(x)g(x))$ has the least degree and we may write $r(x)$ for $f(x) - q(x)g(x)$. Then
Assume that \( \deg r(x) \geq \deg g(x) \). Write \( r(x) = a_0 + a_1x + \cdots + a_nx^n \), \( a_n \neq 0 \), and \( g(x) = b_0 + b_1x + \cdots + b_mx^m \) with \( b_n \) a unit. Since \( \deg r(x) \geq \deg g(x) \), \( n - m \geq 0 \). Let \( s(x) = r(x) - a_nb_m^{-1}x^{n-m}g(x) \). Thus, \( \deg s(x) < n \) and

\[
s(x) = f(x) - q(x)g(x) = f(x) - (q(x) - a_nb_m^{-1}x^{n-m}g(x)),
\]

so \( s(x) \in S \) and \( \deg s(x) < \deg r(x) \), a contradiction.

To prove that \( q(x) \) and \( r(x) \) are unique, suppose that \( q_2(x) \) and \( r_2(x) \) are polynomials such that

\[
f(x) = g(x)q_2(x) + r_2(x) \quad \text{where } r_2(x) = 0 \text{ or } \deg r_2(x) < \deg g(x).
\]

Then \( g(x)q_2(x) + r(x) = f(x) = g(x)q_2(x) + r_2(x) \). Subtracting yields

\[
g(x)[q(x) - q_2(x)] = r_2(x) - r(x).
\]

Since the leading coefficient of \( g(x) \) is assumed to be a unit, we have

\[
\deg(g(x)[q(x) - q_2(x)]) = \deg g(x) + \deg(q(x) - q_2(x)).
\]

Since \( \deg(r_2(x) - r(x)) < \deg g(x) \), this relation can hold only if \( q(x) - q_2(x) \) is zero, i.e., \( q(x) = q_2(x) \), and hence finally \( r(x) = r_2(x) \).

For a field \( R \), the leading coefficient of a nonzero polynomial \( g(x) \) in \( R[x] \) is always a unit in \( R \), so the division algorithm above gives:

**2.5.7. Corollary.** If \( F \) is a field, then \( F[x] \) is a Euclidean domain with valuation \( d(p(x)) = \deg p(x) \) for all \( p(x) \in F[x] \setminus \{0\} \). Moreover, \( F[x] \) is a PID and a UFD.

**2.5.8. Theorem.** [Remainder Theorem] Let \( R \) be a ring and \( f(x) \in R[x] \). Then for all \( c \in R \), the remainder when \( x - c \) divides \( f(x) \) is \( f(c) \).

**Proof.** Let \( c \in R \). By Theorem 2.5.6, there exist unique \( q(x) \in R[x] \) and \( r \in R \) such that \( f(x) = q(x)(x - c) + r \). Then \( f(c) = q(c)(c - c) + r = r \).

**2.5.9. Corollary.** Let \( R \) be a ring.

1. If \( f(x) \in R[x] \), \( c \in R \) and \( f(c) = 0 \), then \( f(x) = q(x)(x - c) \) for some \( q(x) \in R[x] \).
2. If \( R \) is commutative, \( f(x) \in R[x] \) and \( c \in R \), then \( f(c) = 0 \iff (x - c) \mid f(x) \text{ in } R[x] \).
3. Let \( R \) be an integral domain, \( f(x) \in R[x] \), \( \deg f(x) = 2 \text{ or } 3 \) and the leading coefficient of \( f(x) \) is a unit in \( R \). Then \( f(x) \) has a root in \( R \iff f(x) \) is reducible in \( R[x] \).

**Proof.** (1) and (2) are clear. For (3), assume that \( c \) is a root of \( f(x) \). Then \( f(x) = q(x)(x - c) \) for some \( q(x) \in R[x] \). Since \( \deg f(x) \) is 2 or 3, \( \deg q(x) \) is 1 or 2, so \( f(x) \) is reducible. Conversely, suppose that \( f(x) = g(x)h(x) \), where \( g(x), h(x) \in R[x] \) of degree \( \geq 1 \). Since \( \deg f(x) = 2 \text{ or } 3 \), \( \deg g(x) = 1 \text{ or } \deg h(x) = 1 \). Hence, \( f(x) \) has a root in \( R \).

**2.5.10. Examples.**

1. \( x^2 - 3 \) is irreducible over \( \mathbb{Q} \) but not over \( \mathbb{R} \).
2. \( x^2 + 1 \) is irreducible over \( \mathbb{R} \) but not over \( \mathbb{C} \) since \( x^2 + 1 = (x - i)(x + i) \).
3. \( x^3 - x + 1 \) is irreducible over \( \mathbb{Z}_3 \) but reducible over \( \mathbb{R} \) by the intermediate value theorem. In general, every polynomial of odd degree over \( \mathbb{R} \) has a root in \( \mathbb{R} \).
4. \( x^4 + 4 \) has no roots in \( \mathbb{R} \) but it can be factored as \((x^2 - 2x^2 + 2)(x^2 + 2x^2 + 2)\) in \( \mathbb{R}[x] \).
2.5.11. Corollary. Let $F$ be a field.

1. If $f(x)$ is a polynomial over $F$ of degree $n$, then $f(x)$ has at most $n$ roots in $F$.
2. If $f(x)$ and $g(x)$ are polynomials over $F$ of degree $\leq n$ such that $f(\alpha_1) = g(\alpha_1), \ldots, f(\alpha_{n+1}) = g(\alpha_{n+1})$ where $\alpha_1, \ldots, \alpha_{n+1}$ are distinct elements of $F$, then $f(x) = g(x)$.
3. If $F$ is infinite and $f(x)$ and $g(x)$ are polynomials over $F$ such that $f(\alpha) = g(\alpha)$ for all $\alpha \in F$, then $f(x) = g(x)$.

Proof. We shall prove (1) by induction on $k = \deg f(x)$. It is clear when $f(x)$ is linear. Assume that $k > 1$ and any polynomials of degree $k$ have at most $k$ roots in $F$. Suppose that $f(x)$ is of degree $k + 1$. The statement is true when $f(x)$ has no root in $F$. Otherwise, let $\alpha$ be a root of $f(x)$ in $F$. Then $f(x) = (x - \alpha)q(x)$ for some polynomial $q(x) \in F[x]$ of degree $k$. By the inductive hypothesis, $q(x)$ has at most $k$ roots. Hence, $f(x)$ has at most $k + 1$ roots. The remaining statements follow from the first one.

2.5.12. Remarks.

1. $f(x) = x^2 - 1$ has four roots in $\mathbb{Z}_{12}$, namely $1, -1, 5, -5$.
2. Corollary 2.5.11 says that two polynomials over an infinite field $F$ which defined the same function on $F$ are identical. This is NOT true if $F$ is finite. Let $F = \mathbb{Z}_p$, $f(x) = x$ and $g(x) = x^p$. Then $f(\alpha) = g(\alpha)$ for all $\alpha \in \mathbb{Z}_p$ but $f(x) \neq g(x)$.

2.5.13. Theorem. Let $F$ be a field and $F[x]$ the polynomial ring over $F$. Then linear polynomials are the only atoms in $F[x]$ if and only if each polynomial $f(x) \in F[x]$ of positive degree has a root in $F$.

Proof. Suppose that linear polynomials are the only atoms in $F[x]$. Let $f(x)$ be a polynomial of positive degree over $F$. Since $F[x]$ is a UFD, $f(x) = \alpha_1(x) \cdots \alpha_k(x)$, a product of atoms. Each $\alpha_i(x)$ is linear, so $\alpha_i(x) = b_i(x - c_i)$ ($b_i, c_i \in F$ with $b_i \neq 0$). Then $(x - c_i) \mid f(x)$, so $c_1, \ldots, c_k$ are roots of $f(x)$ in $F$. Conversely, assume that every $f(x) \in F[x]$ of positive degree has a root in $F$. Let $\alpha(x)$ be an atom in $F[x]$. We claim that $\alpha(x)$ is linear. For, let $b \in F$ be a root of $\alpha(x)$. Then $(x - b) \mid \alpha(x)$ so $\alpha(x) = (x - b)\beta(x)$ for some $\beta(x) \in F[x]$. Since $\alpha(x)$ is an atom, $\beta(x)$ must be a unit. That is, $\beta(x)$ is a constant lying in $F \setminus \{0\}$. Thus, $\alpha(x)$ is a linear polynomial.

2.5.14. Definition. A field $F$ is an algebraically closed field if every non-constant polynomial has a root in $F$.

2.5.15. Example. By the fundamental theorem of algebra, the only atoms in $\mathbb{C}[x]$ are linear polynomials. Thus, $\mathbb{C}$ is an algebraically closed field.

2.5.16. Theorem. Let $R$ be an integral domain and $f(x) \in R[x]$ a nonzero polynomial. If $\alpha_1, \ldots, \alpha_k$ are distinct roots of $f(x)$, then $(x - \alpha_1) \cdots (x - \alpha_k)$ divides $f(x)$.

Proof. We shall prove this result by induction of $k$. Corollary 2.5.9 (1) gives the basis step. Assume $k > 1$. By the inductive hypothesis $(x - \alpha_1) \cdots (x - \alpha_{k-1})$ divides $f(x)$, so let $f(x) = (x - \alpha_1) \cdots (x - \alpha_{k-1})g(x)$ for some $g(x) \in R[x]$. Then

$$0 = f(\alpha_k) = (\alpha_k - \alpha_1) \cdots (\alpha_k - \alpha_{k-1})g(\alpha_k).$$

Thus, $g(\alpha_k) = 0$ since $R$ is an integral domain, so $(x - \alpha_k) \mid g(x)$. It follows that $(x - \alpha_1) \cdots (x - \alpha_k)$ divides $f(x)$.
2.5.2 Factorizations in Polynomial Rings

When factor a polynomial, we first look for some common factors on its coefficients. For example, \(2x^3 + 4 = 2(x^3 + 2)\). Taking the gcd of the coefficients out allows us to concentrate on polynomials with no common factor on their coefficients and leads to the following definitions.

2.5.17. Definition. Let \(R\) be a UFD and suppose that \(f(x) = a_0 + a_1x + \cdots + a_nx^n\) is a nonzero polynomial in \(R[x]\). The content of \(f(x)\) is the gcd of \(a_0,\ldots,a_n\). We say that \(f(x)\) is primitive if the content of \(f(x)\) is unit in \(R\), i.e., \(a_0,\ldots,a_n\) have no common factor except units.

2.5.18. Theorem. [Gauss’ lemma] Let \(R\) be a UFD and \(f(x), g(x) \in R[x]\). If \(f(x)\) and \(g(x)\) are primitive, so is \(f(x)g(x)\).

Proof. Let

\[
\begin{align*}
    f(x) &= a_0 + a_1x + \cdots + a_m x^m \\
    g(x) &= b_0 + b_1x + \cdots + b_n x^n \\
    f(x)g(x) &= c_0 + c_1x + \cdots + c_{m+n} x^{m+n}.
\end{align*}
\]

We shall suppose that \(f(x)g(x)\) is not primitive and obtain a contradiction. Let \(a \in R\) be an atom of \(R\) which divides all of \(c_0,\ldots,c_{m+n}\). Since \(R\) is a UFD, every atom is a prime, so \(Ra\) is a prime ideal. Then \((R/Ra)[x]\) is an integral domain. Since \(R[x]/Ra = (R/Ra)[x]\), \(R[x]/Ra\) is a prime ideal. Let

\[
\gamma : R[x] \to R[x]/Ra
\]

be the canonical map. Since \(a\) divides \(c_0,\ldots,c_{m+n}\), \(f(x)\bar{g}(x) = 0\). But \(a\) does not divide all of \(a_0,\ldots,a_m\) or all of \(b_0,\ldots,b_n\), since \(f\) and \(g\) are primitive. Thus, \(f(x) \neq 0\), \(\bar{g}(x) \neq 0\). This is a contradiction since \(f(x)\bar{g}(x) = 0\) and \(f(x), \bar{g}(x)\) lie in \(R[x]/P\) which is an integral domain. Hence, \(f(x)g(x)\) is primitive, as claimed.

2.5.19. Theorem. Let \(R\) be a UFD and \(f(x), g(x)\) nonzero polynomials of \(R[x]\). Then:

1. \(f(x)\) is primitive \(\iff\) the content of \(f(x)\) is 1.
2. If \(a\) is the content of \(f\), then \(f(x) = af_1(x)\) where \(f_1(x)\) is primitive.
3. If \(f(x) = af_1(x)\) and \(f_1(x)\) is primitive, then \(a\) is the content of \(f(x)\).
4. If \(a\) and \(b\) are the contents of \(f(x)\) and \(g(x)\), respectively, then \(ab\) is the content of \(f(x)g(x)\).

Proof. (1), (2) and (3) are immediate from the definition of gcd. For the last statement, by (2), we write \(f(x) = af_1(x)\) and \(g(x) = bg_1(x)\) where \(f_1(x)\) and \(g_1(x)\) are primitive. By Gauss’ lemma, \(f_1(x)g_1(x)\) is primitive, and

\[
f(x)g(x) = af_1(x)bg_1(x) = (ab)(f_1(x)g_1(x)).
\]

Hence, \(ab\) is the content of \(f(x)g(x)\), by (3).

2.5.20. Theorem. Let \(R\) be a UFD and let \(F = Q(R) = \{r/s : r, s \in R, s \neq 0\}\) be its field of quotients. Suppose \(f(x)\) is an irreducible polynomial in \(R[x]\). Then \(f(x)\), considered as a polynomial in \(F[x]\), is irreducible in \(F[x]\). In particular, if \(f(x) \in \mathbb{Z}[x]\) is irreducible over \(\mathbb{Z}\), it is irreducible over \(\mathbb{Q}\).

Proof. Suppose \(f(x) = g(x)h(x)\) where \(g(x)\) and \(h(x)\) are polynomials of positive degree in \(F[x]\). Let \(g(x) = a_0/b_0 + (a_1/b_1)x + \cdots + (a_m/b_m)x^m\) and \(h(x) = c_0/d_0 + (c_1/d_1)x + \cdots + (c_n/d_n)x^n\). Let \(b\) be a least common multiple of the \(b_i\) and \(d\) a least common multiple of the \(d_j\) so that

\[
g_1(x) = bg(x) \quad \text{and} \quad h_1(x) = dh(x)
\]
lie in \( R[x] \). Then
\[
bd'(x) = bg(x)dh(x) = g_1(x)h_1(x).
\]
By Theorem 2.5.19, let \( g_1(x) = u g_2(x) \) and \( h_1(x) = v h_2(x) \) where \( u \) is the content of \( g_1(x) \) and \( v \) is the content of \( h_1(x) \), and \( g_2(x) \) and \( h_2(x) \) are primitive polynomials in \( R[x] \). Thus,
\[
bd'(x) = g_1(x)h_1(x) = uv g_2(x)h_2(x).
\]
Since \( g_2(x) \) and \( h_2(x) \) are primitive, so is \( g_2(x)h_2(x) \) and hence the equation above implies that \( bd | uv \) in \( R \). Canceling, we obtain
\[
f(x) = wg_2(x)h_2(x) \quad \text{where } w = \frac{uv}{bd} \in R.
\]
Therefore, \( f(x) \) is reducible in \( R[x] \), which proves the theorem. \( \square \)

Let \( R \) be a UFD and \( F = \mathbb{Q}(R) \) its field of quotients. Suppose
\[
h(x) = a_0/b_0 + (a_1/b_1)x + \cdots + (a_n/b_n)x^n \in F[x],
\]
where \( a_0/b_0, a_1/b_1, \ldots, a_n/b_n \) are in “lowest terms”. That is, \( a_i \) and \( b_i \) have no common factor. Let
\[
b = \text{lcm}(b_0, \ldots, b_n).
\]
Then
\[
bh(x) = a_0(b/b_0) + a_1(b/b_1)x + \cdots + a_n(b/b_n)x^n
\]
is in \( R[x] \). Let \( a \) be the content of \( bh(x) \). It happens that \( a = \text{gcd}(a_0, \ldots, a_n) \), although knowing this is not essential. The main point is that
\[
h_1(x) = \left(\frac{b}{a}\right) h(x)
\]
is a primitive polynomial in \( R[x] \). Moreover, the proof of Theorem 2.5.20 shows that if \( f(x) \in R[x] \), then \( h(x) \mid f(x) \) in \( F[x] \Leftrightarrow h_1(x) \mid f(x) \) in \( R[x] \). In particular, suppose \( f(x) \in R[x] \), and \( r/s \in F \) is a root of \( f(x) \) where \( r \) and \( s \) are relatively prime. Then \( h(x) = x - (r/s) \) divides \( f(x) \) in \( F[x] \), so \( h_1(x) = sx - r \) divides \( f(x) \) in \( R[x] \). Thus, we have:

### 2.5.21. Theorem

Let \( R \) be a UFD and \( F \) its field of quotients. Suppose \( f(x) \in R[x] \) where
\[
f(x) = a_0 + a_1 x + \cdots + a_n x^n
\]
and \( r/s \in F \) is a root of \( f(x) \) where \( r \) and \( s \) are relatively prime. Then \( s \mid a_n \) and \( r \mid a_0 \) if \( r \neq 0 \).

**Proof.** The remarks above show that \( (sx - r) \mid (a_0 + a_1 x + \cdots + a_n x^n) \) in \( R[x] \). It is easy to see that this implies our results. \( \square \)

### 2.5.22. Remarks

1. Suppose \( f(x) = a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{Z}[x] \), where \( a_0 \neq 0 \). Then there are only finitely many rationals which can possibly be roots of \( f(x) \), namely the fractions \( r/s \) where \( r \mid a_0 \) and \( s \mid a_n \).
2. Note that if \( a_n = 1 \) above, then \( s = \pm 1 \) and \( r/s \in \mathbb{Z} \). In other words, if \( a_n = 1 \), then every rational root of \( f(x) \) is an integer.

Another important criterion on irreducibility in \( \mathbb{Q}[x] \) is the next theorem.

### 2.5.23. Theorem

[**Eisenstein’s Criterion**] Let \( f(x) = a_0 + a_1 x + \cdots + a_n x^n \) lies in \( \mathbb{Z}[x] \), and suppose that there is a prime number \( p \) such that

1. \( p \nmid a_n \),
2. \( p \mid a_0, \ldots, a_{n-1} \), and
3. \( p^2 \nmid a_0 \).

Then \( f(x) \) is irreducible in \( \mathbb{Q}[x] \). Moreover, if \( f(x) \) is primitive, then \( f(x) \) is irreducible in \( \mathbb{Z}[x] \).
Proof. We shall suppose that \( f(x) \) is reducible in \( \mathbb{Q}[x] \) and obtain a contradiction. By dividing \( f(x) \) by its content, we may assume that \( f(x) \) is primitive, this does not affect either the hypothesis or the reducibility of \( f(x) \) in \( \mathbb{Q}[x] \). By Theorem 2.5.20, \( f(x) \) is reducible in \( \mathbb{Z}[x] \), so let \( f(x) = g(x)h(x) \) where \( g(x) = b_0 + b_1x + \cdots + b_mx^m \) and \( h(x) = c_0 + c_1x + \cdots + c_{n-m}x^{n-m} \) are in \( \mathbb{Z}[x] \). Note that since \( f(x) \) is primitive, neither \( g(x) \) nor \( h(x) \) is constant. That is, \( m \geq 1 \) and \( n - m \geq 1 \). Let \( \bar{x} : \mathbb{Z}[x] \to \mathbb{Z}_p[x] \) be the canonical projection. Then \( \bar{f}(x) = \bar{a}_nx^n \) where \( \bar{a}_n \neq \bar{0} \) since \( p \nmid a_n \), so \( \bar{g}(x)\bar{h}(x) = \bar{f}(x) = \bar{a}_n x^n \). Since \( \mathbb{Z}_p[x] \) is a UFD, this forms \( \bar{g}(x) = \bar{b}_m x^m, \bar{h}(x) = \bar{c}_{n-m} x^{n-m} \), so that \( \bar{b}_0 = \bar{c}_0 = \bar{0} \) (i.e., \( p \) divides \( b_0 \) and \( c_0 \)). But then \( p^2 \mid a_0 \) since \( a_0 = b_0c_0 \), which contradicts part (3) of the hypotheses. Hence, \( f(x) \) is irreducible in \( \mathbb{Q}[x] \) as claimed. \( \square \)

2.5.24. Example. \( f(x) = 2x^5 - 6x^3 + 9x^2 - 15 \) is irreducible in \( \mathbb{Q}[x] \) and in \( \mathbb{Z}[x] \).

2.5.25. Corollary. The \( p \)th cyclotomic polynomial

\[
\Phi_p(x) = \frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \cdots + x + 1
\]

is irreducible in \( \mathbb{Q}[x] \) for any prime \( p \).

Proof. The polynomial

\[
g(x) = \Phi_p(x + 1) = \frac{(x + 1)^p - 1}{(x + 1) - 1} = x^{p-1} + \binom{p}{1} x^{p-2} + \cdots + p
\]

satisfies the Eisenstein criterion for the prime \( p \) and is thus irreducible in \( \mathbb{Q}[x] \). But clearly if \( \Phi_p(x) = h(x)r(x) \) were a nontrivial factorization of \( \Phi_p(x) \) in \( \mathbb{Z}[x] \), then

\[
\Phi_p(x + 1) = g(x) = h(x + 1)r(x + 1)
\]

would give a nontrivial factorization of \( g(x) \) in \( \mathbb{Z}[x] \). Thus, \( \Phi_p(x) \) must also be irreducible in \( \mathbb{Q}[x] \). \( \square \)

We next wish to prove a famous theorem of Gauss: if \( R \) is a UFD, so is \( R[x] \). Recall the criterion given in Theorem 2.4.11:

1. an integral domain is a UFD \( \iff \) every nonzero nonunit is a product of atoms and
2. every atom is prime.

Suppose \( R \) is a UFD. We first observe that \( R[x] \) is an integral domain, so this presents no problem. We shall establish the criteria above for \( R[x] \) (and these show that \( R[x] \) is a UFD) by doing three things:

(a) We determine all atoms of \( R[x] \) (Theorem 2.5.26).
(b) We show that they are primes (Theorem 2.5.27).
(c) We show that every nonzero nonunit of \( R[x] \) is a product of atoms and conclude that \( R[x] \) is a UFD (Theorem 2.5.28).

2.5.26. Theorem. Let \( R \) be a UFD, \( F \) its field of quotients and \( f(x) \in F[x] \). Then \( f(x) \) is an atom of \( R[x] \) \( \iff \) either

1. \( f(x) \in R \) and \( f(x) \) is an atom of \( R \) \quad or
2. \( f(x) \) is a primitive polynomial of degree \( n \geq 1 \) and \( f(x) \) is irreducible in \( F[x] \).
2.5. Polynomial Rings

Proof. Assume that \( f(x) \) is an atom of \( R[x] \). If \( \deg f(x) = 0 \), then \( f(x) \in R \), and clearly \( f(x) \) must be an atom of \( R \). Otherwise, suppose that \( \deg f(x) = n \geq 1 \), and let \( a \) be the content of \( f(x) \). Then \( f(x) = a f_1(x) \) where \( f_1(x) \) is primitive. Since \( f(x) \) is irreducible in \( R[x] \), \( a \) must be a unit in \( R \), so \( f(x) \) is primitive. Again, since \( f(x) \) is irreducible in \( R[x] \), it is also irreducible in \( F[x] \) by Theorem 2.5.20.

Conversely, assume that (1) and (2) hold. If \( f(x) \) is an atom of \( R \), it is clearly an atom of \( R[x] \) (Theorem 2.5.5). Suppose \( f(x) \) is a primitive polynomial of degree \( n \geq 1 \) and \( f(x) \) is irreducible in \( F[x] \). We claim that \( f(x) \) is an atom of \( R[x] \). For, suppose not, and let

\[
f(x) = g(x)h(x),
\]

where \( g(x) \) and \( h(x) \) are nonunits of \( R[x] \).

(a) If \( g(x) \) or \( h(x) \) lies in \( R \), then \( f(x) \) is not primitive, a contradiction.

(b) If \( g(x) \) and \( h(x) \) both have positive degree, then \( f(x) \) is reducible in \( F[x] \), again a contradiction.

Hence, if \( f(x) \) is an atom of \( R[x] \), it has the form (1) or (2), as required. \( \square \)

2.5.27. Theorem. Let \( R \) be a UFD and \( f(x) \) an atom of \( R[x] \). Then \( R[x]/f(x) \) is a prime ideal of \( R[x] \). That is, \( f(x) \) is a prime element.

Proof. We consider separately the two types of atoms in \( R[x] \) given in Theorem 2.5.26.

Case 1. \( a \in R \) is an atom of \( R \). Since \( R \) is a UFD, every atom is a prime, so \( Ra \) is a prime ideal. Then \( (R/Ra)[x] \) is an integral domain. Since \( R[x]/R[x]a \cong (R/Ra)[x] \), \( R[x]a \) is a prime ideal, so \( a \) is prime.

Case 2. \( f(x) \) is a primitive polynomial of degree \( n \geq 1 \) and \( f(x) \) is irreducible in \( F[x] \) where \( F \) is the quotient field of \( R \). First we claim that \( F[x]/f(x) \cap R[x] = R[x]/f(x) \). Clearly, \( f(x) \in F[x]/f(x) \cap R[x] \) with \( g(x) = a_0/b_0 + (a_1/b_1)x + \cdots + (a_n/b_n)x^n \in F[x] \). We can find relatively prime \( a, b \in R \) such that \( (b/a)g(x) = g_1(x) \) where \( g_1(x) \) is a primitive polynomial in \( R[x] \). (In fact, \( a = \gcd(a_0, a_1, \ldots, a_n) \) and \( b = \text{lcm}(b_0, b_1, \ldots, b_n) \) will do, provided each \( a_i \) and \( b_i \) are relatively prime.) Thus, \( (b/a)g(x) f(x) = g_1(x) f(x) \in R[x] \). By Gauss’ lemma, \( g_1(x) f(x) \) is a primitive polynomial. In connection with the above equation, this forces \( b \) to be a unit of \( R \), so \( g(x) = (a/b)g_1(x) \in R[x] \). Hence, \( g(x)f(x) \in R[x]f(x) \) which proves our claim.

By the second isomorphism theorem, we have

\[
R[x]/R[x]f(x) = R[x]/(R[x] \cap F[x]/f(x)) \cong (R[x] + F[x]/f(x))/F[x]/f(x).
\]

Since \( (R[x] + F[x]/f(x))/F[x]/f(x) \subseteq F[x]/f(x) \) which is a field because \( f(x) \) is irreducible in \( F[x] \), \( R[x]/R[x]f(x) \) is an integral domain. Thus, \( R[x]/R[x]f(x) \) is an integral domain, so \( R[x]f(x) \) is a prime ideal. Therefore, \( f(x) \) is prime and this proves the theorem. \( \square \)

2.5.28. Theorem. [Gauss] If \( R \) is a UFD, so is \( R[x] \). Hence, if \( R \) is a UFD, so is \( R[x_1, \ldots, x_n] \) for all \( n \in \mathbb{N} \).

Proof. We know all the atoms of \( R[x] \) by Theorem 2.5.26 and Theorem 2.5.27 tells us that each atom of \( R[x] \) is prime. Hence, (by Theorem 2.4.11) to verify that \( R[x] \) is a UFD, it remains to show that each nonzero nonunit \( f(x) \in R[x] \) is a product of atoms.

Case 1. \( \deg f(x) = 0 \), i.e., \( f(x) \in R \). Since \( R \) is a UFD and every atom of \( R \) is an atom of \( R[x] \), we can express \( f(x) \) as a product of atoms in \( R \), and so in \( R[x] \).

Case 2. \( \deg f(x) = n \geq 1 \). Let \( f(x) = f_1(x) \ldots f_k(x) \) where (a) each \( f_i(x) \) has degree \( \geq 1 \) and (b) \( k \) is as large as possible. Such a factorization exists because any factorization which satisfies (a) has at most \( n \) terms since \( n = \deg f(x) = \deg f_1(x) + \cdots + \deg f_k(x) \geq k \). Now, let \( a_i \) be the content of \( f_i(x), \) and let \( f_i(x) = a_i g_i(x) \) where \( g_i(x) \) is a primitive polynomial.
We claim that \( g_i(x) \) is an atom in \( R[x] \) because if \( g_i(x) = r(x)s(x) \) where \( r(x) \) and \( s(x) \) are nonunits, then \( r(x) \) and \( s(x) \) cannot lie in \( R \), since \( g_i(x) \) is primitive. In addition, \( r(x) \) and \( s(x) \) cannot both have positive degree because then we could write \( f(x) \) as a product of \( k+1 \) polynomials of positive degree, which violates (b). Thus, each \( g_i(x) \) is an atom as desired. Hence,

\[
f(x) = f_1(x) \ldots f_k(x) \\
= a_1g_1(x) \ldots a_kg_k(x) \\
= a_1 \ldots a_kg_1(x) \ldots g_k(x) \\
= ag_1(x) \ldots g_k(x), \text{ where } a \in R.
\]

By Case 1, \( a \) can be written as a product of atoms in \( R[x] \) and therefore shows that \( f(x) \) is a product of atoms in \( R[x] \), which proves \( R[x] \) is a UFD. \( \square \)

2.5.29. Example. Since \( \mathbb{Z} \) is a UFD, we have \( \mathbb{Z}[x] \) is also a UFD. However, in the following exercises, we shall know that the ideal \( (x, 2) \) of \( \mathbb{Z}[x] \) is not principal, so \( \mathbb{Z}[x] \) is an example of a UFD which is not a PID.

2.5. Exercises.

1. Let \( R \) be a ring.

(a) Show that \( M_n(R[x]) \cong M_n(R)[x] \) for all \( n \in \mathbb{N} \), where \( x \) is an indeterminate in both cases.

(b) If \( I \) is an ideal of \( R \), prove that \( I[x] \), the set of all polynomials with coefficients in \( I \), is an ideal of \( R[x] \) and \( R[x]/I[x] \cong (R/I)[x] \).

2. Prove the following statements.

(a) If \( R \) is an integral domain, then \( x \) is a prime element in \( R[x] \).

(b) In \( \mathbb{Z}[x] \), \( (x) \) is a prime ideal but not a maximal ideal.

(c) If \( F \) is a field, then \( (x) \) is a maximal ideal in \( F[x] \).

3. Let \( F \) be a field. Show that \( F[x]/(x^2) \cong \{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} : a, b \in F \} \), a subring of \( M_2(F) \).

4. Let \( F \) be a field. Find a ring isomorphism \( F[x]/(x^2 - x) \rightarrow F \times F \).

5. Prove that the ideal \( I = (x, 2) \) is not a principal ideal of \( \mathbb{Z}[x] \). Hence \( \mathbb{Z}[x] \) is not a PID. In addition, show that \( I \) is a maximal ideal in \( \mathbb{Z}[x] \).

6. Construct a field of:
   (a) 125 elements
   (b) 81 elements.

7. Find all odd prime numbers \( p \) such that \( x + 2 \) is a factor of \( x^4 + x^3 + x^2 - x + 1 \) in \( \mathbb{Z}[x] \).

8. Let \( p(x) \in \mathbb{R}[x] \). Prove that if \( p(a + bi) = 0 \), then \( p(a - bi) = 0 \) for all \( a, b \in \mathbb{R} \). Deduce by the fundamental theorem of algebra that there exist real numbers \( c, r_1, \ldots, r_k, a_1, b_1, \ldots, a_m, b_m \) such that

\[
p(x) = c(x - r_1) \ldots (x - r_k)(x^2 - (2a_1)x + (a_1^2 + b_1^2)) \ldots (x^2 - (2a_m)x + (a_m^2 + b_m^2)).
\]

In addition, if \( p(x) \in \mathbb{R}[x] \) is irreducible over \( \mathbb{R} \), then \( \deg p(x) = 1 \) or \( 2 \), namely, \( p(x) = bx + c \) or \( p(x) = ax^2 + bx + c \) with \( b^2 - 4ac < 0 \).

9. Let \( p \) be an odd prime. Prove that \( x^n - p \) is irreducible over \( \mathbb{Z}[x] \).

10. If \( R \) is an integral domain for which every ideal of \( R[x] \) is principal, show that \( R \) must be a field.

11. Let \( D \) be an integral domain. If \( \varphi : D[x] \rightarrow D[x] \) is an automorphism such that \( \varphi(a) = a \) for all \( a \in D \), prove that there exist \( c, d \in D \) with \( c \) a unit in \( D \) such that \( \varphi(x) = cx + d \). Here \( x \) stands for the indeterminate of \( D[x] \).

12. Let \( R \) be a UFD and \( F \) its field of quotients. Let \( f(x) = a_0 + a_1x + \cdots + a_nx^n \) lies in \( R[x] \), and suppose that there is an irreducible element \( p \in R \) such that

(i) \( p \nmid a_n \),

(ii) \( p \mid a_0, \ldots, a_{n-1} \), and

(iii) \( p^2 \nmid a_0 \).

Prove that \( f(x) \) is irreducible in \( F[x] \). Moreover, if \( f(x) \) is primitive, then \( f(x) \) is irreducible in \( R[x] \).

14. Project. (Units in a polynomial ring) Let \( R \) be a commutative ring and \( f(x) = a_0 + a_1x + \cdots + a_nx^n \) in \( R[x] \). Prove that \( f(x) \) is a unit in \( R[x] \) if and only if \( a_0 \) is a unit in \( R \) and \( a_1, \ldots, a_n \) are nilpotent elements in \( R \). (This project generalizes the result in Theorem 2.5.5 (3) to any commutative ring.)
15. Project. (Generalized Eisenstein’s criterion) Let \( P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \) be a polynomial with integer coefficients. If there exist a prime number \( p \) and an integer \( k \in \{0, 1, \ldots, n-1\} \) such that \( p \mid a_0, a_1, \ldots, a_k, p \nmid a_{k+1} \) and \( p^2 \nmid a_0 \), then \( P(x) \) has an irreducible factor in \( \mathbb{Z}[x] \) of degree greater than \( k \). Extend this result to a UFD similar to the last question of Exercises 2.5.

### 2.6 Field Extensions

Let \( F \) be a field and \( f(x) \) a polynomial over \( F \) of degree \( n \in \mathbb{N} \). Then the quotient ring

\[
F[x]/(f(x)) = \{ g(x) + (f(x)) : g(x) \in F[x] \} \\
= \{ g(x) + (f(x)) : g(x) \in F[x] \text{ and } g(x) = 0 \text{ or } \deg g(x) < n \} \\
= \{ a_0 + a_1 x + \cdots + a_{n-1} x^{n-1} + (f(x)) : a_i \in F \}
\]

by the division algorithm. Thus, if \( F \) is a finite field, then \( F[x]/(f(x)) \) is a commutative ring of \( |F|^n \) elements. In addition, if \( f(x) \) is irreducible in \( F[x] \), then \( (f(x)) \) is a maximal ideal, so \( F[x]/(f(x)) \) is a field. Note that \( F \) is isomorphic to \( \{ c + (f(x)) : c \in F \} \), so we may embed \( F \) into \( F[x]/(f(x)) \) by using the inclusion map.

**2.6.1 Examples.**

1. \( \mathbb{R}[x]/(x^2 + 1) \) is a field isomorphic to \( \mathbb{C} \) (with the map \( f(x) \mapsto f(i) \)).

2. \( \mathbb{Z}_{11}[x]/(x^2 + 3) \) is a field of 121 elements.

Under the above idea, if a polynomial over a field does not have a root in its own field, we shall create a bigger field where we can find a root of it.

### 2.6.1 Algebraic and Transcendental Extensions

Before we extend a field, we first determine the smallest one possible according to its characteristic.

**2.6.2 Definition.** Let \( F \) be a field. The intersection of all subfields of \( F \) is the smallest subfield of \( F \), called the **prime field** of \( F \).

**2.6.3 Theorem.** Let \( F \) be a field with the prime subfield \( P \) and \( 1_F \) denote the identity of \( F \).

1. If \( \text{char } F = p \), a prime, then \( P = \{ n \cdot 1_F : n = 0, 1, \ldots, p-1 \} \cong \mathbb{Z}/p\mathbb{Z} \).

2. If \( \text{char } F = 0 \), then \( P = \{(m \cdot 1_F)(n \cdot 1_F)^{-1} : m, n \in \mathbb{Z}, n \neq 0 \} \cong \mathbb{Q} \).

**Proof.** Since \( P \) is a field, \( 1_F \in P \), so \( \{ n \cdot 1_F : n \in \mathbb{Z} \} \subseteq P \). Define \( \varphi : \mathbb{Z} \to P \) by \( \varphi(n) = n \cdot 1_F \) for all \( n \in \mathbb{Z} \). Then \( \varphi \) is a ring homomorphism and \( \text{im } \varphi = \{ n \cdot 1_F : n \in \mathbb{Z} \} \), so \( \mathbb{Z}/\ker \varphi \cong \text{im } \varphi \).

(1) Assume that \( \text{char } F = p \) is a prime. Then \( \text{im } \varphi = \{ n \cdot 1_F : n = 0, 1, \ldots, p-1 \} \) and \( p \) is the smallest positive integer such that \( p \in \ker \varphi \), so \( \ker \varphi = p\mathbb{Z} \). Hence, \( \text{im } \varphi \cong \mathbb{Z}/p\mathbb{Z} \) which is a field, so \( P = \text{im } \varphi \cong \mathbb{Z}/p\mathbb{Z} \).

(2) Assume that \( \text{char } F = 0 \). Then \( \varphi \) is a monomorphism. Since \( \{ n \cdot 1_F : n \in \mathbb{Z} \} \subseteq P \) and \( P \) is a subfield of \( F \), \( \{(m \cdot 1_F)(n \cdot 1_F)^{-1} : m, n \in \mathbb{Z}, n \neq 0 \} \subseteq P \). Define \( \tilde{\varphi} : \mathbb{Q} \to P \) by \( \tilde{\varphi}(m/n) = \varphi(m/\varphi(n)^{-1}) \) for all \( m, n \in \mathbb{Z}, n \neq 0 \). Then \( \tilde{\varphi} \) is a monomorphism and \( \tilde{\varphi}|\mathbb{Z} = \varphi \). Thus, \( \mathbb{Q} \cong \text{im } \tilde{\varphi} = \{(m \cdot 1_F)(n \cdot 1_F)^{-1} : m, n \in \mathbb{Z}, n \neq 0 \} \) which is a subfield of \( P \), and hence they are equal.

In this section, we require some background in vector spaces.
2.6.4. Definition. A field $K$ is said to be an extension of a field $F$ if $F$ is a subring of $K$. The degree of $K$ over $F$, $[K : F]$, is the dimension of $K$ as a vector space over $F$. More generally, if a field $F$ is a subring of a ring $R$, then $[R : F]$ is the dimension of $R$ as a vector space over $F$.

2.6.5. Remark. By Theorem 2.6.3, any field can be considered as an extension field of the field $\mathbb{Q}$ or $\mathbb{Z}_p$ for some prime $p$.

For example, $[\mathbb{C} : \mathbb{R}] = 2$ and $[\mathbb{R} : \mathbb{Q}]$ is infinite (in fact $[\mathbb{R} : \mathbb{Q}] = [\mathbb{R}]$).

2.6.6. Theorem. If $[L : K]$ and $[K : F]$ are finite, then $[L : F]$ is finite and

$$[L : F] = [L : K][K : F].$$

In fact, $[L : F] = [L : K][K : F]$ whenever $F \subseteq K \subseteq L$.

Proof. With $F \subseteq K \subseteq L$, let $\{\beta_j\}_{j \in J}$ be a basis of $K$ over $F$ and $\{\alpha_i\}_{i \in I}$ a basis of $L$ over $K$. Every element of $L$ can be written uniquely as a finite linear combination of the elements of $\{\alpha_i\}_{i \in I}$ with coefficients in $K$, and every such coefficient can be written uniquely as a finite linear combination of the elements of $\{\beta_j\}_{j \in J}$ with coefficients in $F$. Hence, every element of $L$ can be written uniquely as a finite linear combination of the elements of $\{\alpha_i\beta_j\}_{i \in I, j \in J}$ with coefficients in $K$: $\{\alpha_i\beta_j\}_{i \in I, j \in J}$ is a basis of $L$ over $F$, and $[L : F] = |I \times J| = [L : K][K : F]$. $\square$

1. Notation. Let $K$ be an extension field of $F$.

1. If $t_1, \ldots, t_n$ are indeterminates over $F$, then $F(t_1, \ldots, t_n)$ denotes the field of quotients of the polynomial ring $F[t_1, \ldots, t_n]$.
2. If $u_1, \ldots, u_n \in K$ (or $S \subseteq K$), then $F[u_1, \ldots, u_n]$ (or $F[S]$) denotes the subring of $K$ generated by $F$ and $u_1, \ldots, u_n$ (or $S$), and $F(u_1, \ldots, u_n)$ (or $F(S)$) denotes its field of quotients.

2.6.7. Theorem. [Classification of Elements in an Extension Field] Let $K$ be a field extension of a field $F$ and let $u \in K$. Then EITHER

(a) $[F(u) : F] = \infty$ and $F[u] \cong F[t]$ where $t$ is an indeterminate OR
(b) $[F(u) : F]$ is finite and $F[u] = F(u)$.

Proof. Let $t$ be an indeterminate and consider the ring homomorphism

$$F[t] \xrightarrow{\varphi} K$$

defined by $\varphi(t) = u$ (or $\varphi(f(t)) = f(u)$). Note that the kernel of $\varphi$ is a prime ideal, since the image of $\varphi$ has no zero divisors. There are two possibilities.

1. $\ker \varphi = \{0\}$. Then we have (a).
2. $\ker \varphi \neq \{0\}$. Then $\varphi = F[t]g(t)$ where $g(t)$ is a monic prime (i.e., irreducible) polynomial. Since $F[t]$ is a PID, $F[t]g(t)$ is a maximal ideal. Thus,

$$F[u] \cong F[t]/F[t]g(t)$$

is a field, so $F[u] = F(u)$. $\square$

2.6.8. Remarks. 1. If $g(t) = g_0 + g_1 t + \cdots + g_{n-1} t^{n-1} + t^n$, then $[F(u) : F] = n$ and $\{1, u, \ldots, u^{n-1}\}$ is a basis for $F(u)$ over $F$.

2. Consider $\mathbb{R} \subset \mathbb{C}$ and $g(t) = g_0 + g_1 t + t^2 \in \mathbb{R}[t]$. We distinguish three cases.

(a) If $g_0^2 - 4 g_0 > 0$, then $g(t) = (t - a)(t - b)$ where $a, b \in \mathbb{R}$, $a \neq b$ and $\mathbb{R}[t]/\mathbb{R}[t]g(t)$ is a ring without nonzero nilpotent elements.
(b) If \( g_1^2 - 4g_0 = 0 \), then \( g(t) = (t - a)^2 \) and \( \mathbb{R}[t]/\mathbb{R}[t]g(t) \) is a ring with nonzero nilpotent elements.

(c) If \( g_1^2 - 4g_0 < 0 \), then \( \mathbb{R}[t]/\mathbb{R}[t]g(t) \cong \mathbb{C} \).

3. If \( p \) is a prime, then \( t^2 - p \) is irreducible over \( \mathbb{Q} \) and the fields \( \mathbb{Q}(\sqrt{p}) \cong \mathbb{Q}[t]/(t^2 - p) \) are all distinct.

**Proof.** Let \( p \) and \( q \) be distinct primes. Assume that \( \varphi : \mathbb{Q}(\sqrt{p}) \rightarrow \mathbb{Q}(\sqrt{q}) \) is an isomorphism. Then \( \varphi(1) = 1 \), and so \( \varphi(r) = r \) for all \( r \) in \( \mathbb{Q} \). Let \( \varphi(\sqrt{p}) = a + b\sqrt{q} \) for some \( a, b \in \mathbb{Q} \). Thus, \( p = \varphi(p) = (\varphi(\sqrt{p}))^2 = (a + b\sqrt{q})^2 = (a^2 + b^2q) + 2ab\sqrt{q} \). Since \( \sqrt{q} \) is not rational, \( ab = 0 \). However, if \( a = 0 \), then \( p = bq \) which implies \( q \mid p \). If \( b = 0 \), then \( \varphi(\sqrt{p}) = a = \varphi(a) \), so \( \sqrt{p} = a \) is rational. Hence, both cases lead to a contradiction. Therefore, \( \mathbb{Q}(\sqrt{p}) \) and \( \mathbb{Q}(\sqrt{q}) \) are not isomorphic.

An element in an extension field can be classified according to Theorem 5.1.6 as follows.

2.6.9. **Definition.** Let \( K \) be an extension field of a field \( F \). An element \( u \in K \) is **algebraic** over \( F \) in case there exists a nonzero polynomial \( f(t) \in F[t] \) such that \( f(u) = 0 \) and **transcendental element** over \( F \) otherwise.

For example, every complex number is algebraic over \( \mathbb{R} \); \( \sqrt{2} \) and \( 1 + \sqrt{5} \in \mathbb{R} \) are algebraic over \( \mathbb{Q} \). It has been proved that \( e \) and \( \pi \in \mathbb{R} \) are transcendental over \( \mathbb{Q} \) (by the Lindemann-Weierstrass theorem).

2.6.10. **Corollary.** Let \( K \) be an extension field of a field \( F \) and \( u \in K \). The following conditions on \( u \) are equivalent:

(i) \( u \) is transcendental over \( F \) (if \( f(t) \in F[t] \) and \( f(u) = 0 \), then \( f = 0 \));

(ii) \( F(u) \cong F(t) \);

(iii) \( [F(u) : F] \) is infinite.

2.6.11. **Corollary.** Let \( K \) be an extension field of a field \( F \) and \( u \in K \). The following conditions on \( u \) are equivalent:

(i) \( u \) is algebraic over \( F \) (there exists a polynomial \( 0 \neq f(t) \in F[t] \) such that \( f(u) = 0 \));

(ii) there exists a monic irreducible polynomial \( g(t) \in F[t] \) such that \( g(u) = 0 \);

(iii) \( [F(u) : F] \) is finite.

Moreover, in part (ii), we have \( g(t) \) is unique; \( f(u) = 0 \) if and only if \( g(t) \mid f(t) \); \( F(u) \cong F[t]/(g(t)) \); and \( [F(u) : F] = \deg g(t) \).

2.6.12. **Definition.** When \( u \) is algebraic over \( F \), the unique **monic irreducible** polynomial \( g(t) \in F[t] \) in part (ii) is the **minimal polynomial** of \( u \). The **degree of \( u \) over \( F \)** is \( \deg g(t) \).

2.6.13. **Definition.** An extension field \( K \) of a field \( F \) is **algebraic** in case every element of \( K \) is algebraic over \( F \).

For example, \( \mathbb{C} \) is an algebraic extension of \( \mathbb{R} \), but \( \mathbb{R} \) is not algebraic over \( \mathbb{Q} \). Note that if \( [K : F] \) is finite, then \( K \) is an algebraic extension because \( [F(u) : F] \leq [K : F] < \infty \) for all \( u \in K \).

2.6.14. **Definition.** An extension field \( E \) of a field \( F \) is said to be a **simple extension** of \( F \) if \( E = F(\alpha) \) for some \( \alpha \in E \).

2.6.15. **Example.** Prove that \( \mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3}) \) is a simple extension.
Solution. Let \( K = \mathbb{Q}(\sqrt{2} + \sqrt{3}) \). Since \( \sqrt{2} + \sqrt{3} \in \mathbb{Q}(\sqrt{2}, \sqrt{3}) \), we have \( K \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3}) \). For another inclusion, note that \( (\sqrt{2} + \sqrt{3})^2 = 5 + 2\sqrt{2}\sqrt{3} \), so \( \sqrt{2}\sqrt{3} \in K \). Thus,
\[
\sqrt{2} = (\sqrt{2} + \sqrt{3}\sqrt{2}\sqrt{3} - 2(\sqrt{2} + \sqrt{3})) \quad \text{and} \quad \sqrt{3} = 3(\sqrt{2} + \sqrt{3}) - (\sqrt{2} + \sqrt{3})\sqrt{2}\sqrt{3}
\]
are in \( K \). Hence, \( \mathbb{Q}(\sqrt{2}, \sqrt{3}) = K \).

\[\square\]

2.6.16. Theorem. If \( L \) is an algebraic extension of \( K \) and \( K \) is an algebraic extension of \( F \),
then \( L \) is algebraic extension over \( F \).

\[\square\]

Proof. Let \( u \in L \). Since \( L \) is algebraic over \( K \), there exists \( f(x) = a_0 + a_1x + \cdots + a_nx^n \in K[x] \) such that \( f(u) = 0 \). Since \( K \) is algebraic over \( F \), \( a_0, a_1, \ldots, a_n \) are algebraic over \( F \), so \( [F(a_0, a_1, \ldots, a_n) : F] \) is finite. For, let \( E = F(a_0, a_1, \ldots, a_n) \). Then
\[
[E : F] = [F(a_0) : F] \prod_{i=1}^{n} [F(a_0, a_1, \ldots, a_i : F(a_0, a_1, \ldots, a_{i-1})],
\]
a_0 \ is algebraic over \( F \) and \( a_i \) is algebraic over \( F(a_0, \ldots, a_{i-1}) \) for all \( i \in \{1, \ldots, n\} \). Since \( f(x) \in E[x] \), \( u \) is algebraic over \( E \), so \([E(u) : E] \) is finite by Corollary 5.1.10. Thus,
\[
[F(u) : F] \leq [E(u) : E] = [E(u) : E][E : F] < \infty.
\]
Hence, \( u \) is algebraic over \( F \).

\[\square\]

2.6.17. Corollary. For \( a, b \in K \), if \( a \) and \( b \) are algebraic over \( F \) of degree \( m \) and \( n \), respectively, then \( a \pm b, ab \) and \( a/b \) (if \( b \neq 0 \)) are all algebraic over \( F \) of degree \( \leq mn \). Hence, the set of all algebraic elements of \( K \) over \( F \) is a subfield of \( K \) and is an algebraic extension over \( F \).

Proof. By Corollary 5.1.10, \([F(a) : F] = m \) and \([F(b) : F] = n \). Since \( b \) is algebraic over \( F \), \( b \) is algebraic over \( F(a) \), so \([F(a)(b) : F(a)] \leq n \). Thus, by Theorem 5.1.5, \([F(a)(b) : F] = [F(a)(b) : F(a)][F(a) : F] \leq mn \). Since \( a \pm b, ab \) and \( a/b \) (if \( b \neq 0 \)) are in \( F(a, b) \) which is a finite extension, they are all algebraic over \( F \) of degree \( \leq mn \).

\[\square\]

2.6.18. Example. Consider \( Q \subseteq C \). Let \( A = \{z \in C : z \text{ is algebraic over } Q\} \). By Corollary 5.1.15, \( A \) is algebraic over \( Q \). Assume that \([A : Q] = n \) is finite. Let \( f(x) = x^n + 3 \). It is irreducible over \( Q \) by Eisenstein's criterion. Let \( \alpha \in C \) be such that \( f(\alpha) = 0 \). Then \( \alpha \in A \) ans so \( Q \subseteq Q[\alpha] \subseteq A \). But \([Q[\alpha] : Q] = n + 1 > [A : Q] \), which is a contradiction. Hence, \([A : Q] \) is infinite. This provides an example of infinite algebraic field extensions.

2.6.2 More on Roots of Polynomials

We conclude this chapter by working more on roots of polynomials. The theorem of Kronecker assures us that we may obtain an extension field of \( F \) in which the polynomial \( p(x) \in F[x] \) has a root.

2.6.19. Theorem. If \( F \) is a field and \( G \) is a finite subgroup of the multiplicative group of nonzero elements of \( F \), then \( G \) is a cyclic group. In particular, the multiplicative group of all nonzero elements of a finite field is cyclic.
2.6. Field Extensions

Proof. If \( G = \{1\} \), then \( G \) is cyclic. Assume that \( G \neq \{1\} \). Since \( G \) is a finite abelian group,
\[
G \cong \mathbb{Z}/(m_1) \oplus \cdots \oplus \mathbb{Z}/(m_k)
\]
where \( k \geq 1, m_1 > 1 \) and \( m_1 | \cdots | m_k \). Since \( m_k \sum_{i=1}^{k} \mathbb{Z}/(m_i) = 0 \), \( u \) is a root of the polynomial \( x^{m_k} - 1 \in F[x] \) for all \( u \in G \). By Corollary 2.5.11, this polynomial has at most \( m_k \) distinct roots in \( F \), so \( |G| \leq m_k \). Hence, we must have \( k = 1 \) and \( G \cong \mathbb{Z}/(m_1) \), which is a cyclic group.

\( \square \)

2.6.20. Remark. The finite multiplicative subgroup of a division ring may not be cyclic. E.g., \( Q_8 = \{\pm 1, \pm i, \pm j, \pm k\} \) is a subgroup of the ring of real quaternions \( \mathbb{H} \) and \( Q_8 \) is not cyclic.

2.6.21. Definition. Let \( R \) be an integral domain and \( f(x) \in R[x] \). If \( \alpha \) is a root of \( f(x) \), then there exist \( m \in \mathbb{N} \) and \( g(x) \in R[x] \) such that \( f(x) = (x - \alpha)^m g(x) \) and \( g(\alpha) \neq 0 \). \( m \) is called the multiplicity of the root \( \alpha \) of \( f(x) \) and if \( m > 1 \), \( \alpha \) is called a multiple root of \( f(x) \).

2.6.22. Definition. If \( f(x) = a_0 + a_1 x + \cdots + a_n x^n \in R[x] \), we define \( f'(x) \in R[x] \), the derivative of \( f(x) \), to be the polynomial \( f'(x) = a_1 + a_2 x + \cdots + n a_n x^{n-1} \).

We record the immediate properties of the derivative of polynomials in the next lemma.

2.6.23. Lemma. If \( f(x) \) and \( g(x) \) are polynomials over an integral domain \( R \) and \( c \in R \), then
1. \((cf(x))' = cf'(x) \),
2. \((f(x) + g(x))' = f'(x) + g'(x) \),
3. \((f(x)g(x))' = f(x)g'(x) + f'(x)g(x) \),
4. \(((f(x))^n)' = n(f(x))^{n-1}f'(x) \) where \( n \in \mathbb{N} \).

Characterizations of polynomials with multiple roots using derivatives are as follows.

2.6.24. Theorem. Let \( E \) be an extension of a field \( F \) and \( f(x) \in F[x] \).

1. For \( \alpha \in E \), \( \alpha \) is a multiple root of \( f(x) \) if and only if \( \alpha \) is a root of both \( f(x) \) and \( f'(x) \).
2. If \( f(x) \) and \( f'(x) \) are relatively prime, then \( f(x) \) has no multiple root.
3. If \( f(x) \) is irreducible over \( F \) having a root in \( E \), then \( f(x) \) has no multiple root in \( E \) if and only if \( f'(x) \neq 0 \).

Proof. (1) is clear.
(2) Since \( f(x) \) and \( f'(x) \) are relatively prime, there exist \( h(x) \) and \( k(x) \) in \( F[x] \) such that \( 1 = f(x)h(x) + f'(x)k(x) \). If \( \alpha \in E \) is a multiple root of \( f(x) \), by (1), \( f(\alpha) = 0 = f'(\alpha) \), so \( 1 = 0 \), a contradiction.
(3) Since \( f(x) \) is irreducible, \( f'(x) \neq 0 \) and \( \deg f'(x) < \deg f(x) \), we have \( f(x) \) and \( f'(x) \) are relatively prime, so \( f(x) \) has no multiple roots. Conversely, if \( f'(x) = 0 \), then \( f(\alpha) = 0 = f'(\alpha) \) for some \( \alpha \in E \) since \( f(x) \) has a root in \( E \). Hence, by (1), \( \alpha \) is a multiple root of \( f(x) \).

\( \square \)

2.6.25. Theorem. [Number of Roots] If \( f(x) \in F[x] \) and \( \deg f(x) = n \geq 1 \), then \( f(x) \) can have at most \( n \) roots counting multiplicities in any extension field of \( F \).

Proof. We shall prove the theorem by induction on the degree of \( f(x) \). If \( \deg f(x) = 1 \), then \( f(x) = ax + b \) for some \( a, b \in F \) and \( a \neq 0 \). Then \(-b/a\) is the unique root of \( f(x) \) and \(-b/a \in F \), so we are done.

Let \( \deg f(x) = n > 1 \) and assume that the result is true for all polynomials of degree \( < n \). Let \( E \) be an extension field of \( F \). If \( f(x) \) has no roots in \( E \), then we are done. Let \( r \in E \) be a root of \( f(x) \) of multiplicity \( m \geq 1 \). Then there exists \( q(x) \in E[x] \) such that \( f(x) = (x - r)^m q(x) \) and \( q(r) \neq 0 \). Then \( \deg q(x) = n - m \). By the inductive hypothesis \( q(x) \) has at most \( n - m \) roots in \( E \) counting multiplicities. Hence, \( f(x) \) has at most \( m + (n - m) \) roots in \( E \) counting multiplicities. \( \square \)
2.6.26. Theorem. [Kroncker] If \( p(t) \in F[t] \) is irreducible in \( F[t] \), then there exists an extension field \( E \) of \( F \) such that \( [E : F] = \deg p(t) \) and \( p(t) \) has a root in \( E \).

Proof. We use the discussion at the beginning of the section to prove this theorem. Let \( E = F[x]/(p(x)) \) where \( x \) is an indeterminate. Since \( p(x) \) is irreducible, \( E \) is a field containing \( \{a + (p(x)) : a \in F\} \) as a subfield. But \( F \cong \{a + (p(x)) : a \in F\} \) by \( \varphi : a \mapsto a + (p(x)) \), so \( E \) can be considered an extension field of \( F \) by considering \( a \) as \( a + (p(x)) \) for all \( a \in F \). Then \( E = F[x]/(p(x)) = F(\bar{t}) \) where \( \bar{t} = x + (p(x)) \) is a root of \( p(t) \). Since \( E = F(\bar{t}) \) and \( p(t) \) is irreducible over \( F \), \( [E : F] = [F(\bar{t}) : F] = \deg p(t) \) by Corollary 5.1.10.

2.6.27. Corollary. If \( p(t) \in F[t] \) is a nonconstant polynomial, then there exists a finite extension field \( E \) of \( F \) containing a root of \( p(t) \) and \( [E : F] \leq \deg p(t) \).

Proof. Since \( F[t] \) is a UFD, \( p(t) \) has an irreducible factor in \( F[t] \), say \( p_1(t) \). By Theorem 2.6.26, there exists an extension field \( E \) of \( F \) such that \( E \) contains a root of \( p_1(t) \) and \( [E : F] = \deg p_1(t) \). Hence, \( [E : F] \leq \deg p(t) \) and \( E \) contains a root of \( p(t) \).

2.6. Exercises.

1. If \( u \in K \) is algebraic of odd degree over \( F \), prove that \( F(u^2) = F(u) \).

2. Let \( a, b \in K \) be algebraic over \( F \) of degree \( m \) and \( n \), respectively. Prove that if \( m \) and \( n \) are relatively prime, then \( [F(a, b) : F] = mn \).

3. Show that the degree of \( \sqrt{2} + \sqrt{3} \) over \( \mathbb{Q} \) is 4 and the degree of \( \sqrt{2} + \sqrt{5} \) is 6.

4. Let \( p \) be a prime and let \( v \in \mathbb{C} \) satisfy \( v \neq 1, v^p = 1 \) (e.g., \( v = \cos(2\pi/p) + i\sin(2\pi/p) \)). Show that \( [Q(v) : Q] = p - 1 \).

5. Let \( E = \mathbb{Q}(u) \) where \( u^2 - u^2 + u + 2 = 0 \). Express \((u^2 + u + 1)(u^2 - u)\) and \((u - 1)^{-1}\) in the form \( au^2 + bu + c \) where \( a, b, c \in \mathbb{Q} \).

6. Let \( E \) be an algebraic extension of a field \( F \). Show that any subring of \( E/F \) is a subfield. Hence prove that any subring of a finite dimensional extension field \( E/F \) is a subfield.

7. Let \( E = F(u) \), \( u \) transcendental and let \( K \neq F \) be a subfield of \( E/F \). Show that \( u \) is algebraic over \( K \).

8. Let \( u \) and \( v \) be positive irrational numbers such that \( u \) is algebraic over \( \mathbb{Q} \) and \( v \) is transcendental over \( \mathbb{Q} \).

(a) Show that \( v \) is transcendental over \( \mathbb{Q}[u] \).

(b) Classify whether the following elements are algebraic or transcendental over \( \mathbb{Q} \).

(i) \( \frac{1}{u + v} \)

(ii) \( \sqrt{u} \)

(iii) \( \sqrt{v} \)

9. (a) Show that there are countably many irreducible polynomials in \( \mathbb{Q}[x] \).

(b) Let \( A \) be the set of all real numbers that are algebraic over \( \mathbb{Q} \). Show that \( A \) is countable, so that \( \mathbb{R} \setminus A \) is uncountable.

10. Let \( R \) be an integral domain and \( f(x) \) a nonconstant polynomial. Prove that:

(a) If char \( R = 0 \), then \( f'(x) \neq 0 \).

(b) If char \( R = p \), a prime, then \( f'(x) = 0 \) if and only if \( 3a_0, a_1, \ldots, a_n \in R, f(x) = a_0 + a_1 x^p + \cdots + a_n x^{np} \).

11. Suppose that \( F \) is a finite field and \( f(x) \in F[x] \) a nonconstant. If \( f'(x) = 0 \), prove that \( f(x) \) is reducible over \( F \).

12. Let \( F \) be a finite field with \( q \) elements. Prove that if \( K \) is an extension field of \( F \) and \( b \in K \) is algebraic over \( F \), then \( b^q = b \) for some \( m \in \mathbb{N} \).

13. A complex number \( \alpha \) is called an algebraic integer if it is a root of a monic polynomial

\[ f(x) = x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \]

whose coefficients are in \( \mathbb{Z} \). Let \( A = \{\alpha \in \mathbb{C} : \alpha \text{ is an algebraic integer}\} \). Prove that \( A \cap \mathbb{Q} = \mathbb{Z} \).

14. Let \( f(x) = x^2 + x + 2 \) be a polynomial in \( \mathbb{Z}_3[x] \) and \( E = \mathbb{Z}_3[x]/(f(x)) \).

(a) Show that \( f(x) \) is irreducible in \( \mathbb{Z}_3[x] \), and so \( E \) is a field of 9 elements extending \( \mathbb{Z}_3 \).

(b) Find the characteristic of \( E \) and \( [E : \mathbb{Z}_3] \).

(c) Find the multiplicative inverse of \( 1 + x + (f(x)) \).

(d) How many generators of the cyclic multiplicative group \( E \setminus \{0\} \)?
15. Let $E_1$ and $E_2$ be subfields of a field $K$. The **composite field** of $E_1$ and $E_2$, denoted by $E_1E_2$, is the smallest subfield of $K$ containing both $E_1$ and $E_2$. Prove that if $[K : F]$ is finite, then $[E_1E_2 : F] \leq [E_1 : F][E_2 : F]$.

16. Let $\alpha$ be an irrational number. If $\alpha$ is a common root of $f(x) = x^3 + ax + b$ and $g(x) = x^2 + cx + d$ for some $a, b, c, d \in \mathbb{Q}$, prove that:
   (a) $g(x)$ is a factor of $f(x)$
   (b) $c$ is a root of $f(x)$.

---

16. **Project.** (Galois ring) Let $p$ be a prime and $n, r \in \mathbb{N}$. Let $f(x)$ be a monic irreducible polynomial in $\mathbb{Z}_p[x]$ of degree $r$. Consider this polynomial as a polynomial in $\mathbb{Z}_{p^n}[x]$.

   (a) Prove that the quotient ring $R = \mathbb{Z}_{p^n}[x]/(f(x))$ is a commutative ring of $p^{nr}$ elements that contains the ring $\mathbb{Z}_{p^n} \cong \{ c + (f(x)) : c \in \mathbb{Z}_{p^n}\}$ as a subring.

   (b) Prove by the first isomorphism theorem that
   $$R/(p + (f(x))) \cong \mathbb{Z}_p[x]/(f(x)).$$

   Deduce that $M = (p + (f(x)))$ is a maximal ideal of $R$.

   (c) Prove that $R \setminus M$ is the unit group $R^\times$. Conclude that $M$ is the unique maximal ideal of $R$ and so $R$ is a local ring.

   The ring $\mathbb{Z}_{p^n}[x]/(f(x))$ is called a **Galois ring**. It is a ring extension of the ring $\mathbb{Z}_{p^n}$ similar to a Galois field that is a field extension of the field $\mathbb{Z}_p$. This finite ring has many parallel properties to the finite field and has many applications in algebraic graph theory and algebraic coding theory.