MATH2301/7307
Linear & Abstract Algebra & Number Theory

Abstract Algebra & Number Theory Notes

2022

Contents

Chapter 1

Number Theory 1: Divisibility and the Primes

1.1 Divisibility

Definition 1.1.1. The set of positive integers $\{1, 2, 3, 4, \ldots\}$ is denoted by $\mathbb{N}$, and the set of all integers $\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ is denoted by $\mathbb{Z}$.

Definition 1.1.2. Let $a, b \in \mathbb{Z}$. We say that a **divides** b and write $a \mid b$ if and only if there exists an integer c such that $b = ac$. If no such c exists then we write $a \nmid b$. If a divides b, then we call a a **divisor** of b and we call b a **multiple** of a.

Some basic properties of divisibility are as follows.

Theorem 1.1.3. Let $a, b, c, d \in \mathbb{N}$.

(a) $1 \mid a$ and $a \mid a$.

(b) If $a \mid b$, then $a \leq b$.

(c) If $a \mid b$ and $b \mid a$, then $a = b$.

(d) If $a \mid b$ and $b \mid c$, then $a \mid c$.

(e) If $a \mid c$ and $b \mid d$, then $ab \mid cd$.

(f) If $a \mid b$ and $a \mid c$, then a divides any integer linear combination of b and c. That is, $a \mid (\alpha b + \beta c)$ for any $\alpha, \beta \in \mathbb{Z}$.

Proof

(a) $a = a \cdot 1$ so $1 \mid a$ and $a \mid a$.

(b) If $a \mid b$, then $b = ac$ for some $c \in \mathbb{Z}$, and it follows from $a, b \in \mathbb{N}$ that $c \in \mathbb{N}$. Thus, $b - a = ac - a = a(c - 1) \geq 0$. So $b \geq a$.

(c) If $a \mid b$ and $b \mid a$, then by (b) we have $a \leq b$ and $b \leq a$, which implies $a = b$.

(d) If $a \mid b$ and $b \mid c$, then $b = ad$ for some $d \in \mathbb{Z}$ and $c = be$ for some $e \in \mathbb{Z}$. Thus, $c = be = (ad)e = a(de)$ and de is an integer, so $a \mid c$.

(e) If $a \mid c$ and $b \mid d$, then $c = ae$ and $d = bf$ for some $e, f \in \mathbb{Z}$. Thus, $cd = (ae)(bf) = (ef)(ab)$ and so $ab \mid cd$ (since ef is an integer).

(f) If $a \mid b$ and $a \mid c$, then $b = ad$ and $c = ae$ for some $d, e \in \mathbb{Z}$. Let $\alpha, \beta \in \mathbb{Z}$. Thus, $\alpha b + \beta c = \alpha ad + \beta ae = a(\alpha d + \beta e)$ and $\alpha d + \beta e$ is an integer, so $a \mid \alpha b + \beta c$.

□

1.2 Greatest Common Divisor

Definition 1.2.1. Let $a, b \in \mathbb{N}$. The **greatest common divisor**, or **gcd**, of a and b is the largest integer d such that $d \mid a$ and $d \mid b$. This is denoted $d = \gcd(a, b)$.

Note that since $1 \mid a$ and $1 \mid b$ there exists a positive common divisor of a and b for all $a, b \in \mathbb{N}$. Since any divisor of a is no larger than a, we have that $\gcd(a, b)$ exists and is positive for all $a, b \in \mathbb{N}$.

Definition 1.2.2. Let $a, b \in \mathbb{N}$. We say that a and b are **relatively prime** or **coprime** if $\gcd(a, b) = 1$.

1.3 Division Algorithm

Theorem 1.3.1. Let $a, b \in \mathbb{Z}$ with $b \neq 0$. There exist unique integers q and r such that $a = qb + r$ and $0 \leq r < |b|$.

Proof Consider the set S of integers given by $S = \{a - nb : n \in \mathbb{Z}\}$. Since $b \neq 0$, S contains infinitely many nonnegative integers, and so there is a least nonnegative integer in S. Let r be the least nonnegative integer in S and let q be such that $a - qb = r$. Thus, we have $a = qb + r$ where q and r are integers with $r \geq 0$. Now, $\frac{|b|}{b}$ is an integer and so $a - (q + \frac{|b|}{b})b \in S$. But $a - (q + \frac{|b|}{b})b = a - qb - |b| = r - |b|$, and so we have $r - |b| \in S$. Since $r - |b| < r$, by the definition of r we have $r - |b| < 0$. That is, $r < |b|$.

We now prove uniqueness. Suppose $a = q_1 b + r_1$ and $a = q_2 b + r_2$ where $0 \leq r_1 < |b|$ and $0 \leq r_2 < |b|$. For a contradiction, suppose $r_2 > r_1$. Now, $q_1 b + r_1 = q_2 b + r_2$ gives us $r_2 - r_1 = (q_1 - q_2)b$, and so we have $b \mid (r_2 - r_1)$. Thus, $|b| \mid (r_2 - r_1)$ and so Theorem 1.1.3 (b) tells us that $|b| \leq r_2 - r_1$. However, we also have $r_2 - r_1 \leq r_2 < |b|$. This is a contradiction and we conclude that $r_2 \leq r_1$. We obtain a similar contradiction if we suppose $r_1 > r_2$. So $r_1 = r_2$ and it follows that $q_1 = q_2$. Thus, we have proven uniqueness. □

1.4 Euclidean GCD Algorithm

Theorem 1.4.1. If $a = qb + r$, then $\gcd(a, b) = \gcd(b, r)$.

Proof Let $d = \gcd(a, b)$ and let $c = \gcd(b, r)$. Since $d \mid a$ and $d \mid b$, we have $d \mid (a - qb)$ by Theorem 1.1.3 (f). But $a - qb = r$. So d is a common divisor of b and r, and so $d \leq c$. We also have $c \mid qb + r$ by Theorem 1.1.3 (f), and so we have $c \mid a$. Thus, c is a common divisor of a and b and so $c \leq d$. Since we have $d \leq c$ and $c \leq d$, we have $c = d$. □

Let $a,\ b \in \mathbb{N}$. Since $\gcd(a, b) = \gcd(b, a)$, we may assume $a \geq b$. Use the division algorithm repeatedly to write

$$\begin{array}{rcccll}
a & = & q_0 b & + & r_1 & 0 \leq r_1 < b \\
b & = & q_1 r_1 & + & r_2 & 0 \leq r_2 < r_1 \\
r_1 & = & q_2 r_2 & + & r_3 & 0 \leq r_3 < r_2 \\
r_2 & = & q_3 r_3 & + & r_4 & 0 \leq r_4 < r_3 \\
 & \vdots & & \vdots & & \vdots \\
r_{n-2} & = & q_{n-1} r_{n-1} & + & \boxed{r_n} & 0 \leq r_n < r_{n-1} \\
r_{n-1} & = & q_n r_n & + & 0 &
\end{array}$$

Since $r_1 > r_2 > \cdots > r_n$, and since $r_1, \ldots, r_n \geq 0$, we eventually reach a step where the remainder is 0. Theorem 1.4.1 tells us that

$$\gcd(a, b) = \gcd(b, r_1) = \gcd(r_1, r_2) = \gcd(r_2, r_3) = \cdots = \gcd(r_{n-1}, r_n)$$

and the last line tells us that $\gcd(r_{n-1}, r_n) = r_n$. Thus, we have $\gcd(a, b) = r_n$ where r_n is the last nonzero remainder we obtain via the above procedure.

Example 1.4.2. Find $\gcd(42823, 6409)$.

$$\begin{array}{rcccl}
42823 & = & 6 \cdot 6409 & + & 4369 \\
6409 & = & 1 \cdot 4369 & + & 2040 \\
4369 & = & 2 \cdot 2040 & + & 289 \\
2040 & = & 7 \cdot 289 & + & \boxed{17} \\
289 & = & 17 \cdot 17 & + & 0
\end{array}$$

So $\gcd(42823, 6409) = 17$.

The above procedure can be written more compactly as follows. There is no need to write each remainder three times.

q	r
	42823
6	6409
1	4369
2	2040
7	289
17	17
	0

1.5 Extended Euclidean GCD Algorithm

The Euclidean GCD Algorithm can be used to find solutions to linear Diophantine equations in two unknowns, when they exist. A linear Diophantine equation in two unknowns is an equation of the form

$$ax + by = c$$

where $a, b, c \in \mathbb{N}$ are given, and integer solutions for x and y are sought.

The first step is to find $\gcd(a, b)$ via the Euclidean GCD Algorithm. The steps of the Euclidean GCD Algorithm take the following form.

$$\begin{aligned} a &= q_0 b + r_1 \\ b &= q_1 r_1 + r_2 \\ r_1 &= q_2 r_2 + r_3 \\ &\vdots \\ r_{n-3} &= q_{n-2} r_{n-2} + r_{n-1} \\ r_{n-2} &= q_{n-1} r_{n-1} + r_n \\ r_{n-1} &= q_n r_n + 0 \end{aligned}$$

The next step is to find an expression for $\gcd(a, b)$ as an integer linear combination a and b, and this is achieved as follows. This process can be thought of as working backwards through the Euclidean GCD Algorithm. From the equation $r_{n-2} = q_{n-1} r_{n-1} + r_n$ of the Euclidean Algorithm we can write r_n in terms of r_{n-1} and r_{n-2}.

$$\gcd(a, b) = r_n = -q_{n-1} r_{n-1} + r_{n-2}$$

Then from the equation $r_{n-3} = q_{n-2} r_{n-2} + r_{n-1}$ we can write r_{n-1} in terms of r_{n-2} and r_{n-3}

$$r_{n-1} = -q_{n-2} r_{n-2} + r_{n-3}$$

and substitute this into our expression for r_n. This gives us an expression for $\gcd(a, b) = r_n$ in terms of r_{n-2} and r_{n-3}.

$$\gcd(a, b) = -q_{n-1}(-q_{n-2} r_{n-2} + r_{n-3}) + r_{n-2} = (1 + q_{n-1} q_{n-2}) r_{n-2} - q_{n-1} r_{n-3}.$$

By continuing in this manner, we eventually obtain an expression for $\gcd(a, b)$ in terms of a and b.

Example 1.5.1. Express $\gcd(98, 36)$ as an integer linear combination of 98 and 36.

$$\begin{array}{rcrcr} 98 & = & 2 \cdot 36 & + & 26 \\ 36 & = & 1 \cdot 26 & + & 10 \\ 26 & = & 2 \cdot 10 & + & 6 \\ 10 & = & 1 \cdot 6 & + & 4 \\ 6 & = & 1 \cdot 4 & + & \boxed{2} \\ 4 & = & 2 \cdot 2 & + & 0 \end{array}$$

So $\gcd(98, 36) = 2$.

$$\begin{array}{rcrcr} 2 & = & -4 & + & 6 \\ & = & -(-6 + 10) & + & 6 \\ & = & 2 \cdot 6 & - & 10 \\ & = & 2 \cdot (-2 \cdot 10 + 26) & - & 10 \\ & = & -5 \cdot 10 & + & 2 \cdot 26 \\ & = & -5 \cdot (-26 + 36) & + & 2 \cdot 26 \\ & = & 7 \cdot 26 & - & 5 \cdot 36 \\ & = & 7 \cdot (-2 \cdot 36 + 98) & - & 5 \cdot 36 \\ & = & -19 \cdot 36 & + & 7 \cdot 98 \end{array}$$

So $\gcd(98, 36) = 2$ and

$$2 = 7 \cdot 98 - 19 \cdot 36.$$

An expression for $\gcd(a, b)$ as an integer linear combination of a and b, gives us a solution to $ax + by = c$ whenever c is a multiple of $\gcd(a, b)$. For if we have integers α and β such that $\gcd(a, b) = \alpha a + \beta b$ and $c = \gamma \cdot \gcd(a, b)$, then we have $c = (\gamma\alpha)a + (\gamma\beta)b$. So $x = \gamma\alpha$ and $y = \gamma\beta$ is a solution. If c is not a multiple of $\gcd(a, b)$, then the equation $ax + by = c$ has no integer solution in x and y, as the following theorem shows.

Theorem 1.5.2. Let $a, b, c \in \mathbb{N}$. The equation

$$ax + by = c$$

has a solution in integers x and y if and only if c is a multiple of $\gcd(a, b)$.

Proof Let $d = \gcd(a, b)$. If $ax + by = c$, then $d \mid a$ and $d \mid b$ so $d \mid (ax + by)$ by Theorem 1.1.3 (f). That is, $d \mid c$. Going in the other direction, if $d \mid c$, then $c = de$ for some integer e. Using the extended Euclidean GCD Algorithm we can find integers x' and y' such that $ax' + by' = d$. Multiplying through by e we obtain

$$aex' + bey' = de = c$$

so taking $x = ex'$ and $y = ey'$ is a solution. □

Example 1.5.3. Determine whether there exists a solution in integers x and y to the linear Diophantine equation

$$155x + 403y = 19$$

and give a solution if there exists one.

We begin by applying the Euclidean GCD Algorithm to find $\gcd(403, 155)$.

$$\begin{array}{rcl} 403 &=& 2 \cdot 155 + 93 \\ 155 &=& 1 \cdot 93 + 62 \\ 93 &=& 1 \cdot 62 + 31 \\ 62 &=& 2 \cdot 31 + 0 \end{array}$$

So $gcd(403, 155) = 31$. Since $31 \nmid 19$, there is no solution to this linear Diophantine equation.

Example 1.5.4. Determine whether there exists a solution in integers x and y to the linear Diophantine equation

$$1098x + 131y = 4$$

and give a solution if there exists one.

We begin by applying the Euclidean GCD Algorithm to find $\gcd(1098, 131)$.

$$\begin{array}{rcl} 1098 &=& 8 \cdot 131 + 50 \\ 131 &=& 2 \cdot 50 + 31 \\ 50 &=& 1 \cdot 31 + 19 \\ 31 &=& 1 \cdot 19 + 12 \\ 19 &=& 1 \cdot 12 + 7 \\ 12 &=& 1 \cdot 7 + 5 \\ 7 &=& 1 \cdot 5 + 2 \\ 5 &=& 2 \cdot 2 + 1 \\ 2 &=& 2 \cdot 1 + 0 \end{array}$$

So $\gcd(1098, 131) = 1$ and $1 \mid 4$ so this linear Diophantine equation does have a solution.

We now proceed backwards through the Euclidean GCD Algorithm to express $\gcd(1098, 131) = 1$

as an integer linear combination of 1098 and 131.

$$\begin{aligned}
1 &= -2 \cdot 2 + 5 \\
&= -2(-5+7) + 5 \\
&= 3 \cdot 5 - 2 \cdot 7 \\
&= 3 \cdot (-7+12) - 2 \cdot 7 \\
&= -5 \cdot 7 + 3 \cdot 12 \\
&= -5 \cdot (-12+19) + 3 \cdot 12 \\
&= 8 \cdot 12 - 5 \cdot 19 \\
&= 8 \cdot (-19+31) - 5 \cdot 19 \\
&= -13 \cdot 19 + 8 \cdot 31 \\
&= -13 \cdot (-31+50) + 8 \cdot 31 \\
&= 21 \cdot 31 - 13 \cdot 50 \\
&= 21 \cdot (-2 \cdot 50 + 131) - 13 \cdot 50 \\
&= -55 \cdot 50 + 21 \cdot 131 \\
&= -55 \cdot (-8 \cdot 131 + 1098) + 21 \cdot 131 \\
&= 461 \cdot 131 - 55 \cdot 1098
\end{aligned}$$

So we have $461 \cdot 131 - 55 \cdot 1098 = 1$. Multiplying both sides by 4 and rearranging we obtain

$$1098 \cdot (-220) + 131 \cdot (1844) = 4$$

and see that our solution is $x = -220$ and $y = 1844$.

1.6 Consequences of Euclid's Algorithm

Definition 1.6.1. A natural number $n \geq 2$ is prime if and only if the only positive divisors of n are 1 and n.

The first few primes numbers are $2, 3, 5, 7, 11, 13, 17$. The number 1 is not prime.

Theorem 1.6.2. Let $a, b, c \in \mathbb{N}$. If $\gcd(a, b) = 1$ and $a \mid bc$, then $a \mid c$.

Proof By Theorem 1.5.2, there exist integers x and y such that $ax + by = 1$. Multiplying by c we obtain $acx + bcy = c$. Since $a \mid a$, if $a \mid bc$, then by Theorem 1.1.3 (f) we have $a \mid (acx + bcy)$. That is, $a \mid c$. □

A corollary of Theorem 1.6.2 is the following fundamental property of prime numbers.

Theorem 1.6.3. If p is prime and $p \mid ab$, then $p \mid a$ or $p \mid b$.

Proof If $p \mid a$ then we are finished, so we can assume $p \nmid a$. Since the only positive divisors of p are 1 and p, and since $p \nmid a$, we have $\gcd(p, a) = 1$. It now follows from Theorem 1.6.2 that $p \mid b$. □

Theorem 1.6.3 is false if p is not prime. For example, $6 \mid (4 \cdot 9)$ but $6 \nmid 4$ and $6 \nmid 9$.

Theorem 1.6.4. If p is prime and $p \mid a_1 a_2 \cdots a_k$, then $p \mid a_i$ for some i with $1 \leq i \leq k$.

Proof The proof is by induction on k. The result holds for $k = 2$ by Theorem 1.6.3. Now suppose that the result holds for $k = r$ and that $p \mid a_1 a_2 \cdots a_r a_{r+1}$. That is, $p \mid (a_1 a_2 \cdots a_r) a_{r+1}$. By the $k = 2$ result we have $p \mid a_1 a_2 \cdots a_r$ or $p \mid a_{r+1}$. If $p \mid a_1 a_2 \cdots a_r$, then p divides one of $a_1, a_2, \ldots, a_r$ by the inductive hypothesis. Otherwise $p \mid a_{r+1}$. Either way the result holds for $k = r + 1$ and so by induction it holds for all k. □

Theorem 1.6.5. If $a \mid c$ and $b \mid c$ and $\gcd(a, b) = 1$, then $ab \mid c$.

Proof Suppose $a \mid c$. Then we can write $c = ad$ for some integer d. By Theorem 1.6.2, since we have $b \mid ad$ and $\gcd(a, b) = 1$, we have $b \mid d$. Now $a \mid a$ and $b \mid d$ together imply $ab \mid ad$, by Theorem 1.1.3 (e). That is, $ab \mid c$. □

Theorem 1.6.6. Suppose $m_i \mid a$ for $1 \leq i \leq k$, and suppose the m_i are pairwise relatively prime. Then the product $m_1 \cdots m_k$ divides a.

Proof The proof is by induction on k. The result holds for $k = 2$ by Theorem 1.6.5. Now suppose that the result holds for $k = r$, that $m_i \mid a$ for $1 \leq i \leq r + 1$, and that the m_i are pairwise relatively prime. Since the m_i are pairwise relatively prime, we have $\gcd(m_1 m_2 \cdots m_r, m_{r+1}) = 1$, by the inductive hypothesis we have $m_1 m_2 \cdots m_r \mid a$, and we have $m_{r+1} \mid a$. Thus, by the $k = 2$ result we have $(m_1 m_2 \cdots m_r) m_{r+1} \mid a$, and so by induction the result holds for all k. □

1.7 Fundamental Theorem of Arithmetic

We now prove the so-called Fundamental Theorem of Arithmetic which says that every natural number factors into primes in a unique way (the number 1 is considered to be a product of zero prime numbers, with the empty product being defined to be 1). That is, if $n \in \mathbb{N}$, then n can be written uniquely as a product

$$n = p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s}$$

where $p_1, p_2, \ldots, p_s$ are prime numbers, and $a_1, a_2, \ldots, a_s \geq 1$ are integers. By unique, we mean unique up to re-ordering of the factors. The order in which we write the factors is not uniquely determined, but the primes are often written in increasing order. The expression is uniquely determined if we additionally demand $p_1 < p_2 < \cdots < p_s$.

Theorem 1.7.1 (Fundamental Theorem of Arithmetic)**.** Every natural number factors into primes in a unique way. That is, for each $n \in \mathbb{N}$, there exist unique prime numbers $p_1, p_2, \ldots, p_s$ with $p_1 < p_2 < \cdots < p_s$ and integers $a_1, a_2, \ldots, a_s$ with $a_1, a_2, \ldots, a_s \geq 1$ such that

$$n = p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s}.$$

Proof We first show that every natural number can be written as a product of primes. For a contradiction, suppose that this is not the case. Then there is a smallest counterexample $N > 1$ such that N cannot be written as a product of primes. Now, N cannot be prime (or it would be a product of one prime, itself), and so it must have some divisor d with $1 < d < N$. Let $N = dM$. Then $d < N$ and $M < N$ and so by the minimality of N, both d and M can be written a product of primes. This implies that $N = dM$ can also be written as a product of primes. We have a contradiction and conclude that every natural number can be written as a product of primes.

We now show uniqueness of the product. Suppose

$$n = p_1 p_2 \cdots p_s = q_1 q_2 \cdots q_t$$

where $p_1, p_2, \ldots, p_s$ and $q_1, q_2, \ldots, q_t$ are primes. Without loss of generality, we can assume $s \leq t$. Now, $p_1 \mid q_1 q_2 \cdots q_t$ and so by Theorem 1.6.4 we know that $p_1 \mid q_i$ for some i. Since we can relabel the subscripts on $q_1, q_2, \ldots, q_t$, we can assume that $i = 1$. Since q_1 is prime, $p_1 \mid q_1$ implies $p_1 = q_1$. Thus, we have

$$p_2 p_3 \cdots p_s = q_2 q_3 \cdots q_t.$$

We can now repeat this process and cancel out p_2 and q_2. If $r < s$, then we can repeat again and again until we eventually get

$$1 = q_{r+1} \cdots q_s$$

which is impossible. Thus $r = s$ and (after relabeling the subscripts) we have $p_1 = q_1, p_2 = q_2, \ldots, p_s = q_s$. □

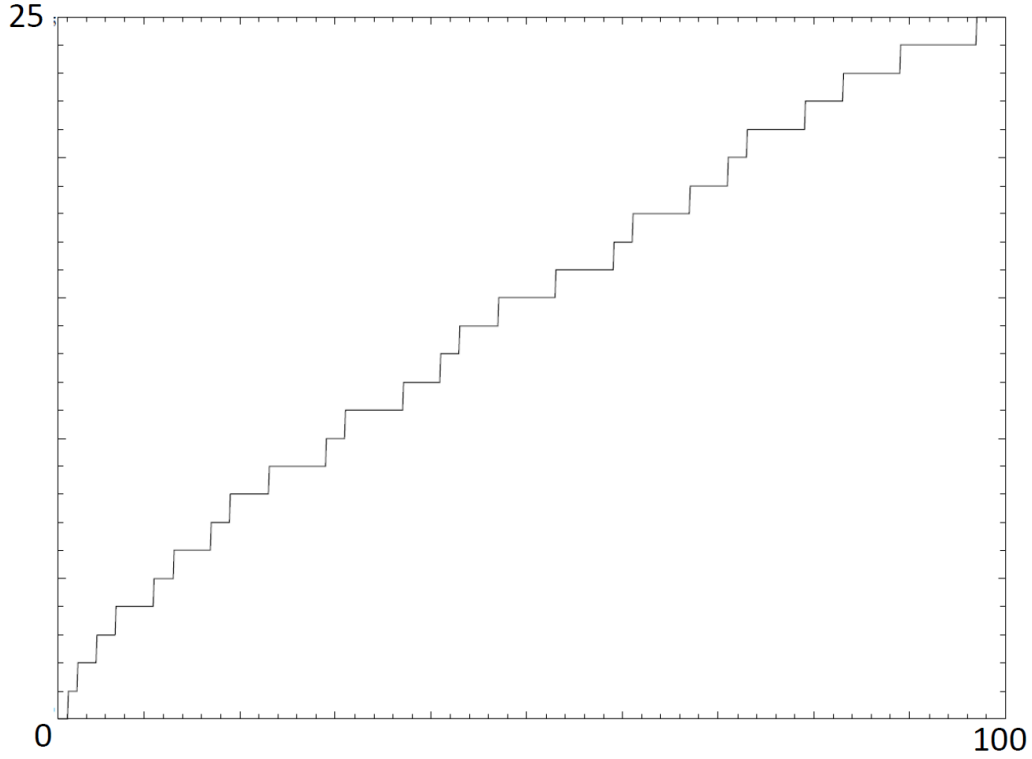
1.8 Distribution of Primes

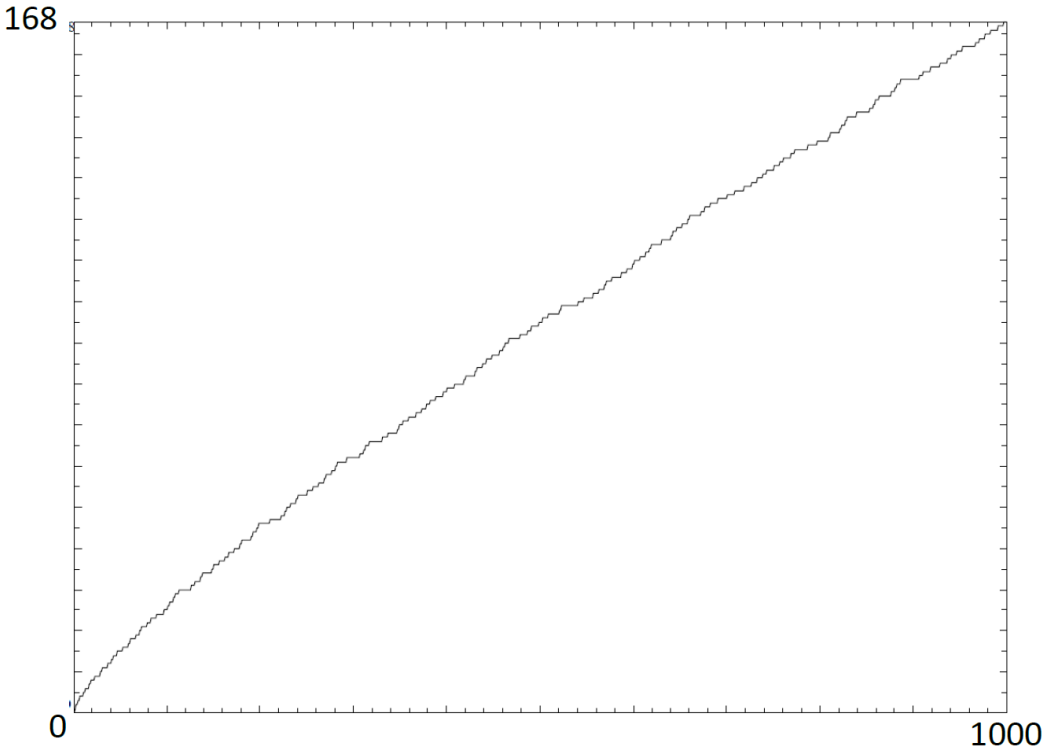
The following theorem is a result of Euclid and is more than 2,300 years old.

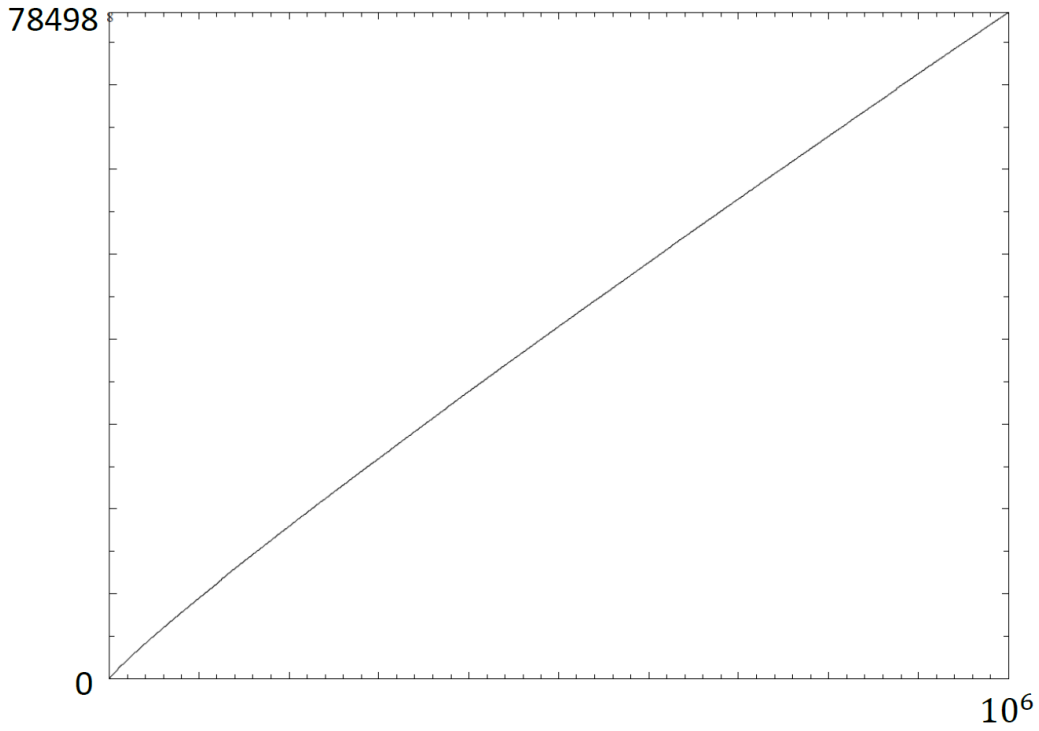
Theorem 1.8.1. There are infinitely many prime numbers.

Proof Suppose $S = \{p_1, p_2, \ldots, p_n\}$ is a given (finite) non-empty set of prime numbers. Let $N = p_1 p_2 \cdots p_n + 1$. According to the Fundamental Theorem of Arithmetic, there exists a prime q such that $q \mid N$. However, all of the primes in S divide $N - 1$ and so $q \notin S$ (if $q \in S$, then $q \mid N - 1$ and $q \mid N$ which implies $q \mid 1$, a contradiction). We have shown that no finite set of prime numbers contains all the prime numbers, and thus conclude that there are infinitely many prime numbers. □

We now briefly discuss the distribution of prime numbers. For any $x \in (0, \infty)$ let $\pi(x)$ be the number of prime numbers less than or equal to x. Below are plots of the function $\pi(x)$ for $1 \leq x \leq 100$, $1 \leq x \leq 1,000$ and $1 \leq x \leq 1,000,000$


25
0
100


168
0
1000



Here is a table showing $\pi(x)$, $\frac{\pi(x)}{x}$ and $\frac{x}{\pi(x)}$ for $x = 10^3, 10^4, 10^5, 10^6, 10^7, 10^8$.

x	10^3	10^4	10^5	10^6	10^7	10^8
$\pi(x)$	168	1229	9592	78498	664579	5761455
$\pi(x)/x$	0.168	0.123	0.0959	0.0785	0.0665	0.0576
$x/\pi(x)$	6.0	8.1	10.4	12.7	15.0	17.4

The value $\pi(x)/x$ represents the probability that an integer chosen at random from $\{1, 2, \ldots, x\}$ is prime. For example, the probability that an integer chosen at random from $\{1, 2, \ldots, 1000\}$ is prime is $\pi(1000)/1000 = 0.168$, and the probability that an integer chosen at random from $\{1, 2, \ldots, 10000\}$ is prime is $\pi(10000)/10000 = 0.123$. It is clear that prime numbers become rarer as we move to larger numbers.

An observation in the above table is that as x increases by a factor of 10, the value of $x/\pi(x)$ seems to increase by about 2.3. This suggests logarithmic growth for $x/\pi(x)$, which the Prime Number Theorem confirms. This theorem was proved independently by Jacques Hadamard and Charles-Jean Étienne Gustave Nicolas de la Vallée Poussin in 1896.

Theorem 1.8.2 (Prime Number Theorem)**.**

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log(x)} = 1.$$

Proof Omitted. Very difficult. □

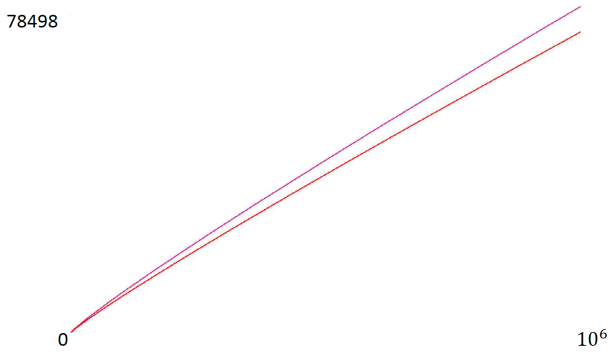
The following table shows the values of $\pi(x)$, $x/\log(x)$ and $\frac{\pi(x)}{x/\log(x)}$ for $x \in \{10^3, 10^4, \ldots, 10^8\}$ and for $x = 10^{24}$.

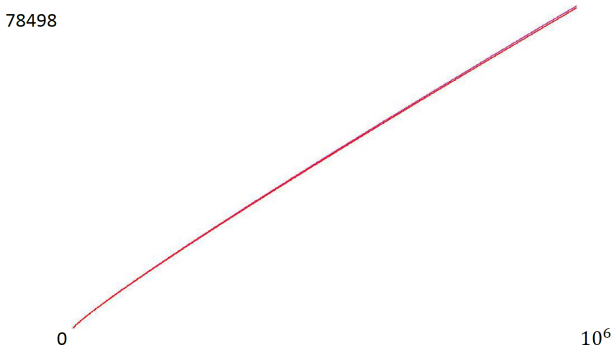
x	$\pi(x)$	$x/\log(x)$	$\frac{\pi(x)}{x/\log(x)}$
10^3	168	145	1.159
10^4	1,229	1,086	1.132
10^5	9,592	8,686	1.104
10^6	78,498	72,382	1.084
10^7	664,579	620,420	1.071
10^8	5,761,455	5,428,661	1.061
$\vdots$	$\vdots$	$\vdots$	$\vdots$
10^{24}			1.019

Although $\lim_{x\to\infty} \dfrac{\pi(x)}{x/\log(x)} = 1$, the difference $\pi(x) - \frac{x}{\log(x)}$ actually grows as x gets large. Better approximations for $\pi(x)$ are known. For example,

$$\pi(x) \simeq \frac{x}{\log(x) - 1}.$$

Shown below is a plot of $\pi(x)$ (top plot at left) and $\frac{x}{\log(x)}$ (bottom plot at left) for $2 \le x \le 10^6$ and a plot of $\pi(x)$ (top plot at right) and $\frac{x}{\log(x)-1}$ (bottom plot at right) for $10 \le x \le 10^6$. The two plots on the right are almost indistinguishable.





The famous *Riemann Hypothesis* is essentially a conjecture about the size of the error term in the Prime Number Theorem.

Chapter 2

Abstract Algebra 1: Semigroups

2.1 Associativity and Semigroups

Definition 2.1.1. Let S be a set. A **binary operation** on S is a function from $S \times S$ to S. If $*$ is a binary operation on S and $a, b \in S$, then $*((a, b))$ is usually denoted by $a * b$. A set S together with a binary operation $*$ on S is often denoted by $(S, *)$, and if several binary operations $*_1, *_2, \ldots, *_t$ are defined on S, then we may write $(S, *_1, *_2, \ldots, *_t)$

Addition, subtraction and multiplication on the integers, the rationals, the reals or the complex numbers are all examples of binary operations. Division is not a binary operation on $\mathbb{R}$ because division by 0 is not defined. Division is a binary operation on $\mathbb{R} \setminus \{0\}$ because for all $a, b \in \mathbb{R} \setminus \{0\}$ we have $a \div b \in \mathbb{R} \setminus \{0\}$. Subtraction is not a binary operation on $\mathbb{N}$ because (for example) $3 - 7$ is not an element of $\mathbb{N}$. Addition and multiplication are binary operations on $\mathbb{N}$. A binary operation, for example multiplication, is often denoted by juxtaposition. That is, we write just ab rather than $a \cdot b$.

Let $\mathcal{F}_A$ be the set of all functions from a set A to itself. Then composition $\circ$ is a binary operation on $\mathcal{F}_A$. Recall that for any set A, if $f : A \to A$ and $g : A \to A$ are functions, then $f \circ g$ is the function from A to A defined by $f \circ g(x) = f(g(x))$ for all $x \in A$.

The set of all n by n matrices with entries from $\mathbb{R}$ is denoted by $M_n(\mathbb{R})$. Matrix addition, matrix subtraction and matrix multiplication are all examples of binary operations on $M_n(\mathbb{R})$.

Observe that in general a binary operation $*$ need not satisfy $a*b = b*a$. Subtraction, composition of functions and matrix multiplication are examples of such binary operations.

Definition 2.1.2. A binary operation $*$ on a set A is **associative** if for all $a, b, c \in A$

$$a * (b * c) = (a * b) * c.$$

Addition and multiplication (of real numbers or matrices) are associative. Composition of functions is also associative because

$$(f \circ (g \circ h))(x) = f((g \circ h)(x)) = f(g(h(x))) = (f \circ g)(h(x)) = ((f \circ g) \circ h)(x).$$

Subtraction is not associative because $(a-b)-c \neq a-(b-c)$. However, it is accepted convention that $a-b-c$ means $(a-b)-c$ rather than $a-(b-c)$.

Most binary operations that we encounter are associative. Associativity allows us to write products and sums such as abc or $a+b+c$ without brackets. For a non-associative binary operation $*$, we need to specify whether $a*b*c$ means $(a*b)*c$ or $a*(b*c)$. The following theorem formalises the preceding claim, and states that it holds for "products" of any length.

Theorem 2.1.3. If $*$ is associative, then any expression of the form $a_1 * a_2 * \cdots * a_n$ is uniquely defined, no matter how parentheses are inserted.

Proof The proof is by (strong) induction. We assume the result for all products of length $m < n$, and show that the result then also holds for products of length n.

If $n \leq 2$ there is nothing to prove, since there are no different ways of associating such products. If $n = 3$ there are two possible ways of inserting parentheses, $(a_1 * a_2) * a_3$ and $a_1 * (a_2 * a_3)$, and associativity guarantees that these give the same result.

Assume the result holds for all products of length $m < n$. Let P be a product $a_1 * a_2 * \cdots * a_n$ where parentheses have been inserted in any legal manner. We shall show that

$$P = (\cdots(((a_1 * a_2) * a_3) * a_4) * \cdots * a_{n-1}) * a_n$$

where the products are grouped (associated) to the left. Thus all parenthesizations of length n are equal.

Now, no matter how the parentheses are inserted in $a_1 * a_2 * \cdots * a_n$, there must be an outermost $*$ operation, the one that is applied last. That is, we can write $P = A * B$ where $A = a_1 * \cdots * a_m$ and $B = a_{m+1} * \cdots * a_n$, both parenthesized in some arbitrary way, with $0 < m < n$.

By the inductive hypothesis,

$$A = (\cdots((a_1 * a_2) * a_3) * \cdots) * a_m$$

and

$$B = (\cdots((a_{m+1} * a_{m+2}) * a_{m+3}) * \cdots) * a_n.$$

If $m = n-1$, then $P = A * a_n = (\cdots((a_1 * a_2) * a_3) * \cdots) * a_{n-1}) * a_n$ and we are done. Otherwise, let $C = (\cdots((a_{m+1} * a_{m+2}) * a_{m+3}) * \cdots) * a_{n-1}$, so $B = C * a_n$. Thus

$$P = A * B = A * (C * a_n) \stackrel{\ddagger}{=} (A * C) * a_n,$$

where $\ddagger$ indicates the use of associativity for 3 arguments.

Since $A * C$ has length $< n$ it is equal to the product with all associations to the left. That is, $A * C = (\cdots((a_1 * a_2) * a_3) * \cdots) * a_{n-1}$, and $P = (\cdots((a_1 * a_2) * a_3) * \cdots a_{n-1}) * a_n$ as claimed. □

Definition 2.1.4. A **semigroup** is an ordered pair $(S, *)$ such that S is a non-empty set, and $*$ is an associative binary operation on S.

Theorem 2.1.5. If $(S, \cdot)$ is a semigroup $a \in S$ and $m, n \in \mathbb{N}$, then

(a) $a^m a^n = a^{m+n}$; and

(b) $(a^m)^n = a^{mn}$.

Proof The expression $a^m a^n$ is a consists of m copies of a followed by n copies of a, which is an expression consisting of $m + n$ copies of a.

$$\overbrace{a \cdot a \cdot \cdots \cdot a}^{m} \cdot \overbrace{a \cdot a \cdot \cdots \cdot a}^{n} = \overbrace{a \cdot a \cdot \cdots \cdot a.}^{m+n}$$

Similarly, the expression $(a^m)^n$ consists of n copies of m copies of a, which is an expression consisting of mn copies of a. □

Note the implicit use of associativity in the proof of Theorem 2.1.5.

2.2 Identities, Inverses and Commutativity

Definition 2.2.1. Let $*$ be a binary operation on a set A. We say that an element $e \in A$ is an **identity** for $(A, *)$ if $a * e = a = e * a$ for all $a \in A$.

Note that an identity must satisfy both conditions $a * e = a$ and $e * a = a$.

Theorem 2.2.2. If an identity exists for $(A, *)$, then it is unique.

Proof Assume that e and e' are both identities. Then $e * e' = e'$ because e is an identity, and $e * e' = e$ because e' is an identity. Hence $e = e'$. □

- 1 is the identity in $(\mathbb{R}, \cdot)$.
- 0 is the identity in $(\mathbb{R}, +)$.
- There is no identity in $(\mathbb{N}, +)$.
- The $n \times n$ identity matrix I_n is the identity in $(M_n(\mathbb{R}), \cdot)$. Recall that $M_n(\mathbb{R})$ is the set of all n by n matrices with entries from $\mathbb{R}$.
- The identity function ι_A is the identity in $(\mathcal{F}_A, \circ)$ where $\mathcal{F}_A$ is the set of all functions from a set A to itself.

In solving equations, and in many other circumstances, an important concept is *cancellation*. That is, if we have $xa = ya$, then we can conclude that $x = y$. We must take care however, for if we are working in $(\mathbb{R}, \cdot)$, then it is not always true that $xa = ya$ implies $x = y$. It is not true when $a = 0$. However, in $(\mathbb{R}, +)$ we can always cancel. For all $x, y, a \in \mathbb{R}$, $x + a = y + a$ implies $x = y$. This is true because we can add the element $-a$ of $\mathbb{R}$ to both sides of the equation $x + a = y + a$ and obtain

$x = y$. The equivalent approach in $(\mathbb{R}, \cdot)$ would be to multiply both sides of $xa = ya$ by $a^{-1} = \frac{1}{a}$, but we cannot do this when $a = 0$ because $\frac{a}{0}$ does not exist.

In $(\mathbb{R}, \cdot)$ we can think of the cancellation process described in the preceding paragraph as follows. Given $xa = ya$, we find an element b such that $ab = 1$. Then we multiply (on the right) both sides of $xa = ya$ by b to obtain $(xa)b = (ya)b$. By associativity of $(\mathbb{R}, \cdot)$ we can rewrite this as $x(ab) = y(ab)$. Since $ab = 1$ this gives us $x \cdot 1 = y \cdot 1$ and hence $x = y$.

Notice that we used only the semigroup properties of $(\mathbb{R}, \cdot)$, the existence of the identity 1, and the existence of b such that $ab = 1$ to achieve the cancellation process of the preceding paragraph. Thus, in a semigroup $(S, *)$ with identity e, if there is an element b such that $ab = e$, then we can cancel out a from an equation $x * a = y * a$ (right cancellation) and deduce $x = y$. Similarly, to cancel out a from an equation $a * x = a * y$ (left cancellation) we need an element b' such that $b' * a = e$. However, if $b' * a = e$ and $a * b = e$ then

$$b' = b' * e = b' * (a * b) = (b' * a) * b = e * b = b.$$

The element b, if it exists is called an *inverse* of a.

Observe that in $(\mathbb{R}, +)$ and in $(\mathbb{R}, \cdot)$, the existence of additive and multiplicative inverses makes the operations of subtraction and division redundant. Instead of $b - a$ we can write $b + (-a)$ where $-a \in \mathbb{R}$ is the additive inverse of a (so there is no subtraction, just addition). Similarly, when $a \neq 0$ instead of $b \div a$ we can write $b \cdot a^{-1}$. Division by 0 is undefined because 0 has no multiplicative inverse.

Definition 2.2.3. If $(S, *)$ has an identity e and $a \in S$, then $b \in S$ is an **inverse** of a if and only if

$$a * b = e = b * a.$$

An element having an inverse is said to be **invertible**.

Theorem 2.2.4. Let $(S, *)$ be a semigroup with identity e. If $a \in S$ has an inverse, then the inverse is unique.

Proof Suppose b and c are inverses of a. Then

$$b = e * b = (c * a) * b = c * (a * b) = c * e = c.$$

□

Definition 2.2.5. Let $(S, *)$ be a semigroup with identity, and let $a \in S$ be invertible. Then a^0 is defined to be the identity, a^{-1} is defined to be the inverse of a, and a^{-n} is defined by $a^{-n} = (a^{-1})^n$ for all $n \in \mathbb{N}$.

An exception to the above definition is that if the binary operation is addition, or some analogue of addition, then the inverse of a is usually denoted by $-a$.

- In $(\mathbb{R}, \cdot)$ every $a \in \mathbb{R}$ except $a = 0$ has inverse $a^{-1} = \frac{1}{a}$.
- In $(\mathbb{R}, +)$ every $a \in \mathbb{R}$ has inverse $-a$.

- In $(M_n(\mathbb{R}), \cdot)$ the invertible elements are the matrices with nonzero determinant.
- In $(\mathcal{F}_A, \circ)$ (where $\mathcal{F}_A$ is the set of all functions from a set A to itself and $\circ$ denotes composition of functions) the invertible elements are precisely the bijections.

The above points illustrate why the same notation, namely an exponent of -1, is used for the multiplicative inverse x^{-1} of a real number x, the inverse A^{-1} of a matrix A, and the inverse f^{-1} of a function f. They are all examples of the same concept when we work in the abstract setting of semigroups with identities.

Theorem 2.2.6. Let $(S, *)$ be a semigroup with identity e.

(a) If a is invertible, then so is a^{-1} and $(a^{-1})^{-1} = a$.

(b) If a and b are invertible, then so is $a * b$ and $(a * b)^{-1} = b^{-1} * a^{-1}$.

Proof

(a) To show that the inverse of a^{-1} is a we need to show that $a^{-1} * a = e = a * a^{-1}$, and this is true by the definition of a^{-1}.

(b) We have

$$(a * b) * (b^{-1} * a^{-1}) = a * (b * b^{-1}) * a^{-1} = a * e * a^{-1} = a * a^{-1} = e.$$

Similarly,

$$(b^{-1} * a^{-1}) * (a * b) = b^{-1} * (a^{-1} * a) * b = b^{-1} * e * b = b^{-1} * b = e.$$

Thus the inverse of $a * b$ exists and is equal to $b^{-1} * a^{-1}$.

□

The next theorem extends Theorem 2.1.5. When a is invertible, a^n is defined for any integer n (not just positive integers), and in this case the result of Theorem 2.1.5 holds for any integer exponents.

Theorem 2.2.7. If $(S, \cdot)$ is a semigroup with identity, $a \in S$ is invertible, and $m, n \in \mathbb{Z}$, then

(a) $a^m a^n = a^{m+n}$; and

(b) $(a^m)^n = a^{mn}$.

Proof

(a) We check all of the possible cases. If m, $n > 0$ the result is Theorem 2.1.5. If m or $n = 0$ the result is obvious. Suppose $m > 0$ but $n < 0$. Note that $|n| = -n > 0$. We have $a^m a^n = a^m (a^{-1})^{|n|}$. If $m \geq |n|$ we may cancel out all the a^{-1} terms one by one, leaving $a^{m-|n|} = a^{m+n}$. If $m < |n|$ we may cancel out all the a terms in a^m, leaving $(a^{-1})^{|n|-m}$. By definition, this is $a^{-(|n|-m)} = a^{m-|n|} = a^{m+n}$ as required. The proof is similar if $m < 0$ but $n > 0$. Finally if m, $n < 0$ let $b = a^{-1}$. Then $a^m a^n = a^{-|m|} a^{-|n|} = (a^{-1})^{|m|} (a^{-1})^{|n|}$ by definition, and this is $b^{|m|} b^{|n|} = b^{|m|+|n|}$ by the positive case. This final expression is $(a^{-1})^{(|m|+|n|)} = a^{-(|m|+|n|)}$ (by definition), which is a^{m+n}.

(b) The proof of (b) is similar to that of (a).

□

Definition 2.2.8. A binary operation $*$ on a set A is **commutative** if for all a, $b \in A$

$$a * b = b * a.$$

Multiplication of real numbers and addition of real numbers are both commutative. None of subtraction, matrix multiplication nor composition of functions is commutative.

2.3 Definition of a Group

Definition 2.3.1. A **group** consists of a nonempty set G together with a binary operation $*$ on G satisfying

Associativity	For all $a, b, c \in G$, $a * (b * c) = (a * b) * c$.
Identity	There exists $e \in G$ such that $e * a = a = a * e$ for every $a \in G$.
Inverses	For every $a \in G$ there exists $b \in G$ such that $a * b = e = b * a$.

Thus, a group is a semigroup with an identity in which every element has an inverse.

The fact that $*$ is a binary operation on G means that $* : G \times G \to G$, so $a * b \in G$ for every a, $b \in G$. Sometimes this property of groups is stated as an additional axiom called "closure", but it holds automatically according to our definition of binary operation.

If the binary operation of a group is understood (clear from the context), then we may refer to the group simply as G. If it is necessary to specify the binary operation, then we may refer to the group $(G, *)$, or we may write "G under $*$" or similar.

Example 2.3.2.

(a) The integers under addition $(\mathbb{Z}, +)$ is a group. We check the axioms. Firstly, $+$ is a binary operation on $\mathbb{Z}$ because the sum of any two integers is an integer. Associativity holds because for all $a, b, c \in \mathbb{Z}$ we have $a + (b + c) = (a + b) + c$. There is an identity, namely 0, because $0 + a = a = a + 0$ for all $a \in \mathbb{Z}$. Finally, for every $a \in \mathbb{Z}$ we have $-a \in \mathbb{Z}$ and $a + (-a) = 0 = (-a) + a$, so every element has an inverse.

(b) The real numbers under addition $(\mathbb{R}, +)$ and the rational numbers under addition $(\mathbb{Q}, +)$ are both groups. The proof is similar to the proof that $(\mathbb{Z}, +)$ is a group.

(c) The natural numbers under addition $(\mathbb{N}, +)$ is not a group. There is no identity.

(d) The real numbers under multiplication $(\mathbb{R}, \cdot)$ is not a group. There is an identity (which is unique), namely 1, but 0 has no inverse.

(e) The nonzero real numbers under multiplication $(\mathbb{R}\setminus\{0\}, \cdot)$ is a group. If $a, b \in \mathbb{R}\setminus\{0\}$, then $ab \in \mathbb{R} \setminus \{0\}$ so $\cdot$ is indeed a binary operation. Associativity is clear, and follows from associativity of $\mathbb{R}$. There is an identity, namely 1, and for every $a \in \mathbb{R} \setminus \{0\}$ we have $aa^{-1} = 1 = a^{-1}a$, and $a^{-1} = \frac{1}{a} \in \mathbb{R} \setminus \{0\}$.

(f) The set G consisting of a single element e is a group. Here the only possible binary operation is defined by $e \cdot e = e$. Associativity is clear because any product is e. Also e is the identity, and e is its own inverse. This group is called the **trivial group**.

(g) The set of all n by n matrices with entries from $\mathbb{R}$, denoted $M_n(\mathbb{R})$, is a group under addition but not under multiplication. Some matrices have no multiplicative inverses, namely those with determinant 0.

The subset of $M_n(\mathbb{R})$ consisting of all invertible matrices forms a group under multiplication. It is denoted $GL_n(\mathbb{R})$, where "GL" stands for "General Linear". If A and B are invertible, then so is AB. The inverse of AB is $(AB)^{-1} = B^{-1}A^{-1}$. Thus matrix multiplication is a binary operation on $GL_n(\mathbb{R})$. Matrix multiplication is associative, and associativity of $(GL_n(\mathbb{R}), \cdot)$ follows from this. The identity matrix is invertible (it is its own inverse) and is the identity of the group. Finally every element of $GL_n(\mathbb{R})$ is invertible, by definition.

(h) For any set A, the set S_A of all bijections from A to A is a group under composition. Composition is indeed a binary operation because the composition of two bijections is a bijection. We saw earlier that composition of functions is associative, so composition of bijections is also associative. The identity function ι_A (given by $\iota_A(x) = x$ for all $x \in A$) is a bijection and is the identity of the group. We have $\iota_a \circ f = f = f \circ \iota_a$ for all $f \in S_A$ because

$$(\iota_a \circ f)(x) = \iota_A(f(x)) = f(x) = f(\iota_A(x)) = (f \circ \iota_A)(x).$$

Finally, for each $f \in S_A$ let f^{-1} be the function defined for all $y \in A$ by $f^{-1}(y) = x$ where x is the unique element of A such that $f(x) = y$. Then we have $f \circ f^{-1}(x) = f(f^{-1}(x)) = x$ and $f^{-1} \circ f(x) = f^{-1}(f(x)) = x$. Thus we have $f \circ f^{-1} = \iota_A = f^{-1} \circ f$ and so f^{-1} is the inverse of f.

In Example 2.3.2, (e), (g) and (h) are special cases of the following theorem.

Theorem 2.3.3. Let $(S, *)$ be a semigroup with identity, and let G be the subset of S consisting of all invertible elements. Then $(G, *)$ is a group.

Proof The identity element e of $(S, *)$ is its own inverse. Thus, $e \in G$ and G is non-empty. If a and b are invertible, then so is $a * b$ by Theorem 2.2.6 (b). Thus, G is closed under $*$ (that is, $*$ is a binary operation on G). Multiplication in S is associative, so it is still associative in the subset G. The element $e \in G$ and is the identity in $(G, *)$. It remains to show that every element of G has an inverse in G. If $a \in G$, then by the definition of G we have $b \in S$ such that $a * b = b * a = e$. So we just need to show that $b \in G$. Theorem 2.2.6 (a) says that in a semigroup the inverse of an invertible element is also invertible. So $b \in G$. □

Chapter 3

Number Theory 2: Modular Arithmetic

3.1 Modular Arithmetic

Definition 3.1.1. For each $n \in \mathbb{N}$, let $\sim_n$ be the relation defined on $\mathbb{Z}$ as follows. For all $a, b \in \mathbb{Z}$, $a \sim_n b$ if and only if $n \mid (a-b)$. If $a \sim_n b$, then we write $a \equiv b \pmod{n}$ and say that a **is congruent to** b **modulo** n.

Recall that an equivalence relation is a relation $\sim$ that is reflexive, symmetric and transitive. The equivalence class of any element $a \in S$ is denoted by $[a]$ and is defined by $[a] = \{x \in S : x \sim a\}$. The set of equivalence classes of an equivalence relation on S forms a partition of S.

Theorem 3.1.2. For all $n \in \mathbb{N}$, the relation $\sim_n$ on $\mathbb{Z}$ is an equivalence relation.

Proof For all $a \in \mathbb{Z}$ we have $n \mid (a-a)$, so $\sim_n$ is a reflexive relation. If $a \sim_n b$, then $n \mid (a-b)$ which implies $n \mid (b-a)$ and hence $b \sim_n a$. So $\sim_n$ is a symmetric relation. If $a \sim_n b$ and $b \sim_n c$, then $n \mid (a-b)$ and $n \mid (b-c)$ so $n \mid ((a-b)+(b-c))$. That is, $n \mid (a-c)$. Thus, $a \sim_n c$ and so $\sim_n$ is a transitive relation. We have shown that $\sim_n$ is reflexive, symmetric and transitive and hence we have shown that it is an equivalence relation. □

Definition 3.1.3. For each $a \in \mathbb{Z}$, the equivalence class of a under the relation $\sim_n$ is denoted by $[a]_n$ and is called the **congruence class of** a **modulo** n.

Theorem 3.1.4. There are n congruence classes modulo n, namely $[0]_n, [1]_n, \ldots, [n-1]_n$.

Proof For distinct $a, b \in \{0, 1, \ldots, n-1\}$, we have $0 < |a-b| < n$ which means that $n \nmid a-b$. Thus, $[0]_n, [1]_n, \ldots, [n-1]_n$ are distinct congruence classes modulo n. Moreover, for any integer a there exist integers q and r such that $a = qn + r$, equivalently $qn = a - r$, and $0 \leq r \leq n-1$. Thus, $n \mid a - r$ which means that $a \equiv r \pmod{n}$ and so a belongs to one of the congruence classes $[0]_n, [1]_n, \ldots, [n-1]_n$. □

Theorem 3.1.5. Let $n \in \mathbb{N}$ and let $a, b, c, d \in \mathbb{Z}$. If $a \equiv c \pmod{n}$ and $b \equiv d \pmod{n}$, then $a + b \equiv c + d \pmod{n}$ and $ab \equiv cd \pmod{n}$.

Proof If $a \equiv c \pmod{n}$ and $b \equiv d \pmod{n}$, then we have $n \mid (a-c)$ and $n \mid (b-d)$. Thus, $n \mid ((a-c)+(b-d))$ so $n \mid ((a+b)-(c+d))$. That is, $a+b \equiv c+d \pmod{n}$. Also, if $a \equiv c \pmod{n}$ and $b \equiv d \pmod{n}$, then we have integers s and t such that $a-c = sn$ and $b-d = tn$. That is, $a = sn+c$ and $b = tn+d$. Thus,

$$ab = (sn+c)(tn+d) = n(stn+sd+tc)+cd$$

and so we have $n \mid (ab-cd)$ and hence $ab \equiv cd \pmod{n}$. □

Theorem 3.1.5 allows us to greatly simplify calculations in many situations. It is used implicitly in the examples of the next section. However, we need to take care when using modular arithmetic with exponents. It is not true that $a \equiv b \pmod{n}$ implies $c^a \equiv c^b \pmod{n}$. For example, $5 \equiv 2 \pmod{3}$, but $2^5 \not\equiv 2^2 \pmod{3}$.

3.2 Modular Arithmetic Examples

Example 3.2.1. What is the remainder when 599×373 is divided by 3?

Since $599 \equiv 2 \pmod{3}$ and $373 \equiv 1 \pmod{3}$, we have $599 \times 373 \equiv 2 \times 1 \pmod{3}$. Thus, the remainder is 2.

Example 3.2.2. Mary leaves home at 9 o'clock and returns after 88 hours. What is the time when Mary returns?

Since $88 \equiv 4 \pmod{12}$, we have $9+88 \equiv 9+4 \equiv 1 \pmod{12}$ so the time when Mary returns is 1 o'clock. If we want to know whether the time is am or pm, then we need to know whether Mary left at 9am or 9pm, and we need to work modulo 24 rather than modulo 12. Suppose Mary left at 9pm, which is 21:00 on a 24 hour clock. Since $88 \equiv 16 \pmod{24}$, we have $21+88 \equiv 21+16 \equiv 13 \pmod{24}$ so the time when Mary returns is 13:00 or 1pm.

Example 3.2.3. You may have encountered the rule that a number is divisible by 3 if and only if the sum of its digits is divisible by 3. For example, to determine whether 3 divides 728 we observe that $7+2+8 = 17$ and $3 \nmid 17$, and conclude that $3 \nmid 728$. However, $3 \mid 729$ because $7+2+9 = 18$ and $3 \mid 18$. In fact, the following stronger statement holds. Any positive integer is equivalent modulo 3 to the sum of its digits. We now prove this statement using modular arithmetic.

Consider the number $N = a_n a_{n-1} \ldots a_0$. By this we mean N has digits $a_n, a_{n-1}, \ldots, a_0$ as read from left to right. So if $N = 4271$, then $n = 3$, $a_3 = 4$, $a_2 = 2$, $a_1 = 7$ and $a_0 = 1$. We have

$$N = a_n \cdot 10^n + a_{n-1} \cdot 10^{n-1} + \cdots + a_2 \cdot 10^2 + a_1 \cdot 10^1 + a_0 \cdot 10^0$$

Now, $10 \equiv 1 \pmod{3}$ and so $10^x \equiv 10 \times 10 \times \cdots \times 10 \equiv 1 \times 1 \times \cdots \times 1 \equiv 1 \pmod{3}$. Thus,

$$N \equiv a_n \cdot 1 + a_{n-1} \cdot 1 + \cdots + a_2 \cdot 1 + a_1 \cdot 1 + a_0 \cdot 1 \equiv a_n + a_{n-1} + \cdots + a_2 + a_1 + a_0 \pmod{3}.$$

That is, N is equivalent modulo 3 to the sum of its digits.

Example 3.2.4. What is the last decimal digit of 3^{2018}?

We have $3^1 \equiv 3 \pmod{10}$, $3^2 \equiv 9 \pmod{10}$, $3^3 \equiv 7 \pmod{10}$ and $3^4 \equiv 1 \pmod{10}$. We make use of the last expression $3^4 \equiv 1 \pmod{10}$. Since $2018 = 4 \cdot 504 + 2$, we have

$$3^{2018} = 3^{4 \cdot 504 + 2} = (3^4)^{504} \cdot 3^2.$$

Since $3^4 \equiv 1 \pmod{10}$, we thus have

$$3^{2018} \equiv 1^{504} \cdot 3^2 \equiv 9 \pmod{10}$$

which means that the last decimal digit of 3^{2018} is 9.

Example 3.2.5. Show that $11 \mid (3^{32} + 2)$.

We repeatedly square mod 11 as follows.

$$\begin{array}{rcccccl} & & 3^2 & \equiv & 9 & & \\ 3^4 & = & (3^2)^2 & \equiv & 9^2 & \equiv & 4 \pmod{11} \\ 3^8 & = & (3^4)^2 & \equiv & 4^2 & \equiv & 5 \pmod{11} \\ 3^{16} & = & (3^8)^2 & \equiv & 5^2 & \equiv & 3 \pmod{11} \\ 3^{32} & = & (3^{16})^2 & \equiv & 3^2 & \equiv & 9 \pmod{11} \end{array}$$

We have $3^{32} \equiv 9 \pmod{11}$ which gives us $3^{32} + 2 \equiv 0 \pmod{11}$. Thus, $11 \mid (3^{32} + 2)$.

Example 3.2.6. Find the last 2 decimal digits of 2^{98}.

We work mod 100.

$$\begin{array}{rcccccl} & & 2^2 & \equiv & 4 & & \\ 2^4 & = & (2^2)^2 & \equiv & 4^2 & \equiv 16 & \pmod{100} \\ 2^8 & = & (2^4)^2 & \equiv & 16^2 & \equiv 56 \equiv -44 & \pmod{100} \\ 2^{16} & = & (2^8)^2 & \equiv & (-44)^2 & \equiv 36 & \pmod{100} \\ 2^{32} & = & (2^{16})^2 & \equiv & 36^2 & \equiv 96 \equiv -4 & \pmod{100} \\ 2^{64} & = & (2^{32})^2 & \equiv & (-4)^2 & \equiv 16 & \pmod{100} \end{array}$$

Now, $98 = 64 + 32 + 2$, so

$$2^{98} = 2^{64} \cdot 2^{32} \cdot 2^2 \equiv 16 \cdot (-4) \cdot 4 \equiv -(16^2) \equiv 44 \pmod{100}.$$

So $2^{98} \equiv 44 \pmod{100}$ and the last 2 decimal digits of 2^{98} are both 4s.

3.3 The Groups $(\mathbb{Z}_n, +)$ and $(\mathbb{Z}_n^*, \cdot)$

Definition 3.3.1. The set of congruence classes modulo n is denoted by $\mathbb{Z}_n$.

Theorem 3.1.5 allows us to define binary operations $\oplus$ and $\odot$ on $\mathbb{Z}_n$ as follows.

Definition 3.3.2. For all $[a]_n, [b]_n \in \mathbb{Z}_n$, we define $[a]_n \oplus [b]_n = [a + b]_n$ and $[a]_n \odot [b]_n = [ab]_n$.

The binary operations $\oplus$ and $\odot$ are well-defined. That is, we get the same answer regardless of which representatives we choose for the congruence classes. Theorem 3.1.5 guarantees that for any $c \in [a]_n$ and any $d \in [b]_n$ if we use c and d rather than a and b in our calculations of $[a]_n \oplus [b]_n$ and $[a]_n \odot [b]_n$, then we get the same result. This is because

$$[c]_n \oplus [d]_n = [c+d]_n = [a+b]_n = [a]_n \oplus [b]_n$$

and

$$[c]_n \odot [d]_n = [cd]_n = [ab]_n = [a]_n \odot [b]_n.$$

The binary operations $\oplus$ and $\odot$ are fundamental attributes of $\mathbb{Z}_n$, and they are usually referred to as addition and multiplication respectively. Often when we talk about $\mathbb{Z}_n$ it is assumed that we mean $\mathbb{Z}_n$ together with its binary operations $\oplus$ and $\odot$, although sometimes $\mathbb{Z}_n$ refers just to the set $\{[0]_n, [1]_n, \ldots, [n-1]_n\}$. The meaning should be clear from the context.

From now on, when working in $\mathbb{Z}_n$, we will often just write a, or sometimes $[a]$ rather than $[a]_n$, $a+b$ rather than $[a]_n \oplus [b]_n$, and ab rather than $[a]_n \odot [b]_n$. We will also often write $a-b$ rather than $[a]_n \oplus [-b]_n$. However, it should be kept in mind that when we are using a to denote the congruence class $[a]_n$, that a is not an integer but rather a set of integers (namely the set of all integers that are congruent to a modulo n). If there is any occasion where the context does not make it clear whether we are working in $\mathbb{Z}$ or $\mathbb{Z}_n$, then we will revert to the unabbreviated notation.

Many algebraic properties of the integers also hold in $\mathbb{Z}_n$, as the following theorem shows.

Theorem 3.3.3. For any classes $[a]$, $[b]$, $[c] \in \mathbb{Z}_n$

(a) $[a] \oplus ([b] \oplus [c]) = ([a] \oplus [b]) \oplus [c]$

(b) $[a] \oplus [0] = [a] = [0] \oplus [a]$.

(c) $[a] \oplus [-a] = [0] = [-a] \oplus [a]$.

(d) $[a] \oplus [b] = [b] \oplus [a]$.

(e) $[a] \odot ([b] \odot [c]) = ([a] \odot [b]) \odot [c]$

(f) $[a] \odot [1] = [a] = [1] \odot [a]$.

(g) $[a] \odot [b] = [b] \odot [a]$.

(h) $[a] \odot ([b] \oplus [c]) = ([a] \odot [b]) \oplus ([a] \odot [c])$.

(i) $([a] \oplus [b]) \odot [c] = ([a] \odot [c]) \oplus ([b] \odot [c])$.

Proof We prove (a) (associativity of $\oplus$) as follows.

$$[a] \oplus ([b] \oplus [c]) = a \oplus [b+c] = [a+(b+c)] = [(a+b)+c] = [a+b] \oplus [c] = ([a] \oplus [b]) \oplus [c].$$

Notice that the step $[a+(b+c)] = [(a+b)+c]$ is just using associativity of addition on the integers. Each of (b)-(i) can be proved similarly to (a), and this is left as an exercise. □

Theorem 3.3.4. For all $n \in \mathbb{N}$, $(\mathbb{Z}_n, +)$ is a group.

Proof Theorem 3.3.3 (a) gives us associativity, Theorem 3.3.3 (b) tells us that $[0]$ is the identity, and Theorem 3.3.3 (c) tells us that $[-a]$ is the inverse of $[a]$ for each $[a] \in \mathbb{Z}_n$. □

Although, $(\mathbb{Z}_n, +)$ is a group, $(\mathbb{Z}_n, \cdot)$ is never a group (except in the trivial case where $n = 1$) because the element $[0]$ has no inverse. Some elements of $(\mathbb{Z}_n, \cdot)$ have inverses and some do not. For example, in $(\mathbb{Z}_{14}, \cdot)$ the element $[2]$ is not invertible whereas $[3]$ has inverse $[5]$.

In $\mathbb{Z}_n$, multiplicative inverses (inverses in $(\mathbb{Z}_n, \cdot)$) are much more interesting than additive inverses (inverses in $(\mathbb{Z}_n, +)$), and if we use the term *inverse* without qualification, then it should be assumed that we are referring to multiplicative inverse. The term *invertible* without qualification similarly refers to elements that have multiplicative inverses.

The following theorem tells us exactly which elements of $\mathbb{Z}_n$ are invertible. The theorem requires $a > 0$ so that $\gcd(a, n)$ is defined. However, if $a \leq 0$ and we wish to know whether $[a]_n$ invertible, then we can simply choose $a' \in [a]_n$ such that $a' > 0$ and calculate $\gcd(a', n)$ instead.

Theorem 3.3.5. Let $a \geq 0$. The congruence class $[a]_n$ is invertible in $(\mathbb{Z}_n, \cdot)$ if and only if $\gcd(a, n) = 1$.

Proof We have $[a]_n \odot [b]_n = [b]_n \odot [a]_n = [ab]_n$ and so we have that $[a]_n$ is invertible if and only if there exists an integer b such that $ab \equiv 1 \pmod{n}$. But $ab = 1 \pmod{n}$ if and only if $n \mid 1 - ab$, and $n \mid 1 - ab$ if and only if there exists a $q \in \mathbb{Z}$ such that that $nq = 1 - ab$, equivalently $ab + nq = 1$. By Theorem 1.5.2, this last equation has a solution in integers b and q if and only if $\gcd(a, n) = 1$. □

Implicit in the proof of Theorem 3.3.5 is the fact that if $a, a' > 0$ and $a \equiv a' \pmod{n}$, then $\gcd(a, n) = 1$ if and only if $\gcd(a', n) = 1$. Otherwise, we might get different answers as to whether $[a]_n$ is invertible depending on which representative we choose for $[a]_n$. In fact, we have the following theorem.

Theorem 3.3.6. Let $a, b > 0$. If $a \equiv b \pmod{n}$, then $\gcd(a, n) = \gcd(b, n)$.

Proof If $a \equiv b \pmod{n}$, then $a - b = qn$ for some $q \in \mathbb{Z}$. That is, $a = qn + b$. By Theorem 1.4.1, this implies that $\gcd(a, n) = \gcd(b, n)$. □

Example 3.3.7. Which elements of $\mathbb{Z}_{12}$ are invertible?

The elements of $\mathbb{Z}_{12}$ are $\{[0], [1], \ldots, [11]\}$. For $a \in \{1, 5, 7, 9\}$ we have $\gcd(a, 12) = 1$ and for $a \in \{2, 3, 4, 6, 8, 9, 10\}$ we have $\gcd(a, 12) > 1$. Thus, the equivalence classes $[1], [5], [7], [11]$ are invertible, and $[2], [3], [4], [6], [8], [9], [10]$ are not invertible. Finally, since $[0] = [12]$ and $\gcd(12, 12) = 12 > 1$, $[0]$ also is not invertible. It can be checked that each of $[1], [5], [7], [11]$ is its own inverse in $\mathbb{Z}_{12}$.

Definition 3.3.8. For each $n \in \mathbb{N}$, we define $\mathbb{Z}_n^*$ to be the set of invertible elements in $(\mathbb{Z}_n, \cdot)$.

Theorem 3.3.9. For all $n \in \mathbb{N}$, $(\mathbb{Z}_n^*, \cdot)$ is a group.

Proof Theorem 3.3.3 (e) and (f) tell us that $(\mathbb{Z}_n, \cdot)$ is a semigroup with identity. It thus follows by Theorem 2.3.3 that $(\mathbb{Z}_n^*, \cdot)$ is a group. □

Theorem 3.3.10. If p is prime, then every nonzero element of $\mathbb{Z}_p$ is invertible. Thus, $(\mathbb{Z}_p \setminus \{0\}, \cdot)$ is a group for each prime p.

Proof The fact that every nonzero element of $\mathbb{Z}_p$ is invertible follows immediately from Theorem 3.3.5 because when p is prime, $\gcd(a, p) = 1$ for each $a \in \{1, 2, \ldots, p-1\}$. Thus, $\mathbb{Z}_p \setminus \{0\} = \mathbb{Z}_p^*$ and Theorem 3.3.9 tells us that $(\mathbb{Z}_p \setminus \{0\}, \cdot) = (\mathbb{Z}_p^*, \cdot)$ is a group. □

Theorem 3.3.10 shows that $\mathbb{Z}_p$ and $\mathbb{R}$ have some interesting similarities. Both $(\mathbb{R}, +)$ and $(\mathbb{Z}_p, +)$ are groups, and so are $(\mathbb{R} \setminus \{0\}, \cdot)$ and $(\mathbb{Z}_p \setminus \{0\}, \cdot)$.

3.4 Finding Inverses in $(\mathbb{Z}_n, \cdot)$

Theorem 3.3.5 tells us that the congruence class $[a]_n$ is invertible if and only if $\gcd(a, n) = 1$. To actually find the inverse of $[a]_n$, we can use the Extended Euclidean GCD Algorithm to find integers x and y such that $ax + ny = 1$. The congruence class $[x]_n$ is then the inverse of $[a]_n$ because we have $ax \equiv 1 \pmod{n}$. The value of y is not needed.

Example 3.4.1. Determine whether 11 has an inverse in $\mathbb{Z}_{80}$, and find the inverse if it exists.

We use the Euclidean GCD Algorithm to find $\gcd(11, 80)$ (it is obvious that $\gcd(11, 80) = 1$ but we will need the calculation of the gcd via the Euclidean GCD Algorithm in what follows).

$$\begin{array}{rcrcr} 80 & = & 7 \cdot 11 & + & 3 \\ 11 & = & 3 \cdot 3 & + & 2 \\ 3 & = & 1 \cdot 2 & + & 1 \\ 2 & = & 2 \cdot 1 & + & 0 \end{array}$$

Thus, $\gcd(80, 11) = 1$ and 11 has an inverse in $\mathbb{Z}_{80}$. We now use the Extended Euclidean GCD Algorithm to determine 11^{-1}.

$$\begin{array}{rcrcr} 1 & = & -2 & + & 3 \\ & = & -(-3 \cdot 3 + 11) & + & 3 \\ & = & 4 \cdot 3 & - & 11 \\ & = & 4 \cdot (-7 \cdot 11 + 80) & - & 11 \\ & = & -29 \cdot 11 & + & 4 \cdot 80 \end{array}$$

Thus, $-29 \cdot 11 \equiv 1 \pmod{80}$ and so $11^{-1} = 51 \pmod{80}$. This is easily checked, $11 \cdot 51 = 561 \equiv 1 \pmod{80}$.

Example 3.4.2. Find x such that $11x \equiv 4 \pmod{80}$.

We have seen in the above example that $11^{-1} = 51 \pmod{80}$. Thus, we multiply both sides by 51 as follows; $11x \equiv 4 \pmod{80}$ if and only if $51 \cdot 11x \equiv 51 \cdot 4 \pmod{80}$. But $51 \cdot 11 \equiv 1 \pmod{80}$ and $51 \cdot 4 \equiv 44 \pmod{80}$, so we have $11x \equiv 4 \pmod{80}$ if and only if $x \equiv 44 \pmod{80}$.

Note that finding an x such that $11x \equiv 4 \pmod{80}$ is equivalent to finding an integer solution to the linear Diophantine equation $11x + 80y = 4$ (except that in the former case we don't care about the value of y). To see this, observe that $11x \equiv 4 \pmod{80}$ if and only if $80 \mid (4 - 11x)$ if and only if there exists an integer y such that $80y = 4 - 11x$.

3.5 Chinese Remainder Theorem

Consider the following problem.

Example 3.5.1. Find an integer x such that

$$\begin{aligned} x &\equiv 4 \,(\text{mod } 7) \\ x &\equiv 9 \,(\text{mod } 11) \\ x &\equiv 3 \,(\text{mod } 13) \end{aligned}$$

We first find an integer e_1 such that

$$e_1 \equiv 1 \,(\text{mod } 7) \qquad e_1 \equiv 0 \,(\text{mod } 11) \qquad e_1 \equiv 0 \,(\text{mod } 13).$$

We have $11 \times 13 = 143$. So $143 \equiv 0 \,(\text{mod } 11)$ and $143 \equiv 0 \,(\text{mod } 13)$. But $143 \equiv 3 \,(\text{mod } 7)$. However, since 3 has inverse 5 in $\mathbb{Z}_7$, if we multiply 143 by 5, then the result will be congruent to $1 \,(\text{mod } 7)$, and still congruent to $0 \,(\text{mod } 11)$ and $0 \,(\text{mod } 13)$. Thus, we take

$$e_1 = 143 \times 5 = 715$$

and we have $e_1 \equiv 1 \,(\text{mod } 7)$, $e_1 \equiv 0 \,(\text{mod } 11)$ and $e_1 \equiv 0 \,(\text{mod } 13)$ as required.

Proceeding in a similar manner, we can find e_2 such that

$$e_2 \equiv 0 \,(\text{mod } 7) \qquad e_2 \equiv 1 \,(\text{mod } 11) \qquad e_2 \equiv 0 \,(\text{mod } 13)$$

and e_3 such that

$$e_3 \equiv 0 \,(\text{mod } 7) \qquad e_3 \equiv 0 \,(\text{mod } 11) \qquad e_3 \equiv 1 \,(\text{mod } 13)$$

For e_2, we have $7 \times 13 = 91$. So $91 \equiv 0 \,(\text{mod } 7)$ and $91 \equiv 0 \,(\text{mod } 13)$. But $91 \equiv 3 \,(\text{mod } 11)$. However, since 3 has inverse 4 in $\mathbb{Z}_{11}$, if we multiply 91 by 4, then the result will be congruent to $1 \,(\text{mod } 11)$, and still congruent to $0 \,(\text{mod } 7)$ and $0 \,(\text{mod } 13)$. Thus, we take

$$e_2 = 91 \times 4 = 364$$

and we have $e_2 \equiv 0 \,(\text{mod } 7)$, $e_2 \equiv 1 \,(\text{mod } 11)$ and $e_2 \equiv 0 \,(\text{mod } 13)$ as required. For e_3, we have $7 \times 11 = 77$. So $77 \equiv 0 \,(\text{mod } 7)$ and $77 \equiv 0 \,(\text{mod } 11)$. But $77 \equiv 12 \,(\text{mod } 13)$. However, since 12 has inverse 12 in $\mathbb{Z}_{13}$, if we multiply 77 by 12, then the result will be congruent to $1 \,(\text{mod } 13)$, and still congruent to $0 \,(\text{mod } 7)$ and $0 \,(\text{mod } 11)$. Thus, we take

$$e_3 = 77 \times 12 = 924$$

and we have $e_3 \equiv 0 \,(\text{mod } 7)$, $e_2 \equiv 0 \,(\text{mod } 11)$ and $e_2 \equiv 1 \,(\text{mod } 13)$ as required.

Now, let $x = 4e_1 + 9e_2 + 3e_3$. Since $e_1 \equiv 1 \,(\text{mod } 7)$, $e_1 \equiv 0 \,(\text{mod } 11)$ and $e_1 \equiv 0 \,(\text{mod } 13)$, we have $x \equiv 4 \,(\text{mod } 7)$. Similarly, since $e_2 \equiv 0 \,(\text{mod } 7)$, $e_2 \equiv 1 \,(\text{mod } 11)$ and $e_2 \equiv 0 \,(\text{mod } 13)$, we have $x \equiv 9 \,(\text{mod } 11)$, and since $e_3 \equiv 0 \,(\text{mod } 7)$, $e_3 \equiv 0 \,(\text{mod } 11)$ and $e_3 \equiv 1 \,(\text{mod } 13)$, we have $x \equiv 3 \,(\text{mod } 13)$. So we have an x as required, namely $x = 4 \cdot 715 + 9 \cdot 364 + 3 \cdot 924 = 8908$.

The product $7 \cdot 11 \cdot 13 = 1001$ is congruent to $0 \,(\text{mod } 7)$, $0 \,(\text{mod } 11)$ and $0 \,(\text{mod } 13)$. So if x is a solution to our system of congruences, then so is $x + 1001t$ for any integer t. Thus, the smallest positive solution is $x = 8908 - 8 \cdot 1001 = 900$ and the smallest, in absolute value, solution is $x = 900 - 1001 = -101$.

Example 3.5.1 generalises and the result is known as the Chinese Remainder Theorem.

Theorem 3.5.2 (Chinese Remainder Theorem)**.** Let $m_1, \ldots, m_k$ be pairwise relatively prime positive integers and let $a_1, a_2, \ldots, a_k \in \mathbb{Z}$. Then the simultaneous system of congruences

$$\begin{aligned} x &\equiv a_1 \,(\text{mod } m_1) \\ x &\equiv a_2 \,(\text{mod } m_2) \\ &\vdots \\ x &\equiv a_k \,(\text{mod } m_k) \end{aligned}$$

is solvable. Moreover the solution is unique modulo $m_1 m_2 \cdots m_n$.

Proof Let $M = m_1 m_2 \cdots m_k$ and for $i = 1, 2, \ldots, k$, let $M_i = \frac{M}{m_i}$. Then $M_i \equiv 0 \,(\text{mod } m_j)$ for each $j \in \{1, 2, \ldots, k\} \setminus \{m_i\}$ and M_i is invertible in $\mathbb{Z}_{m_i}$ (because M_i is a product of integers that are relatively prime to m_i). For $i = 1, 2, \ldots, k$, let N_i be the inverse of M_i in $\mathbb{Z}_{m_i}$, and let $e_i = M_i N_i$. Thus, we have $e_i \equiv 0 \,(\text{mod } m_j)$ for each $j \in \{1, 2, \ldots, k\} \setminus \{i\}$ and $e_i \equiv 1 \,(\text{mod } m_i)$, and if we let $x = a_1 e_1 + a_2 e_2 + \cdots + a_k e_k$, then x is a solution to the system of congruences.

We now prove uniqueness modulo $m_1 m_2 \cdots m_k$. If x and y are both solutions, then for $i = 1, 2, \ldots, k$ we have $x - y \equiv 0 \,(\text{mod } m_i)$. That is, each m_i divides $x - y$. Since the m_i are relatively prime, this implies that their product $m_1 m_2 \cdots m_k$ divides $x - y$ (see Theorem 1.6.6). That is, $y \equiv x \,(\text{mod } m_1 m_2 \cdots m_k)$. □

Example 3.5.3. Solve the equation $x^2 + 1 \equiv 0 \,(\text{mod } 85)$.

We have $85 = 5 \cdot 17$ and $\gcd(5, 17) = 1$. If $x^2 + 1 \equiv 0 \,(\text{mod } 85)$, then $x^2 + 1 \equiv 0 \,(\text{mod } 5)$ and $x^2 + 1 \equiv 0 \,(\text{mod } 17)$. Conversely, if $x^2 + 1 \equiv 0 \,(\text{mod } 5)$ and $x^2 + 1 \equiv 0 \,(\text{mod } 17)$, then $x^2 + 1 \equiv 0 \,(\text{mod } 85)$. Thus, $x^2 + 1 \equiv 0 \,(\text{mod } 85)$ if and only if $x^2 + 1 \equiv 0 \,(\text{mod } 5)$ and $x^2 + 1 \equiv 0 \,(\text{mod } 17)$.

By inspection, the only solutions to $x^2 + 1 \equiv 0 \,(\text{mod } 5)$ are $x = 2$ or $3 \,(\text{mod } 5)$, and the only solutions to $x^2 + 1 \equiv 0 \,(\text{mod } 17)$ are $x = 4$ or $13 \,(\text{mod } 17)$. Thus, there are four possibilities.

(a) $x \equiv 2 \,(\text{mod } 5)$ and $x \equiv 4 \,(\text{mod } 17)$;

(b) $x \equiv 2 \,(\text{mod } 5)$ and $x \equiv 13 \,(\text{mod } 17)$;

(c) $x \equiv 3 \,(\text{mod } 5)$ and $x \equiv 4 \,(\text{mod } 17)$; and

(d) $x \equiv 3 \,(\text{mod } 5)$ and $x \equiv 13 \,(\text{mod } 17)$.

Each of these possibilities can be solved using the Chinese Remainder Theorem.

$$\begin{aligned} x \equiv 2 \,(\text{mod } 5) \text{ and } x \equiv 4 \,(\text{mod } 17) &\xrightarrow{CRT} x \equiv 72 \,(\text{mod } 85); \\ x \equiv 2 \,(\text{mod } 5) \text{ and } x \equiv 13 \,(\text{mod } 17) &\xrightarrow{CRT} x \equiv 47 \,(\text{mod } 85); \\ x \equiv 3 \,(\text{mod } 5) \text{ and } x \equiv 4 \,(\text{mod } 17) &\xrightarrow{CRT} x \equiv 38 \,(\text{mod } 85); \text{ and} \\ x \equiv 3 \,(\text{mod } 5) \text{ and } x \equiv 13 \,(\text{mod } 17) &\xrightarrow{CRT} x \equiv 13 \,(\text{mod } 85). \end{aligned}$$

Thus, $x \equiv 13, 38, 47$ or $72 \,(\text{mod } 85)$.

Chapter 4

Abstract Algebra 2: Groups

4.1 Basic Properties of Groups

Recall that a group consists of a nonempty set G together with a binary operation $*$ on G satisfying

Associativity	For all $a, b, c \in G$, $a * (b * c) = (a * b) * c$.
Identity	There exists $e \in G$ such that $e * a = a = a * e$ for every $a \in G$.
Inverses	For every $a \in G$ there exists $b \in G$ such that $a * b = e = b * a$.

Examples of groups are given in Example 2.3.2 and we have also seen that $(\mathbb{Z}_n, +)$ and $(\mathbb{Z}_n^*, \cdot)$ are groups.

Most groups are thought of in a multiplicative sense and we use *multiplicative notation*. That is, the binary operation of the group is thought of as a "multiplication". Thus, we sometimes drop the symbol representing the binary operation and just use juxtaposition. That is, for a group $(G, *)$ we sometimes write ab rather than $a * b$. Similarly, the notation a^{-1} is used for inverses and the product $a * a \cdots * a$ (n copies of a) is denoted by a^n. The notation a^n is also defined for an integer $n \leq 0$, see Definition 2.2.5.

Definition 4.1.1. A group $(G, *)$ is **abelian** or **commutative** if its binary operation is commutative, that is if $xy = yx$ for all $x, y \in G$.

Examples of abelian groups that we have seen include $(\mathbb{R}, +)$, $(\mathbb{R} \setminus \{0\}, \cdot)$, $(\mathbb{Z}, +)$, $(\mathbb{Z}_n, +)$, $(\mathbb{Z}_n^*, \cdot)$ and $(M_n(\mathbb{R}), +)$. Examples of non-abelian groups include $GL_n(\mathbb{R})$ (the group of $n \times n$ invertible matrices with real entries under matrix multiplication) and $(S_A, \circ)$ where S_A is the set of all bijections from a set A to itself and $\circ$ is composition of functions.

Although multiplicative notation is often used for general groups, the situation is different for abelian groups where *additive notation* is often used. When a group G is abelian, we often think of it additively and denote the binary operation by $+$ (unless it has a well-established other name). Examples of abelian groups where the binary operation has a well-established other name are $(\mathbb{R}^*, \cdot)$ and $(\mathbb{Z}_n^*, \cdot)$, and multiplicative notation is used for these groups. When additive notation is used, we denote the identity by 0, the inverse of x by $-x$, and $x + x + \cdots + x$ by nx.

Theorem 4.1.2. Let G be a group, let $a, b, c \in G$, and let $m, n \in \mathbb{Z}$.

(a) If $ab = ac$, then $b = c$ (left cancellation). If $ba = ca$, then $b = c$ (right cancellation).

(b) $a^m a^n = a^{m+n}$.

(c) $(a^m)^n = a^{mn}$.

(d) $(ab)^{-1} = b^{-1}a^{-1}$

(e) If G is abelian, then $(ab)^n = a^n b^n$.

Proof

(a) If $ab = ac$, then multiplying by a^{-1} on the left we obtain $a^{-1}(ab) = a^{-1}(ac)$ so $(a^{-1}a)b = (a^{-1}a)c$ so $1b = 1c$ so $b = c$. The proof for right cancellation works similarly.

(b) Both (b) and (c) were proved in Theorem 2.2.7.

(d) This was proved in Theorem 2.2.6.

(e) If G is abelian, then $(ab)^n = (ab)(ab)\cdots(ab)$ where the number of copies of ab is n. Using associativity and commutativity, we can rearrange $(ab)(ab)\cdots(ab)$ into $a^n b^n$.

□

Definition 4.1.3. The **order** of a group G is the cardinality $|G|$ of the set G. A group of finite order is called a **finite group** and a group of infinite order is called an **infinite group**.

Definition 4.1.4. Let $G = \{g_1, \ldots, g_n\}$ be a finite group. We can list all possible products $g_i g_j$ in an $n \times n$ table. This table is called the **Cayley table** or **group table** of G.

Example 4.1.5. Give the Cayley table for the group $(\mathbb{Z}_4, +)$.
In position (i, j) of the table the entry is $i + j \pmod 4$.

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

Theorem 4.1.6. If G is a finite group, then the Cayley table of G contains each element of G exactly once in each row and column.

Proof Let $n = |G|$ and let $G = \{g_1, g_2, \ldots, g_n\}$. For $i = 1, 2, \ldots, n$, the entries in the row indexed by g_i are $g_i g_1$, $g_i g_2$, $\ldots$, $g_i g_n$. These are pairwise distinct because if $g_i g_j = g_i g_k$, then $g_j = g_k$. A similar argument works for the columns. □

Definition 4.1.7. Let $(G, *)$ and $(H, \circ)$ be groups. The **direct product** of G and H is the set $G \times H$ with binary operation $\odot$ defined by

$$(g_1, h_1) \odot (g_2, h_2) = (g_1 * g_2,\ h_1 \circ h_2).$$

Note that if G has m elements and H has n elements then $G \times H$ has mn elements:

$$|G \times H| = |G| \cdot |H|$$

Theorem 4.1.8. If G and H are groups, then $G \times H$ is a group.

Proof Let $(G, *)$ and $(H, \circ)$ be groups and let $\odot$ be the binary operation on $G \times H$ given by

$$(g_1, h_1) \odot (g_2, h_2) = (g_1 * g_2,\ h_1 \circ h_2).$$

If $(g_1, h_1), (g_2, h_2) \in G \times H$ then $(g_1, h_1) \odot (g_2, h_2) = (g_1 * g_2, h_1 \circ h_2)$. Since $*$ and $\circ$ are binary operations on G and H, $g_1 * g_2 \in G$ and $h_1 \circ h_2 \in H$. Thus, $(g_1, h_1) \odot (g_2, h_2) \in G \times H$, and so $\odot$ is a binary operation on $G \times H$.
Associativity: If (g_1, h_1), (g_2, h_2), $(g_3, h_3) \in G \times H$, then

$$\begin{aligned}
\Big((g_1, h_1) \odot (g_2, h_2)\Big) \odot (g_3, h_3) &= (g_1 * g_2,\ h_1 \circ h_2) \odot (g_3, h_3) \\
&= \Big((g_1 * g_2) * g_3,\ (h_1 \circ h_2) \circ h_3\Big) \\
&= (g_1 * (g_2 * g_3),\ h_1 \circ (h_2 \circ h_3)) \\
&= (g_1, h_1) \odot (g_2 * g_3, h_2 \circ h_3) \\
&= (g_1,\ h_1) \odot \Big((g_2, h_2) \odot (g_3, h_3)\Big)
\end{aligned}$$

so $\odot$ is associative.
Identity: If $(g, h) \in G \times H$, then

$$\begin{aligned}
(g, h) \odot (1_G, 1_H) &= (g * 1_G,\ h \circ 1_H) \\
&= (g, h) \\
&= (1_G * g,\ 1_H \circ h) \\
&= (1_G,\ 1_H) \odot (g, h)
\end{aligned}$$

Thus $(1_G,\ 1_H)$ is the identity in $G \times H$.
Inverses: If $(g, h) \in G \times H$, then

$$\begin{aligned}
(g, h) \odot (g^{-1}, h^{-1}) &= (g * g^{-1},\ h \circ h^{-1}) \\
&= (1_G,\ 1_H) \\
&= (g^{-1} * g,\ h^{-1} \circ h) \\
&= (g^{-1},\ h^{-1}) \odot (g, h)
\end{aligned}$$

So the inverse of (g, h) is (g^{-1}, h^{-1}). □

Definition 4.1.9. The direct product of groups $(G_1, *_1), (G_2, *_2), \ldots, (G_n, *_n)$ is denoted by

$$G_1 \times G_2 \times \cdots \times G_n$$

and is the set $G_1 \times G_2 \times \cdots \times G_n$ together with the binary operation $*$ defined by

$$(a_1, a_2, \ldots, a_n) * (b_1, b_2, \ldots, b_n) = (a_1 *_1 b_1, a_2 *_2 b_2, \ldots, a_n *_n b_n).$$

The direct product $G \times G \times \cdots \times G$ of n copies of G is denoted G^n.

Theorem 4.1.10. If $G_1, G_2, \ldots, G_n$ are groups, then $G_1 \times G_2 \times \cdots \times G_n$ is a group.

Proof The proof is similar to the proof for the case $n = 2$. □

4.2 Symmetric Groups

We saw in Example 2.3.2 (h) that the set of all bijections from A to A is a group under composition. A bijection from A to A is also called a **permutation** of A. A convenient way to describe and work with permutations is by using their *cycle representation*, which we now define and use hereafter.

Let A be a non-empty finite set and let θ be a permutation of A. Since A is finite and since θ is a permutation, for any $a \in A$, there is a smallest positive integer k such that $\theta^k(a) = a$. Moreover, $a,\ \theta(a),\ \theta^2(a),\ \ldots,\ \theta^{k-1}(a)$ are pairwise distinct elements of A. Thus, the elements of A can be partitioned into cycles where each cycle $(a_0\ a_1\ \ldots\ a_{k-1})$ satisfies

$$a_0 \mapsto a_1 \mapsto a_2 \mapsto \cdots \mapsto a_{k-1} \mapsto a_0.$$

A cycle $(a_0\ a_1\ \ldots\ a_{k-1})$ is said to have **length** k. A **cycle representation** for a permutation is given by listing all its cycles in this manner. The **cycle structure** of a permutation is the sequence of the lengths of its cycles (in non-increasing order).

If a is a fixed point of θ, then the cycle of θ containing a is (a), and has length 1. Sometimes cycles of length 1 are omitted from a cycle representation of a permutation, with the understanding that any elements of A not appearing in the cycle representation of θ are fixed points of θ (however, any cycles of length 1 must still be counted in the cycle structure of θ). The notation () or (1) may be used to represent the identity permutation.

As an example, the permutation θ of $\{1, 2, 3, 4, 5, 6\}$ given by

$$1 \mapsto 4,\ 2 \mapsto 6,\ 3 \mapsto 2,\ 4 \mapsto 1,\ 5 \mapsto 5,\ 6 \mapsto 3$$

has cycle representation $\theta = (1\ 4)(2\ 6\ 3)(5)$ or just $\theta = (1\ 4)(2\ 6\ 3)$. The cycles of a cycle representation of a permutation can be written in any order, and the elements within each cycle can be cyclically permuted, with any one of the elements of the cycle appearing first. So, if the cycle of length 1 is omitted, then there are twelve different equivalent ways of writing a cycle representation of the permutation θ given above, namely

$$\begin{array}{llllll} (1\ 4)(2\ 6\ 3), & (1\ 4)(6\ 3\ 2), & (1\ 4)(3\ 2\ 6), & (4\ 1)(2\ 6\ 3), & (4\ 1)(6\ 3\ 2), & (4\ 1)(3\ 2\ 6), \\ (2\ 6\ 3)(1\ 4), & (6\ 3\ 2)(1\ 4), & (3\ 2\ 6)(1\ 4), & (2\ 6\ 3)(4\ 1), & (6\ 3\ 2)(4\ 1), & (3\ 2\ 6)(4\ 1). \end{array}$$

It should be clear that the cycle structure of a permutation is independent of which particular cycle representation is used. The partition given by the cycles is also unique to the permutation.

Note that if $\theta = (x_1\ x_2\ \cdots\ x_k)(y_1\ y_2\ \cdots\ y_\ell)$, then θ is equal to the composition

$$\theta = (x_1\ x_2\ \cdots\ x_k) \circ (y_1\ y_2\ \cdots\ y_\ell)$$

of the two permutations having cycle representations $(x_1\ x_2\ \cdots\ x_k)$ and $(y_1\ y_2\ \cdots\ y_\ell)$. Thus, each cycle within any given cycle representation, may be thought of as a permutation in its own right, and the cycle representation may be thought of as a composition of the permutations corresponding to the individual cycles.

When composing cycle representations, it must be remembered that the permutation on the right acts first and the permutation on the left acts last. For example,

$$(1\ 2\ 3) \circ (1\ 4\ 5)(2\ 3) = (1\ 4\ 5\ 2).$$

However, if $(a_1\ a_2\ \cdots\ a_k)$ and $(b_1\ b_2\ \cdots\ b_\ell)$ are disjoint cycles (that is, $\{a_1, a_2, \ldots, a_k\} \cap \{b_1, b_2, \ldots, b_\ell\} = \emptyset$), then $(a_1\ a_2\ \cdots\ a_k) \circ (b_1\ b_2\ \cdots\ b_\ell) = (b_1\ b_2\ \cdots\ b_\ell) \circ (a_1\ a_2\ \cdots\ a_k)$. This is why the order of the cycles in a cycle representation of a permutation does not matter.

Note that there are two subtly different meanings of the term permutation in common usage. First, a permutation of A is a bijective function from A to itself, and this is the definition that we will be using. The second is that a permutation is simply a listing of the elements of A in some order.

Definition 4.2.1. Let A be a finite non-empty set. The group of all permutations of A under composition is called the **symmetric group on** A and is denoted $\mathrm{Sym}(A)$. If $|A| = n$, then the group $\mathrm{Sym}(A)$ is a **symmetric group of degree** n and the notation S_n may be used.

Theorem 4.2.2. S_n is a group of order $n!$.

Proof The fact that S_n is a group was proved in Example 2.3.2 (h). It has $n!$ elements because there are $n!$ distinct permutations of a set with n elements. □

Example 4.2.3. Let θ and τ be permutations of $\{1, 2, 3\}$ with $\theta = (1\ 2\ 3)$ and $\tau = (2\ 3)$. Then $\sigma, \tau \in S_3$. The group operation is composition. For example $\sigma\tau$ maps 2 to $\sigma(\tau(2)) = \sigma(3) = 1$. Thus, we have

$$\sigma\tau = (1\ 2\ 3)(2\ 3) = (1\ 2)$$

and

$$\tau\sigma = (2\ 3)(1\ 2\ 3) = (1\ 3).$$

Note that $\sigma\tau \neq \tau\sigma$ and so S_3 is not abelian. With further calculation we determine that

$$S_3 = \{\iota, \sigma, \sigma^2, \tau, \sigma\tau, \tau\sigma\} = \{(), (1\ 2\ 3), (1\ 3\ 2), (2\ 3), (1\ 2), (1\ 3)\}.$$

It can be checked that $\tau\sigma = \sigma^2\tau$ and so equally we could write

$$S_3 = \{\iota, \sigma, \sigma^2, \tau, \sigma\tau, \sigma^2\tau\}.$$

The Cayley Table for S_3 is as follows.

S_3	1	σ	σ^2	τ	$\sigma\tau$	$\sigma^2\tau$
1	1	σ	σ^2	τ	$\sigma\tau$	$\sigma^2\tau$
σ	σ	σ^2	1	$\sigma\tau$	$\sigma^2\tau$	τ
σ^2	σ^2	1	σ	$\sigma^2\tau$	τ	$\sigma\tau$
τ	τ	$\sigma^2\tau$	$\sigma\tau$	1	σ^2	σ
$\sigma\tau$	$\sigma\tau$	τ	$\sigma^2\tau$	σ	1	σ^2
$\sigma^2\tau$	$\sigma^2\tau$	$\sigma\tau$	τ	σ^2	σ	1

4.3 Subgroups

Often we encounter a smaller group within a larger one. For example, the group $(\mathbb{Z}, +)$ is contained in the larger group $(\mathbb{R}, +)$.

Definition 4.3.1. Let G be a group. If H is a non-empty subset of G such that

(a) For all x, $y \in H$, $xy \in H$;

(b) $1 \in H$; and

(c) For all $x \in H$, $x^{-1} \in H$;

then H is a **subgroup** of G. Property (a) is called **closure**. If H is a subgroup of G, then we write $H \leq G$.

These properties imply that H is a group in its own right under the group operation of G: (a) ensures that the group operation of G restricts to a binary operation on H. That is, when we multiply elements in H we stay in H, not just in G. Since the operation is associative on all of G, it will certainly be associative on H. Finally (b) and (c) ensure that the identity and inverses are in H.

Example 4.3.2.

(a) $(\mathbb{Z}, +) \leq (\mathbb{Q}, +)$ and $(\mathbb{Q}, +) \leq (\mathbb{R}, +)$.

(b) $\mathbb{N}$ is not a subgroup of $(\mathbb{Z}, +)$. It is closed under the group operation $+$, but it lacks the identity 0 and inverses.

(c) Let G be a group with identity 1. Then $G \leq G$ and $\{1\} \leq G$.

(d) Let $G = \mathbb{Z}_4 = \{0, 1, 2, 3\}$. Let $H = \{0, 2\}$ and $K = \{0, 3\}$. Then H is a subgroup of G, but K is not. The set H is non-empty and if x, $y \in H$ then $x + y = 0 + 0$ or $0 + 2$ or $2 + 0$ or $2 + 2$, all of which are elements of H. Thus H is closed. Also, H contains the identity 0 and inverses ($-0 = 0$ and $-2 = 2$). However $3 + 3 = 2 \notin K$. So K is not closed and hence is not a subgroup of G.

We can combine the three subgroup properties into a single test:

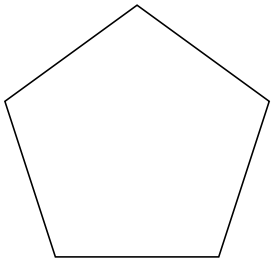
Theorem 4.3.3. Let G be a group and let H be a non-empty subset of G. Then $H \leq G$ if and only if for all $x, y \in H$ we have $xy^{-1} \in H$.

Proof If $H \leq G$ and $x, y \in H$ then $y^{-1} \in H$, and hence $xy^{-1} \in H$.

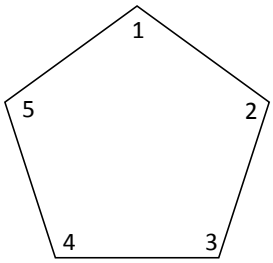
Now suppose that for all $x, y \in H$ we have $xy^{-1} \in H$. Since H is non-empty, there exists some $x \in H$. Thus $xx^{-1} = 1 \in H$ and so the identity of G is in H. Also, if $x \in H$, then $1x^{-1} \in H$ and so $x^{-1} \in H$. Thus the inverse of any element of H is in H. Finally if $x, y \in H$ then (by what we have just proved) $x, y^{-1} \in H$ and so $x(y^{-1})^{-1} = xy \in H$. This proves H is closed. □

4.4 Dihedral Groups

A regular pentagon has lots of symmetry. There are several rotations and reflections that leave the pentagon unchanged. It can be rotated in the plane by $0, \frac{2\pi}{5}, \frac{4\pi}{5}, \frac{6\pi}{5}$ or $\frac{8\pi}{5}$ (that is $0°, 72°, 144°, 216°$ or $288°$), or it can be reflected about a line through a vertex and the midpoint of the opposite side. There are 5 such reflections, one for each vertex.



Now label the positions of vertices of the pentagon $1, 2, 3, 4$ and 5. For each $x \in \{1, 2, 3, 4, 5\}$, each rotation or reflection θ of the pentagon moves the vertex in position x to a (possibly) new position $\theta(x)$. Thus, each rotation or reflection induces a permutation of $\{1, 2, 3, 4, 5\}$. For example, with the labelling below,



a clockwise rotation in the plane by $\frac{2\pi}{5}$ induces the permutation

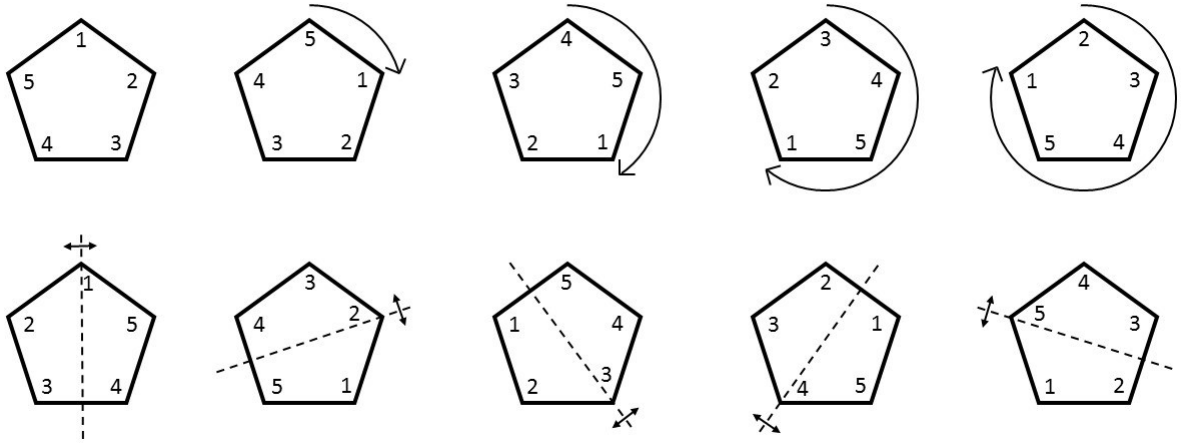
$$\sigma = (1\ 2\ 3\ 4\ 5)$$

and a reflection about the line through 1 and the midpoint of the side opposite 1 induces the permutation

$$\tau = (2\ 5)(3\ 4)$$

The permutations of $\{1, 2, 3, 4, 5\}$ induced by rotations and reflections of the pentagon are called *symmetries* of the pentagon.

The 10 symmetries mentioned above, 5 rotations and 5 reflections, are indicated in the figure below. For each symmetry, the corresponding position to which the labels are moved is shown.



The 10 symmetries of a regular pentagon listed above are the only symmetries. To see this, observe that the relative locations of the labels must be preserved. That is, 1 must always be adjacent to 2 and 5 and not adjacent to 3 and 4, and so on. There are 5 possibilities for the position of 1, and once the position of 1 is fixed, there are two possibilities for the position of 2 (either to the left or to the right of 1). Once the positions of 1 and 2 are fixed, the positions of the remaining labels $3, 4$ and 5 are all determined. Thus, there can be at most $5 \times 2 = 10$ distinct symmetries.

Let D_5 denote the set of symmetries of a regular pentagon. Then D_5 is a subgroup of the group $\mathrm{Sym}(\{1, 2, 3, 4, 5\})$, and is called the *symmetry group of the regular pentagon*. It is clear that D_5 is a nonempty subset of $\mathrm{Sym}(\{1, 2, 3, 4, 5\})$, the composition of any two symmetries in D_5 is in D_5 (a symmetry followed by another symmetry is another symmetry), the identity is in D_5, and the inverse of any symmetry in D_5 is in D_5 (the inverse of any symmetry is a symmetry). Thus, D_5 is a subgroup of $\mathrm{Sym}(\{1, 2, 3, 4, 5\})$, see Definition 4.3.1.

It can be checked that

$$D_5 = \{1, \sigma, \sigma^2, \sigma^3, \sigma^4, \tau, \sigma\tau, \sigma^2\tau, \sigma^3\tau, \sigma^4\tau\},$$

where σ and τ are as defined above. Geometrically, σ is a clockwise rotation by $\frac{2\pi}{5}$ and τ is a reflection about the line through 1 and the midpoint of the side opposite 1.

The construction of the group D_5 as the symmetry group of the regular pentagon generalises as follows.

Definition 4.4.1. Let $n \geq 3$. The set of symmetries of the regular n-gon is denoted D_n.

Theorem 4.4.2. For all $n \geq 3$, D_n is a group of order $2n$ and is a subgroup of S_n. Indeed,

$$D_n = \{1, \sigma, \sigma^2, \ldots, \sigma^{n-1}, \tau, \sigma\tau, \sigma^2\tau, \ldots, \sigma^{n-1}\tau\}$$

where

$$\sigma = (1\ 2\ 3\ \cdots\ n-1\ n),$$

$$\tau = (1)(2\ n)(3\ n-1)(4\ n-2)\cdots(\tfrac{n+1}{2}\ \tfrac{n+3}{2}) \text{ when } n \text{ is odd,}$$

and

$$\tau = (1)(2\ n)(3\ n-1)(4\ n-2)\cdots(\tfrac{n}{2}\ \tfrac{n+4}{2})(\tfrac{n+2}{2}) \text{ when } n \text{ is even.}$$

Definition 4.4.3. Let $n \geq 3$. The group D_n of symmetries of a regular n-gon is called the **dihedral group of degree** n.

In general, a subgroup of S_n is called a *permutation group of degree* n.

Example 4.4.4. The group $D_3 = \{1, \sigma, \sigma^2, \tau, \sigma\tau, \sigma^2\tau\}$ has 6 elements and so must contain every permutation of $\{1, 2, 3\}$. Thus, $D_3 = S_3$, see Example 4.2.3. This makes sense because the symmetries of a triangle admit every permutation of its vertices.

Many puzzles can be viewed as problems about elements of S_n.

Example 4.4.5 (Rubik's Cube)**.** Consider the $3 \times 3 \times 3$ Rubik's cube. Fix an orientation of the cube in space by fixing where the central squares are (eg blue centre on top, white on right, green in front, or whatever). Label the positions of the $6 \cdot 9 = 54$ coloured squares (stickers) in some fixed way. (1 to 9 across the top face, 10 to 18 on the front face etc).

Any move of the cube can be represented as a bijection $\{1, \ldots, 54\} \to \{1, \ldots, 54\}$ sending $j \mapsto k$ if the sticker in position j moves to position k. The collection of all cube moves forms a subgroup C of S_{54}.

There are 6 basic moves of the cube; a clockwise rotation by $\pi/2$ for each of the 6 faces. These moves are denoted by F, U, L, R, B, D for a clockwise rotation by $\pi/2$ of the front, upper, left, right, back and down (bottom) faces respectively. Thus $F^4 = 1$ etc. We say that the cube group is *generated* by these 6 elements.

Given a scrambled cube, it is easy to write down the permutation α of the original identity state that led to its current state. To unscramble the cube, we need to write α^{-1} (efficiently) in terms of combinations of powers of F, U, L, R, B, D. This is not so easy ...

It can be shown that C is "equal to" (isomorphic to) the group

$$C = \Big((\mathbb{Z}_3^7) \times (\mathbb{Z}_2^{11})\Big) \rtimes \Big((A_8 \times A_{12}) \rtimes \mathbb{Z}_2\Big).$$

Here $A \rtimes B$ denotes the "semidirect product" of A and B, which is the set $A \times B$ with a certain new operation defined on it, and A_m is the alternating group of degree m, which we will encounter in Section 4.12. The alternating group A_m is a subgroup of S_m consisting of half the elements of S_m.

Thus $|C| = 3^7 \cdot 2^{11} \cdot \frac{8!}{2} \cdot \frac{12!}{2} \cdot 2 = 2^{27} \cdot 3^{14} \cdot 5^3 \cdot 7^2 \cdot 11 = 43252003274489856000 \simeq 4.3 \times 10^{19}$. (Still much smaller than $|S_{54}| = 54! \simeq 2.3 \times 10^{71}$.)

4.5 Order of Group Elements

The order of a finite group G is the cardinality of G. We now define the order of a group element.

Definition 4.5.1. Let G be a group with identity 1 and let $a \in G$. We say that a has **finite order** if $a^n = 1$ for some positive integer n. In this case, the **order** of a in G is the smallest such positive integer. If no such n exists, then we say that a has infinite order.

In an abelian group written additively, a has order n if $\overbrace{a + \cdots + a}^{n} = 0$ with n copies of a added together, and no smaller positive number of a's adds to 0.

Example 4.5.2.

(a) In any group, the identity has order 1, and no other element has order 1 (because the identity is unique).

(b) In the group $(\mathbb{R} \setminus \{0\}, \cdot)$, 1 has order 1, -1 has order 2, and every other element has infinite order. If $x \in \mathbb{R} \setminus \{0\}$ and $|x| < 1$, then $|x^n| < 1$ for any positive integer n. Similarly, if $x \in \mathbb{R} \setminus \{0\}$ and $|x| > 1$, then $|x^n| > 1$ for any positive integer n.

(c) In the group $\mathbb{Z}_2 \times \mathbb{Z}_3$, the element $(1, 1)$ has order 6. Repeatedly adding $(1, 1)$ we get: $(1, 1)$, $(1, 1) + (1, 1) = (2, 2) = (0, 2)$, 3 copies of $(1, 1)$ add to $(1, 3) = (1, 0)$, 4 copies add to $(2, 1) = (0, 1)$, 5 copies to $(1, 2)$, and 6 to $(2, 3) = (0, 0)$. So the first time we get to the identity $(0, 0)$ is with 6 copies of $(1, 1)$.

(d) $2^6 = 64 \equiv 1 \pmod 7$, but this does not mean that 2 has order 6 in $\mathbb{Z}_7^*$. In fact its order is 3 because $2^3 = 8 \equiv 1 \pmod 7$ and neither 2^2 nor 2^1 is equivalent to $1 \pmod 7$.

Theorem 4.5.3. Let G be a group, $a \in G$ and $m \in \mathbb{N}$. Then $a^m = 1$ if and only if m is a multiple of the order of a.

Proof Let the order of a be n. First suppose $a^m = 1$. Use the division algorithm to write $m = qn + r$ with $0 \leq r < n$. Then

$$1 = a^m = a^{qn+r} = (a^n)^q \cdot a^r = 1^q a^r = a^r.$$

Since $r < n$, the definition of order implies that $r = 0$. Thus n divides m. Going in the other direction, we have that if $m = qn$, then $a^m = (a^n)^q = 1^q = 1$. □

Theorem 4.5.4. Let G be a group and let $a \in G$ have finite order n in G.

(a) $a^r = a^s$ if and only if $r \equiv s \pmod n$.

(b) $1, a, a^2, \ldots, a^{n-1}$ are pairwise distinct elements of G.

Proof

(a) Without loss of generality assume $r \geq s$. We have $a^r = a^s$ if and only if $a^r a^{-s} = 1$ if and only if $a^{r-s} = 1$. Thus, we have $a^r = a^s$ if and only if $n \mid r - s$ by Theorem 4.5.3. That is, $a^r = a^s$ if and only if $r \equiv s \pmod{n}$.

(b) This follows immediately from (a) because $n \nmid r - s$ for any distinct $r, s \in \{0, 1, \ldots, n-1\}$.

□

4.6 Group Generators

Definition 4.6.1. Let G be a group and S a non-empty subset of G. A **word** formed from S is either the identity element, or a finite product $s_1 \cdots s_m$ where for $1 \leq i \leq m$ at least one of s_i or $s_i^{-1} \in S$.

A product of length zero is usually defined to be 1. So a word formed from S is simply a finite product of zero or more terms s_i such that s_i or $s_i^{-1} \in S$. If we collect together any repeated adjacent letters (replace $s \cdot s$ by s^2 etc), then a word formed from S is a product

$$\prod_{i=1}^{m} s_i^{n_i} \qquad s_i \in S, \quad n_i \in \mathbb{Z}, \quad m \geq 0.$$

Example 4.6.2. If $S = \{a, b, c\}$, then words formed from S include 1, b^{-1}, a^2, $c^3a^{-2}ba$, $a^2b^{-3}a^2c^5a^{-7}$ and so on. (We cannot simplify these words without further knowledge of G.)

Definition 4.6.3. Let G be a group, and S a non-empty subset of G. The **group generated by S**, denoted $\langle S \rangle$ is the subset of G containing all words formed from S. If $S = \{a_1, \ldots, a_n\}$, then we may write $\langle a_1, \ldots, a_n \rangle$ instead of $\langle \{a_1, \ldots, a_n\} \rangle$.

The next theorem shows that $\langle S \rangle$ is indeed a group.

Theorem 4.6.4. Let S be a non-empty subset of a group G. Then $\langle S \rangle$ is a subgroup of G, containing the set S.

Proof The set $\langle S \rangle$ is non-empty and indeed contains 1, by definition. If w and v are words formed from S, so is wv. Finally if z is a word formed from S then $z = s_1 \cdots s_m$ with each s_i or $s_i^{-1} \in S$. So $z^{-1} = s_m^{-1} \cdots s_1^{-1} \in \langle S \rangle$ also, and $\langle S \rangle$ is a subgroup of G. □

Definition 4.6.5. Let G be a group and let S be a subset of G. If $\langle S \rangle = G$ then we say that G is **generated by** S.

Example 4.6.6.

(a) The group $(\mathbb{Z}, +)$ is generated by $\{1\}$.

(b) The group $(\mathbb{Z}_n, +)$ is generated by $[1]_n$.

(c) Let $G = S_3$ and $S = \{\sigma, \tau\}$. Then $\langle S \rangle$ contains 1 and also all products involving powers (positive, negative and zero) of σ and τ, such as σ, $\sigma\tau$, $\sigma^2\tau$ etc. We saw in Example 4.2.3 that all elements of S_3 are obtained in this way (in fact $S_3 = \{1, \sigma, \sigma^2, \tau, \sigma\tau, \sigma^2\tau\}$) Thus S_3 is generated by S and $\langle \sigma, \tau \rangle = S_3$.

We have shown that $\langle S \rangle$ is a subgroup of G, containing the set S. In fact, we shall see that it is the smallest subgroup of G containing S. In other words, $\langle S \rangle$ adds to S the fewest possible extra elements that result in a group.

To make this precise and talk about the smallest group containing some set we need the following.

Theorem 4.6.7. Let G be a group. Any intersection of subgroups of G is a group.

Proof Let $\{H_i\}$ be a collection of subgroups of G, where i runs over some indexing set I. Let

$$H = \bigcap_{i \in I} H_i.$$

All of the H_i contain 1, so H contains 1, and hence is non-empty. If $a, b \in H$ then a, b are in each H_i which means that ab^{-1} is in each H_i. Thus ab^{-1} is in H and H is a subgroup of G. □

Theorem 4.6.8. Let G be a group and let S be a non-empty subset of G. Then $\langle S \rangle$ is equal to the intersection of all subgroups of G that contain S.

Proof There is at least one subgroup of G containing S, namely G itself. Thus we may define K to be the intersection of all subgroups of G that contain S. The intersection of subgroups of G is a group, and so K is a subgroup of G. We show $K = \langle S \rangle$.

We know that $\langle S \rangle$ is a subgroup of G containing S. Thus $\langle S \rangle$ is one of the groups being intersected in the definition of K. Hence $K \subseteq \langle S \rangle$.

Now let w be any element of $\langle S \rangle$. Thus w is a word formed from S. Let H be any subgroup of G containing S. Because H is a group and contains S, it must contain w, so $w \in H$. This holds for any such H. Hence w is in the intersection of all such H. That is, $w \in K$. But $w \in \langle S \rangle$ was arbitrary. Hence $\langle S \rangle \subseteq K$ and we have $K = \langle S \rangle$ (because we have already shown that $K \subseteq \langle S \rangle$). □

4.7 Cyclic Groups

Consider a group generated by a single element:

$$\langle a \rangle = \{a^n \mid n \in \mathbb{Z}\}$$

or when using additive notation, $\langle a \rangle = \{na : n \in \mathbb{Z}\}$.

Definition 4.7.1. If $G = \langle a \rangle$ then we say that G is **cyclic,** generated by a.

Thus G is cyclic exactly if every element of G can be written as a power (repeated sum in the additive case) of some fixed generating element a.

Example 4.7.2.

(a) $\mathbb{Z}$ is cyclic, generated by 1.

(b) $\mathbb{Z}_n$ is cyclic, generated by $[1]_n$.

(c) $\mathbb{Z}_{10}^*$ is cyclic. $\mathbb{Z}_{10}^* = \{1, 3, 7, 9\} = \langle 3 \rangle$.

(d) $\mathbb{Z}_8^*$ is not cyclic. $\mathbb{Z}_8^* = \{1, 3, 5, 7\}$ and $\langle 1 \rangle = \{1\}$, $\langle 3 \rangle = \{1, 3\}$, $\langle 5 \rangle = \{1, 5\}$, $\langle 7 \rangle = \{1, 7\}$.

(e) Dihedral groups are not cyclic. In D_n, rotation by $\frac{2\pi}{n}$ is denoted by σ and $\langle \sigma \rangle$ is a subgroup of order n (recall that $|D_n| = 2n$). Any element of this subgroup cannot generate D_n, and any reflection generates a group of order 2.

Theorem 4.7.3. If $a \in G$ has order n, then $\langle a \rangle = \{1, a, a^2, \ldots, a^{n-1}\}$.

Proof By definition, $\langle a \rangle$ consists of all words of the form

$$\prod_{i=1}^{m} a^{n_i} \qquad n_i \in \mathbb{Z},\ m \geq 0.$$

Since we have $a^x a^y = a^{x+y}$ for all $x, y \in \mathbb{Z}$, we thus have $\langle a \rangle = \{a^i : i \in \mathbb{Z}\}$. Now, for any $i \in \mathbb{Z}$ we have $i = qn+r$ where $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, n-1\}$. Thus, $a^i = a^{qn+r} = (a^n)^q \cdot a^r = 1^q \cdot a^r = 1 \cdot a^r = a^r$ and so we have $\langle a \rangle = \{1, a, a^2, \ldots, a^{n-1}\}$. □

Theorem 4.7.3 explains the two uses of the word order: *the order of an element is the order of the cyclic subgroup it generates.*

Theorem 4.7.4. Let $G = \langle a \rangle$ be a cyclic group of order n and let $x \in \mathbb{N}$. Then a^x has order n if and only if $\gcd(x, n) = 1$.

Proof First suppose a^x has order n and let d be a positive common divisor of x and n. Then we have $(a^x)^{\frac{n}{d}} = (a^n)^{\frac{x}{d}} = 1^{\frac{x}{d}} = 1$. Thus, by Theorem 4.5.3 we have that $\frac{n}{d}$ is a multiple of n, which implies that $d = 1$, and hence that $\gcd(x, n) = 1$.

Now suppose that $\gcd(x, n) = 1$ and let m be the order of a^x. Then $a^{xm} = (a^x)^m = 1$ and so $n \mid xm$ by Theorem 4.5.3. Since $\gcd(x, n) = 1$, this implies $n \mid m$. Thus, $m = n$. □

Theorem 4.7.5. Every cyclic group is abelian.

Proof Let $\langle a \rangle$ be a cyclic group. If we pick two arbitrary elements a^n and a^m from this group then $a^n \cdot a^m = a^{n+m} = a^{m+n} = a^m \cdot a^n$. □

Theorem 4.7.6. A subgroup of a cyclic group is cyclic.

Proof Let G be a cyclic group and let $H \leq G$. Since G is cyclic, there exists $a \in G$ such that $\langle a \rangle = G$. If $H = \{1\}$, then $H = \langle 1 \rangle$ is cyclic, so we can assume $H \neq \{1\}$. Thus, a^x and its inverse a^{-x} are in H for some $x \in \mathbb{Z} \setminus \{0\}$. Let m be the smallest positive integer such that $a^m \in H$. We will show that $H = \langle a^m \rangle$ which shows that H is cyclic. Since $a^m \in H$ we have $\langle a_m \rangle \subseteq H$. It remains to show that $H \subseteq \langle a_m \rangle$.

Consider an arbitrary element $b \in H$ with $b \neq 1$. We have $b = a^n$ for some integer n. By Theorem 1.3.1, there exist integers q and r with $0 \leq r < m$ such that $n = qm + r$. So we have $a^n = a^{qm+r}$ from which it follows that $a^r = (a^n)(a^m)^{-q}$. Since $a^n \in H$ and $a^m \in H$ this implies that $a^r \in H$. But $0 \leq r < m$ and so by the definition of m as the smallest positive integer such that $a^m \in H$, we have $r = 0$. Thus, $n = qm$ and $a^n = (a^m)^q$. This means that $a^n \in \langle a^m \rangle$. Thus, $b \in H$ and we have proved $H \subseteq \langle a_m \rangle$. □

Theorem 4.7.7. A cyclic group of order n has a unique subgroup of order d for each positive divisor d of n.

Proof Let G be a cyclic group of order n and let d be a positive divisor of n. By Theorem 4.7.3 we can write $G = \langle a \rangle = \{1, a, a^2, \ldots, a^{n-1}\}$. Consider the subgroup H of G generated by $a^{n/d}$. It can be seen that $H = \{1, a^{n/d}, a^{2n/d}, \ldots, a^{(d-1)n/d}\}$ has order d. It remains to show that H is the only subgroup of order d.

Suppose that H' is another subgroup of order d. By Theorem 4.7.6, H' is cyclic and so contains an element a^x of order d. Thus we have $(a^x)^d = a^{xd} = 1$ and so by Theorem 4.5.3 we have $xd = qn$ for some integer q. Since $d \mid n$, this means that $\frac{n}{d} \mid x$. Thus, $a^x \in H$. Since a^x generates H', this means that every element of H' is in H and $H' = H$. That is, H is the only subgroup of order d. □

4.8 Group Homomorphisms

Consider the following well-known identity from linear algebra. The det function $\det : GL_n(\mathbb{R}) \to \mathbb{R} \setminus \{0\}$ satisfies

$$\det(AB) = \det(A)\det(B).$$

Here the product AB on the left is the product of two matrices, while the product $\det(A)\det(B)$ on the right is the product of two real numbers. It makes no difference whether we first multiply in $GL_n(\mathbb{R})$ and then apply det, or whether we first apply det and then multiply in $\mathbb{R}$.

Abstracting, we have the following definition:

Definition 4.8.1. Let $(G, *)$ and $(H, \odot)$ be two groups. A **(group) homomorphism** from G to H is a function $f : G \to H$ satisfying

$$f(x * y) = f(x) \odot f(y) \qquad \text{for all } x, y \in G.$$

Here x and y are in G, so we can form $x * y$, which is again in G. Then we can apply f to obtain an element in H. Or, we can first apply f to x and to y. Now we have two elements $f(x)$, $f(y) \in H$,

so we can form $f(x) \odot f(y)$. The defining property of a homomorphism is that we must get the same result either way. We say that f *respects* the group operation.

$$\begin{array}{ccccc} G & & (x,y) & \xrightarrow{*} & x*y \\ f\downarrow & & f\downarrow & & f\downarrow \\ H & & (f(x),f(y)) & \xrightarrow{\odot} & f(x)\odot f(y)=f(x*y) \end{array}$$

Example 4.8.2.

(a) Consider the groups $GL_n(\mathbb{R})$ and $\mathbb{R}\setminus\{0\}$ under multiplication. The determinant function $\det : GL_n(\mathbb{R}) \to \mathbb{R}\setminus\{0\}$ is a homomorphism.

(b) Consider the groups $(\mathbb{R},+)$ and $(\mathbb{R}\setminus\{0\},\cdot)$. Let $f:\mathbb{R}\to\mathbb{R}\setminus\{0\}$ be the exponential function. Then
$$f(x+y) = e^{x+y} = e^x \cdot e^y = f(x)\cdot f(y)$$
so f is a homomorphism.

(c) Recall the reduction modulo n map $\pi : \mathbb{Z}\to\mathbb{Z}_n$ given by $\pi(x) = [x]_n$ for all $x\in\mathbb{Z}$. Then $\pi(a+b) = [a+b]_n = [a]_n \oplus [b]_n = \pi(a)\oplus\pi(b)$. So π is a homomorphism from the group $(\mathbb{Z},+)$ to $(\mathbb{Z}_n,\oplus)$.

Group homomorphisms are required to respect the group operation, but in fact they also respect the identity element and inverses:

Theorem 4.8.3. Let $(G,*)$ and $(H,\odot)$ be groups, and let $f:G\to H$ be a homomorphism. Then:

(a) $f(1_G) = 1_H$ (homomorphisms preserve the identity)

(b) $f(x^{-1}) = f(x)^{-1}$ (homomorphisms preserve inverses).

Proof

(a) $f(1_G) = f(1_G * 1_G) = f(1_G)\odot f(1_G)$, and the result follows by cancelling.

(b) By (a), $1_H = f(1_G) = f(x*x^{-1}) = f(x)\odot f(x^{-1})$ and similarly $f(x^{-1})\odot f(x) = 1_H$ so $f(x^{-1})$ is the inverse of $f(x)$.

□

Theorem 4.8.4. Let $(G,*)$, $(H,\odot)$ and $(K,\boxtimes)$ be groups. If $f:G\to H$ and $g:H\to K$ are homomorphisms, then $g\circ f : G\to K$ is a homomorphism.

Proof Let $x, y \in G$ and let $h = g \circ f$. Then $h(x)$, $h(y) \in K$, and

$$\begin{aligned} h(x * y) &= (g \circ f)(x * y) \\ &= g\Big(f(x * y)\Big) \\ &= g\Big(f(x) \odot f(y)\Big) && \text{f is a homomorphism} \\ &= g\Big(f(x)\Big) \boxtimes g\Big(f(y)\Big) && g \text{ is a homomorphism} \\ &= (g \circ f)(x) \boxtimes (g \circ f)(y) \\ &= h(x) \boxtimes h(y). \end{aligned}$$

□

Definition 4.8.5. Let $f : G \to H$ be a homomorphism. The **kernel** of f is the set

$$\ker f = \{g \in G : f(g) = 1_H\}.$$

The **image** of f is the set

$$\operatorname{Im} f = \{f(g) : g \in G\}.$$

Example 4.8.6. Consider the reduction mod n homomorphism $\pi : \mathbb{Z} \to \mathbb{Z}_n$ given by $\pi(x) = [x]_n$ for all $x \in \mathbb{Z}$. The kernel of π is the set of all integers a with $[a] = 0$, that is, all a with $a \equiv 0 \pmod{n}$. Thus $\ker \pi = \{nx \mid x \in \mathbb{Z}\} = n\mathbb{Z}$.

Theorem 4.8.7. Let $f : G \to H$ be a homomorphism.

(a) $\ker f$ is a subgroup of G.

(b) $\operatorname{Im} f$ is a subgroup of H.

Proof

(a) $1_G \in \ker f$ by Theorem 4.8.3, so $\ker f$ is non-empty. If $x, y \in \ker f$ then $f(x) = f(y) = 1_H$. So

$$\begin{aligned} f(xy^{-1}) &= f(x)f(y^{-1}) && f \text{ is a homomorphism} \\ &= f(x)f(y)^{-1} && \text{by Theorem 4.8.3} \\ &= 1_H\, 1_H^{-1} && \text{Since } x, y \in \ker f \\ &= 1_H. \end{aligned}$$

So $\ker f$ is a subgroup of G, by Theorem 4.3.3.

(b) We have $f(1_G) \in \operatorname{Im} f$ so $\operatorname{Im} f$ is a non-empty subset of H. Also, if $h_1, h_2 \in \operatorname{Im} f$, then there exist $g_1, g_2 \in G$ such that $f(g_1) = h_1$ and $f(g_2) = h_2$. Thus, $h_1 h_2^{-1} = f(g_1)f(g_2)^{-1} = f(g_1)f(g_2^{-1}) = f(g_1 g_2^{-1}) \in \operatorname{Im} f$, and so $\operatorname{Im} f \leq H$ by Theorem 4.3.3.

□

If $f : G \to H$ is an injective function (not necessarily a homomorphism) then the number of elements of G that map to 1_H is at most one. If f is a homomorphism, this condition turns out to be sufficient to ensure f is injective. That is, we can check if a homomorphism is injective by looking at just one value in the codomain.

Theorem 4.8.8. If $f : G \to H$ is a homomorphism, then $\ker f = \{1_G\}$ if and only if f is injective.

Proof Suppose f is injective. We know $1_G \in \ker f$. Suppose $x \in \ker f$. Then $f(x) = 1_H = f(1_G)$ by Theorem 4.8.3. Since f is injective $x = 1_G$. Thus $\ker f = \{1_G\}$.

Now suppose $\ker f = \{1_G\}$. If $f(x) = f(y)$ then

$$1_H = f(y)f(y)^{-1} = f(x)f(y)^{-1} = f(x)f(y^{-1}) = f(xy^{-1}),$$

so $xy^{-1} \in \ker f$. Thus $xy^{-1} = 1_G$, so $x = y$. Hence f is injective. □

4.9 Group Isomorphisms

Suppose that there exists a bijective homomorphism f from $(G, *)$ to $(H, \odot)$. Then G and H are essentially the same group, their elements have just been named differently. The element called x in G is called $f(x)$ in H. If we have $x * y = z$ in G, then we have $f(x) \odot f(y) = f(z)$ in H (because $f(x) \odot f(y) = f(x * y) = f(z)$).

Definition 4.9.1. An **isomorphism** is a bijective group homomorphism. If there exists an isomorphism $f : G \to H$ we say that G is **isomorphic** to H, and write $G \simeq H$.

If $G \simeq H$ then any theorem we can prove about G is true for H, because we can just relabel all the elements, using our isomorphism. This perfectly matches elements between G and H, and products in G are matched with products in H. Similarly 1_G matches with 1_H and inverses match also (Theorem 4.8.3). So the entire proof in G translates to a proof in H.

Example 4.9.2.

(a) Let $2\mathbb{Z} = \{2x : x \in \mathbb{Z}\}$. It is easy to check that $2\mathbb{Z}$ is a group under addition. We have $2\mathbb{Z} \subseteq \mathbb{Z}$ and for all $x, y \in 2\mathbb{Z}$ we have $x - y \in 2\mathbb{Z}$ so $(2\mathbb{Z}, +)$ is a group by Theorem 4.3.3. We now show that $Z \simeq 2\mathbb{Z}$. Let $f : \mathbb{Z} \to 2\mathbb{Z}$ be defined by $f(x) = 2x$ for all $x \in \mathbb{Z}$. It is easy to check that f is a bijection and for all $x, y \in \mathbb{Z}$ we have

$$f(x + y) = 2(x + y) = 2x + 2y = f(x) + f(y)$$

and so f is an isomorphism. Thus, $Z \simeq 2\mathbb{Z}$.

(b) If G is a group and $a \in G$ has order n, then $\langle a \rangle \simeq \mathbb{Z}_n$.

Define $f : \langle a \rangle \to \mathbb{Z}_n$ by $f(a^i) = [i]_n$ for all $i \in \mathbb{Z}$. Then

$$f(a^i a^j) = f(a^{i+j}) = [i + j]_n = [i]_n + [j]_n = f(a^i) + f(a^j)$$

so f is a homomorphism from $\langle a \rangle$ to $\mathbb{Z}_n$. To see that f is a bijection, recall that $\langle a \rangle = \{a^0, a^1, \ldots, a^{n-1}\}$ (see Theorem 4.7.3).

(c) $\mathbb{Z}_6 \simeq \mathbb{Z}_2 \times \mathbb{Z}_3$.

Because we are working with several different groups with operations that we usually denote by the same symbol, namely $(\mathbb{Z}_6, \oplus)$, $(\mathbb{Z}_2, \oplus)$, $(\mathbb{Z}_3, \oplus)$ and $(\mathbb{Z}_2 \times \mathbb{Z}_3, \oplus)$, we will denote $\oplus$ for $\mathbb{Z}_n$ by $\oplus_n$ and $\oplus$ for $\mathbb{Z}_m \times \mathbb{Z}_n$ by $\oplus_{m\times n}$. Thus, our groups are $(\mathbb{Z}_6, \oplus_6)$, $(\mathbb{Z}_2, \oplus_2)$, $(\mathbb{Z}_3, \oplus_3)$ and $(\mathbb{Z}_2 \times \mathbb{Z}_3, \oplus_{2\times 3})$.

We will show that

$$f([x]_6) = ([x]_2, [x]_3)$$

is an isomorphism from $\mathbb{Z}_6$ to $\mathbb{Z}_2 \times \mathbb{Z}_3$. We have

$$\begin{array}{lll} f([0]_6) = ([0]_2, [0]_3) & f([1]_6) = ([1]_2, [1]_3) & f([2]_6) = ([0]_2, [2]_3) \\ f([3]_6) = ([1]_2, [0]_3) & f([4]_6) = ([0]_2, [1]_3) & f([5]_6) = ([1]_2, [2]_3) \end{array}$$

and so f is a bijection. We now show that f is a homomorphism.

$$\begin{aligned} f\Big([n]_6 \oplus_6 [m]_6\Big) &= f\Big([n+m]_6\Big) \\ &= \Big([n+m]_2,\ \ [n+m]_3\Big) \\ &= \Big([n]_2 \oplus_2 [m]_2,\quad [n]_3 \oplus_3 [m]_3\Big) \\ &= \Big([n]_2,\ [n]_3\Big) \oplus_{2\times 3} \Big([m]_2,\ [m]_3\Big) \\ &= f\Big([n]_6\Big) \oplus_{2\times 3} f\Big([m]_6\Big). \end{aligned}$$

Thus, f is a homomorphism and hence (because it is a bijection) also an isomorphism.

The only thing special about 2 and 3 in Example 4.9.2 is that $\gcd(2,3) = 1$. Otherwise, the function f is not a bijection.

Theorem 4.9.3. Let m, $n \in \mathbb{N}$ with $\gcd(m,n) = 1$. Then

$$\mathbb{Z}_{mn} \simeq \mathbb{Z}_m \times \mathbb{Z}_n.$$

Proof Define a map $f : \mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n$ by $f([a]_{mn}) = ([a]_m, [a]_n)$. We need to check that f is well defined. If $[a]_{mn} = [b]_{mn}$ then $mn \mid (a-b)$ so $a \equiv b \pmod m$ and $a \equiv b \pmod n$. Hence $([a]_m, [a]_n) = ([b]_m, [b]_n)$. Thus f is well defined.

We now show that f is a homomorphism. As in Example 4.9.2, for all $m, n \in \mathbb{N}$ we denote the binary operation of $\mathbb{Z}_n$ by $\oplus_n$ and the binary operation of $\mathbb{Z}_m \times \mathbb{Z}_n$ by $\oplus_{m\times n}$. We have $f([a]_{mn} \oplus_{mn} [b]_{mn}) = f([a+b]_{mn}) = ([a+b]_m, [a+b]_n) = ([a]_m \oplus_m [b]_m, [a]_n \oplus_n [b]_n) = ([a]_m, [a]_n) \oplus_{m\times n} ([b]_m, [b]_n) = f([a]_{mn}) \oplus_{m\times n} f([b]_{mn})$ so f is a homomorphism.

It remains to show that f is a bijection. To do this, it is enough to show that f is injective because $|\mathbb{Z}_{mn}| = |\mathbb{Z}_m \times \mathbb{Z}_n| = mn$. To show f is injective we use Theorem 4.8.8. If $f([a]_{mn}) = ([0]_m, [0]_n)$, then $[a]_m = [0]_m$ so $m \mid a$. Similarly $n \mid a$. Since $\gcd(m,n) = 1$, $m \mid a$ and $n \mid a$ implies $mn \mid a$ (see Theorem 1.6.5). So $[a]_{mn} = [0]_{mn}$. Hence the kernel of f is $\{[0]_{mn}\}$, and so f is injective by Theorem 4.8.8. □

Theorem 4.9.3 generalises as follows.

Theorem 4.9.4. Let $m_1, \ldots, m_k \in \mathbb{N}$ with all the m_i pairwise relatively prime. Then

$$\mathbb{Z}_{m_1 \cdots m_k} \simeq \mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_k}.$$

Proof The proof is similar to that of Theorem 4.9.3 and is left as an exercise. □

Theorem 4.9.4 is an algebraic formulation of the Chinese Remainder Theorem (Theorem 3.5.2). Actually the result is still true if the $\mathbb{Z}_{m_1 \cdots m_k}$ and $\mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_k}$ are viewed as rings (see later) and the isomorphism as a ring isomorphism. It is really this ring result that is an algebraic formulation of the Chinese Remainder Theorem.

It should not be surprising that isomorphism is an equivalence relation (on any given set of groups), as the following theorem shows.

Theorem 4.9.5. Let G, H, K be groups.

(a) $G \simeq G$.

(b) If $G \simeq H$, then $H \simeq G$.

(c) If $G \simeq H$ and $H \simeq K$, then $G \simeq K$.

Proof

(a) The identity function $f : G \to G$ defined by $f(x) = x$ for all x is clearly a bijection. It is also a homomorphism, since $f(xy) = xy = f(x)f(y)$.

(b) Let $f : G \to H$ be an isomorphism. We need an isomorphism $H \to G$. Since f is a bijection, $f^{-1} : H \to G$ is also a bijection. We check that f^{-1} is also a homomorphism. Let h_1, $h_2 \in H$. Since f is surjective, $h_1 = f(g_1)$ and $h_2 = f(g_2)$ for some g_1, $g_2 \in G$. We can also write this as $g_1 = f^{-1}(h_1)$ and $g_2 = f^{-1}(h_2)$. Thus

$$\begin{aligned} f^{-1}(h_1h_2) &= f^{-1}\Big(f(g_1)f(g_2)\Big) \\ &= f^{-1}\Big(f(g_1g_2)\Big) \qquad \text{f is a homomorphism} \\ &= g_1g_2 \\ &= f^{-1}(h_1)\ f^{-1}(h_2). \end{aligned}$$

Hence $f^{-1} : H \to G$ is an isomorphism.

(c) Let $f : G \to H$ and $g : H \to K$ be isomorphisms. Then $g \circ f : G \to K$ is bijective (the composition of bijections is a bijection), and is a homomorphism by Theorem 4.8.4. Hence $g \circ f$ is an isomorphism from G to K and we have $G \simeq K$.

□

Sometimes there is a very easy way to tell that two groups are not isomorphic. If $|G| \neq |H|$, then $G \not\simeq H$. However, two groups with the same cardinality need not be isomorphic, and in this case it may be very difficult to prove that there is no function from G to H that is a group isomorphism. However, it should be intuitive that if $G \simeq H$, then G and H have the same structural properties, and so different structural properties can be used to show groups are not isomorphic.

Examples of structural properties of a group G are

- G is abelian;
- G has an element of order n;
- G has exactly t elements of order n;

and so on. Properties of G that depend on the names of elements are not structural properties and cannot be used to show non-isomorphism. Examples of properties that cannot be used to prove non-isomorphism include

- G contains the number 2;
- The elements of G are permutations;
- G is a subgroup of $\mathbb{Z}$;

and so on.

Example 4.9.6.

(a) Let $2\mathbb{Z} = \{2x : x \in \mathbb{Z}\}$. We have $5 \in \mathbb{Z}$ and $5 \notin 2\mathbb{Z}$, but this certainly does not mean $\mathbb{Z} \not\simeq 2\mathbb{Z}$. In fact, we saw in Example 4.9.2 that $\mathbb{Z} \simeq 2\mathbb{Z}$.

(b) If $G \simeq H$ and G is abelian, then H is abelian. To see this let f be an isomorphism from G to H. Then for all $x, y \in H$ we have $x = f(a)$ and $y = f(b)$ for some $a, b \in G$ (in fact $a = f^{-1}(x)$ and $b = f^{-1}(y)$). Thus,

$$xy = f(a)f(b) = f(ab) = f(ba) = f(b)f(a) = yx$$

and so H is abelian.

(c) If $G \simeq H$ and G has an element of order n, then so does H. To see this let f be an isomorphism from G to H and let a be an element of order n in G. We will show that $f(a)$ has order n in H. There exist distinct elements $a^0, a^1, \ldots, a^{n-1}$ in G and corresponding distinct elements $f(a^0), f(a^1), \ldots, f(a^{n-1})$ in H. Note in particular that, $f(a^0) = 1$ and that none of $f(a^1), \ldots, f(a^{n-1})$ is 1.

We have $f(a^2) = f(aa) = f(a)f(a) = (f(a))^2$. By repeating the same argument we have $f(a^i) = (f(a))^i$ for $i = 1, 2, \ldots, n$. Thus, in particular, we have shown that $(f(a))^n = f(a^n) = f(1) = 1$ and that $(f(a))^i \neq 1$ for $i = 1, 2, \ldots, n-1$. That is, we have shown that $f(a)$ has order n.

(d) $\mathbb{Z}_2 \times \mathbb{Z}_4 \not\simeq \mathbb{Z}_8$. The element 1 has order 8 in $\mathbb{Z}_8$ ($\mathbb{Z}_8$ is cyclic). However, no element of $\mathbb{Z}_2 \times \mathbb{Z}_4$ has order 8. The orders of the elements in $\mathbb{Z}_2 \times \mathbb{Z}_4$ are as follows: $(0,0)$ has order 1, $(0,2), (1,0), (1,2)$ have order 2, and $(0,1), (0,3), (1,1), (1,3)$ have order 4. Thus, $\mathbb{Z}_2 \times \mathbb{Z}_4 \not\simeq \mathbb{Z}_8$.

4.10 Cosets and Lagrange's Theorem

Definition 4.10.1. Let G be a group, H a subgroup of G and $a \in G$. The **left coset** aH is the subset of G given by

$$aH = \{ah : h \in H\}.$$

The collection of all left cosets of H is denoted G/H.

The **right coset** Ha is the set $Ha = \{ha : h \in H\}$, but we will not be dealing with right cosets. Thus, we usually refer to left cosets simply as **cosets**.

When $a = 1$, we have the coset $1H = \{1h : h \in H\} = H$.

If our group is abelian with additive notation, then we have the following analogous definition of cosets. Left and right cosets are the same in an abelian group because $a + h = h + a$.

Definition 4.10.2. Let G be an abelian group written additively, H be a subgroup of G and $a \in G$. The **coset** $a + H$ is the subset of G given by

$$a + H = \{a + h : h \in H\}.$$

Example 4.10.3. Let $G = \mathbb{Z}$ and $H = 3\mathbb{Z} = \{3x : x \in \mathbb{Z}\}$. Then H is a subgroup of G and the cosets of H are

$$\begin{aligned} 0 + H = H &= \{\ldots, -6, -3, 0, 3, 6, \ldots\} \\ 1 + H &= \{\ldots, -5, -2, 1, 4, 7, \ldots\} \\ 2 + H &= \{\ldots, -4, -1, 2, 5, 8, \ldots\} \end{aligned}$$

So we see that for $i = 0, 1, 2$, the coset $i + H$ is the congruence class of integers congruent to i modulo 3. That is, $\mathbb{Z}/3\mathbb{Z} = \mathbb{Z}_3$. This is an equality of sets, but we will soon see that this extends to a group equality. There is nothing special about 3, in general we have $\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n$.

Just like there can be more than one representative for a congruence class (or an equivalence class), cosets can also be represented in more than one way. For example, when $G = \mathbb{Z}$ and $H = 3\mathbb{Z}$ we have $1 + H = 4 + H$.

Theorem 4.10.4. Let G be a group, let H be a subgroup of G, and let $a, b, c \in G$.

(a) $aH = bH$ if and only if $a^{-1}b \in H$.

(b) $aH = bH$ if and only if $b \in aH$.

(c) $aH = bH$ if and only if $caH = cbH$.

(d) $a \in bH$ if and only if $ca \in cbH$.

(e) G/H is a partition of G.

(f) If G is finite then any two left cosets of H have the same number of elements, equal to the number of elements in H.

Proof

(a) If $aH = bH$ then $b \in bH = aH$ so $b = ah$ for some $h \in H$, so $a^{-1}b = h \in H$.

Going in the other direction, if $a^{-1}b = h \in H$, then

$$b = ah \qquad \text{and} \qquad a = bh^{-1}.$$

Consider any element $bh_1 \in bH$. Then $bh_1 = a(hh_1) \in aH$. Thus $bH \subseteq aH$. And if $ah_2 \in aH$, then $ah_2 = b(h^{-1}h_2) \in bH$ so $aH \subseteq bH$. Thus, $aH = bH$.

(b) We have $b \in aH$ if and only if $b = ah$ for some $h \in H$ if and only if $a^{-1}b = h \in H$. Thus, by (a) we have $aH = bH$ if and only if $b \in aH$.

(c) By (a) we have $caH = cbH$ if and only if $(cb)^{-1}ca \in H$. But $(cb)^{-1}ca = b^{-1}c^{-1}ca = b^{-1}a$, and so we have $caH = cbH$ if and only if $b^{-1}a \in H$. Thus, by (a) we have $caH = cbH$ if and only if $aH = bH$.

(d) By (b) we have $a \in bH$ if and only if $aH = bH$, by (c) we have $aH = bH$ if and only if $caH = cbH$, and by (b) we have $caH = cbH$ if and only if $ca \in cbH$. Thus, we have $a \in bH$ if and only if $ca \in cbH$.

(e) We first show that if $aH \cap bH \neq \emptyset$, then $aH = bH$. Let aH and bH be left cosets and let $x \in aH \cap bH$. Then $x = ah_1 = bh_2$ for some $h_1, h_2 \in H$. But

$$ah_1 = bh_2 \quad \rightarrow \quad a = bh_2h_1^{-1} \quad \rightarrow \quad b^{-1}a = h_2h_1^{-1} \in H$$

and so $aH = bH$ by (a). We also have $a \in aH$ for each $a \in G$ (because $a = a \cdot 1$ and $1 \in H$), which means that every element of G occurs in at least one coset. Thus, G/H is a partition of G.

(f) Fix $a \in G$ and define a function $f : H \rightarrow aH$ by $f(h) = ah$. If $f(h) = f(h')$ then $ah = ah'$ and so we have $h = h'$. Thus f is injective. And f is surjective by the definition of aH. Thus f is a bijection and we have $|H| = |aH|$.

□

Theorem 4.10.5 (Lagrange)**.** Let G be a finite group, and let H be a subgroup of G. Then the order of H divides the order of G.

Proof Theorem 4.10.4 (e) and (f) imply that $|G| = t|H|$ where t is the number of cosets of H. □

Definition 4.10.6. Let G be a group, H a subgroup of G. If the number of left cosets of H in G is a finite number, then we denote this number by $[G : H]$ and say that H has **index** $[G : H]$ in G. Thus, if G is finite, then $|G| = [G : H] \cdot |H|$.

Theorem 4.10.7. If G is a finite group, then the order of any element of G divides the order of G.

Proof If $a \in G$, then the order of a is the order of the subgroup $\langle a \rangle$. By Lagrange's Theorem (Theorem 4.10.5) this divides the order of G. □

Theorem 4.10.8. If G is a group of prime order p, then $G \simeq \mathbb{Z}_p$.

Proof If G is a group of prime order, then G contains an element $a \neq 1$. By Theorem 4.10.7, the order of a divides p, and hence is equal to p. Thus, $\langle a \rangle = G$ and so G is cyclic, which means that $G \simeq \mathbb{Z}_p$ (see Example 4.9.2 (b)). □

4.11 Normal Subgroups and Quotient Groups

It is natural to ask whether the set G/H of all left cosets of G forms a group. The natural way to define a group operation on left cosets aH and bH is

$$(aH)(bH) = (ab)H.$$

This is analogous to our definition of the group operation in $\mathbb{Z}_n \simeq \mathbb{Z}/n\mathbb{Z}$ where we define

$$[a]_n + [b]_n = [a+b]_n$$

or equivalently

$$(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a+b)n\mathbb{Z}.$$

However, the definition $(aH)(bH) = (ab)H$ is not always well defined, as the following example shows.

Example 4.11.1. Let $G = S_3 = D_3 = \{1, \sigma, \sigma^2, \tau, \sigma\tau, \sigma^2\tau\}$ and $H = \langle \tau \rangle = \{1, \tau\}$. Recall that $S_3 = D_3$ (see Example 4.4.4) and recall from Example 4.2.3 that the Cayley Table of S_3 is

S_3	1	σ	σ^2	τ	$\sigma\tau$	$\sigma^2\tau$
1	1	σ	σ^2	τ	$\sigma\tau$	$\sigma^2\tau$
σ	σ	σ^2	1	$\sigma\tau$	$\sigma^2\tau$	τ
σ^2	σ^2	1	σ	$\sigma^2\tau$	τ	$\sigma\tau$
τ	τ	$\sigma^2\tau$	$\sigma\tau$	1	σ^2	σ
$\sigma\tau$	$\sigma\tau$	τ	$\sigma^2\tau$	σ	1	σ^2
$\sigma^2\tau$	$\sigma^2\tau$	$\sigma\tau$	τ	σ^2	σ	1

The cosets of H in G are

$$H = \{1, \tau\} \qquad \sigma H = \{\sigma, \sigma\tau\} \qquad \sigma^2 H = \{\sigma^2, \sigma^2\tau\}.$$

Now, if we "define" an operation on G/H by $(aH)(bH) = (ab)H$, then we have

$$(\sigma H)(\sigma H) = \sigma^2 H.$$

However, since $\sigma H = \sigma\tau H$ and $\sigma\tau\sigma = \tau$, we also have

$$(\sigma H)(\sigma H) = (\sigma\tau H)(\sigma H) = \sigma\tau\sigma H = \tau H$$

Thus, our operation is not well defined. We get different answers, namely $\sigma^2 H$ and τH ($\sigma^2 H \neq \tau H$), for $(\sigma H)(\sigma H)$ when we choose different representatives for σH.

In order for our group operation $(aH)(bH) = (ab)H$ to make sense, we will show that H must have the following property.

Definition 4.11.2. A subgroup H of G is said to be **normal** if for every g in G and $h \in H$, we have $g^{-1}hg \in H$. We write

$$H \trianglelefteq G$$

to mean H is a normal subgroup of G.

Theorem 4.11.3. Let $H \trianglelefteq G$. Then G/H is a group under the group operation defined by

$$(aH)(bH) = (ab)H.$$

If $[G : H]$ is finite, then G/H is a group with $[G : H]$ elements.

Proof We first show that the operation $(aH)(bH) = (ab)H$ is well defined. That is, we show that if $a_1H = a_2H$ and $b_1H = b_2H$, then $(a_1H)(b_1H) = (a_2H)(b_2H)$. By definition $(a_1H)(b_1H) = (a_1b_1)H$ and $(a_2H)(b_2H) = (a_2b_2)H$, so what we need to show is that if $a_1H = a_2H$ and $b_1H = b_2H$, then

$$(a_1b_1)H = (a_2b_2)H.$$

Note that by Theorem 4.10.4 (a) we have $(a_1b_1)H = (a_2b_2)H$ if and only if $(a_1b_1)^{-1}(a_2b_2) \in H$.

Suppose $a_1H = a_2H$ and $b_1H = b_2H$. Thus we have $a_1^{-1}a_2 = h_1 \in H$ and $b_1^{-1}b_2 = h_2 \in H$ (see Theorem 4.10.4 (a)). Now,

$$\begin{aligned} (a_1b_1)^{-1}a_2b_2 &= b_1^{-1}a_1^{-1}a_2b_2 \\ &= b_1^{-1}b_2b_2^{-1}a_1^{-1}a_2b_2 \qquad \text{Inserting factor } b_2b_2^{-1} = 1 \\ &= h_2b_2^{-1}h_1b_2. \end{aligned}$$

We have $(b_2^{-1}h_1b_2) = h_3 \in H$ because $H \trianglelefteq G$, and so we have $(a_1b_1)^{-1}a_2b_2 = h_2h_3 \in H$. This proves $(a_1b_1)H = (a_2b_2)H$ and hence that multiplication is well defined in G/H.

It remains to check the group axioms.

Associativity: $(aH)\Big((bH)(cH)\Big) = (aH)(bcH) = a(bc)H = (ab)cH = \Big((aH)(bH)\Big)(cH)$.

Identity: $(aH)(1H) = (a \cdot 1)H = aH$ and $(1H)(aH) = (1 \cdot a)H = aH$. Thus, H is the identity of G/H.

Inverses: $(aH)(a^{-1}H) = (aa^{-1})H = H$ and $(a^{-1}H)(aH) = (a^{-1}a)H = H$. Thus, aH has inverse $a^{-1}H$. □

Definition 4.11.4. If $H \trianglelefteq G$, then G/H is called the **quotient group**, or **factor group**, of G by H.

The factor group G/H is not a subgroup of G - the elements of G/H are cosets, not elements of G. The factor group G/H is a "new" group and need not be isomorphic to any subgroup of G. For example, $\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n$ has n elements, but the only finite subgroup of $\mathbb{Z}$ is the trivial group $\{0\}$.

Theorem 4.11.5. Let G be a group.

(a) $\{1_G\} \trianglelefteq G$ and $G/\{1_G\} \simeq G$.

(b) $G \trianglelefteq G$ and G/G is the trivial group.

(c) If G is abelian and $H \leq G$, then $H \trianglelefteq G$.

Proof

(a) We have $g^{-1}1_G g = g^{-1}g = 1_G \in \{1_G\}$ so $\{1_G\} \trianglelefteq G$. Also, $G/\{1_G\} = \{\{g\} : g \in G\}$, and the function $f : G \to G/\{1_G\}$ given by $f(g) = \{g\}$ is an isomorphism from G to $G/\{1_G\}$.

(b) We have $g^{-1}g'g \in G$ for all $g, g' \in G$ so $G \trianglelefteq G$. Also, $G/G = \{G\}$ and so is the trivial group.

(c) Suppose G is abelian and $H \leq G$. If $g \in G$ and $h \in H$, then we have $g^{-1}hg = hg^{-1}g = h \in H$, so $H \trianglelefteq G$.

□

Theorem 4.11.5 (c) explains why $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ is a group. In $\mathbb{Z}_n$ we denote the coset $a + n\mathbb{Z}$ by $[a]_n$. The addition we defined for $\mathbb{Z}_n$ is exactly the addition of cosets in $\mathbb{Z}/n\mathbb{Z}$. That is, the addition $[a]_n + [b]_n = [a + b]_n$ is exactly the coset addition $(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a + b) + n\mathbb{Z}$

Example 4.11.6.

(a) Consider the subgroup $\langle\tau\rangle = \{1, \tau\}$ of S_3, see Example 4.2.3. Referring to the Cayley table for S_3, we have $\sigma^{-1}\tau\sigma = \sigma^2\tau\sigma = \sigma\tau \notin \langle\tau\rangle$. Thus, $\langle\tau\rangle \not\trianglelefteq S_3$. This explains why the operation in Example 4.11.1 is not well defined.

(b) Consider the subgroup $\langle\sigma\rangle = \{1, \sigma, \sigma^2\}$ of S_3, see Example 4.2.3. We will show that $\langle\sigma\rangle \trianglelefteq S_3$. Let $h \in H = \langle\sigma\rangle$ and $g \in G = S_3$. We need to show that $g^{-1}hg \in H$. If $g \in H$ then we certainly have $g^{-1}hg \in H$ so we can assume $g \in G \setminus H = \{\tau, \sigma\tau, \sigma^2\tau\}$. Thus, $g = \sigma^i\tau$ for some $i \in \{0, 1, 2\}$ and $h = \sigma^j$ for some $j \in \{0, 1, 2\}$, and so we have

$$g^{-1}hg = (\sigma^i\tau)^{-1}\sigma^j(\sigma^i\tau) = \tau\sigma^{-i}\sigma^j\sigma^i\tau = \tau\sigma^j\tau.$$

Referring to the Cayley table for S_3, it can be checked that $\tau\sigma^j = \sigma^{-j}\tau$ for $j = 0, 1, 2$. Thus, we have $g^{-1}hg = \sigma^{-j}\tau\tau = \sigma^{-j} \in H$, and so $\langle\sigma\rangle \trianglelefteq S_3$. It can be easily checked that $G/H \simeq \mathbb{Z}_2$ with isomorphism $H \mapsto [0]_2$, $\tau H \mapsto [1]_2$.

(c) Consider $\mathbb{Z}$ and $\mathbb{Q}$ under addition. $(\mathbb{Q}, +)$ is an abelian group so $\mathbb{Z}$ is a normal subgroup. In $\mathbb{Q}/\mathbb{Z}$, $b + \mathbb{Z} = a + \mathbb{Z}$ if and only if $a - b \in \mathbb{Z}$. So we have one coset $q + \mathbb{Z}$ for each $q \in [0, 1) \cap \mathbb{Q}$.

Now, $1/2 + \mathbb{Z}$ has order 2 in $\mathbb{Q}/\mathbb{Z}$ because $(1/2 + \mathbb{Z}) + (1/2 + \mathbb{Z}) = (1 + \mathbb{Z}) = 0 + \mathbb{Z}$. In general $m/n + \mathbb{Z}$ has order n (provided $\gcd(m, n) = 1$). Thus $\mathbb{Q}/\mathbb{Z}$ is an infinite group, in which every element has finite order.

Theorem 4.11.7 (First Isomorphism Theorem)**.** Let G and H be groups and let $f : G \to H$ be a homomorphism. Then

(a) $\ker f \trianglelefteq G$;

(b) $\operatorname{Im} f \leq H$; and

(c) $G/\ker f \simeq \operatorname{Im} f$.

Proof

(a) Let $g \in G$ and $h \in \ker f$. We need to show $g^{-1}hg \in \ker f$. But $f(g^{-1}hg) = f(g)^{-1}f(h)f(g) = f(g)^{-1}f(g)$ because $f(h) = 1_H$, and $f(g)^{-1}f(g) = 1$, so $g^{-1}hg \in \ker f$ and we are done.

(b) This was proved in Theorem 4.8.7.

(c) Let $K = \ker f$ and define $\theta : G/K \to \operatorname{Im} f$ by

$$\theta(aK) = f(a).$$

We will show that θ is an isomorphism. First we need to show that θ is well defined. That is, we need to show that if $aK = bK$, then $f(a) = f(b)$. Suppose $aK = bK$. Then we have $a^{-1}b \in K$ which means that $f(a^{-1}b) = 1_H$. But $f(a^{-1}b) = f(a^{-1})f(b) = f(a)^{-1}f(b)$ and so we have $f(a)^{-1}f(b) = 1_H$, which means that $f(a) = f(b)$. Thus, θ is well defined.

We now show that θ is a homomorphism. We have

$$\theta((aK)(bK)) = \theta((ab)K) = f(ab) = f(a)f(b) = \theta(aK)\theta(bK)$$

and so θ is indeed a homomorphism.

It remains to show that θ is a bijection. If $h \in \operatorname{Im} f$, then $h = f(g)$ for some $g \in G$ (by definition of $\operatorname{Im} f$). Thus, $h = \theta(gK)$ and so θ is surjective. Since we know that θ is a homomorphism, to show that θ is also injective, it is sufficient to show that $\ker\theta = \{1_{G/K}\} = \{K\}$, see Theorem 4.8.8. We have

$$aK \in \ker\theta \quad \leftrightarrow \quad \theta(aK) = 1_H \quad \leftrightarrow \quad f(a) = 1_H \quad \leftrightarrow \quad a \in \ker f = K$$

and so we do indeed have $\ker\theta = \{K\}$. Thus, θ is a bijection and hence is an isomorphism. This completes the proof that $G/\ker f \simeq \operatorname{Im} f$.

□

4.12 Alternating Groups

In this section we show that for each integer $n \geq 2$, the symmetric group S_n has a subgroup of index 2 (and order $n!/2$). This group is called the **alternating group of degree** n and is denoted by A_n.

Definition 4.12.1. Let $g \in S_n$ and define $c(g)$ to be the number of cycles, including any cycles of length 1, in any cycle representation of g (this equals the number of terms in the cycle structure of g). The **parity** of g is defined to be the parity, odd or even, of the integer $n - c(g)$. A permutation of odd parity is called an **odd permutation** and a permutation of even parity is called an **even permutation**. The parity of a permutation is considered an element $\mathbb{Z}_2$. So even permutations have parity $0 \in \mathbb{Z}_2$, and odd permutations have parity $1 \in \mathbb{Z}_2$.

Definition 4.12.2. A **transposition** is a permutation that interchanges two elements and fixes every other element.

Thus, if A is a set and $a, b \in A$ are distinct, then $(a\ b)$ is a transposition. Thus, the cycle structure of a transposition is $2, 1, 1, \ldots, 1$. Note that if t is a transposition in S_n, then $c(t) = n - 1$, and this implies that t has parity $n - (n-1) = 1$, which is odd.

Lemma 4.12.3. If $t, g \in S_n$ and t is a transposition, then tg has opposite parity to g.

Proof Let $t = (a\ b)$. There are two cases to consider, the case where a and b are in the same cycle of g, and the case where a and b are in different cycles of g. In the first case, we can assume

$$g = (a_1\ a_2\ \cdots\ a_k\ b_1\ b_2\ \cdots\ b_\ell)\cdots$$

and

$$t = (a_1\ b_1),$$

which gives us

$$tg = (a_1\ a_2\ \cdots\ a_k)(b_1\ b_2\ \cdots\ b_\ell)\cdots.$$

In the second case, we can assume

$$g = (a_1\ a_2\ \cdots\ a_k)(b_1\ b_2\ \cdots\ b_\ell)\cdots$$

and

$$t = (a_1\ b_1),$$

which gives us

$$tg = (a_1\ a_2\ \cdots\ a_k\ b_1\ b_2\ \cdots\ b_\ell)\cdots.$$

In both cases the number of cycles changes by 1, and so we see that tg has opposite parity to g. □

Theorem 4.12.4. Any element of S_n is a product of transpositions.

Proof Since any permutation of S_n is a product of cycles (the cycles in its cycle representation), it is sufficient to show that an arbitrary cycle $(a_1\ a_2\ \cdots\ a_k)$ is a product of transpositions, and we have

$$(a_1\ a_2\ \cdots\ a_k) = (a_1\ a_m)(a_1\ a_{m-1})\cdots(a_1\ a_3)(a_1\ a_2).$$

□

Theorem 4.12.5. Let $n \geq 2$ be an integer. The set of all even permutations of S_n forms a normal subgroup of index 2.

Proof We show that parity is a homomorphism from S_n to $\mathbb{Z}_2$. That is, we show that the function $p : S_n \to \mathbb{Z}_2$ given by $p(g) = 0$ if g is even, and $p(g) = 1$ if g is odd, is a homomorphism. Let $g, h \in S_n$. By Theorem 4.12.4, we can write $g = t_1 t_2 \cdots t_r$ where each of $t_1, t_2, \ldots, t_r$ is a transposition. Now consider the product $gh = t_1 t_2 \cdots t_r h$. By Lemma 4.12.3, each multiplication on the left by a transposition changes the parity, and so we have

$$p(gh) = p(t_1 t_2 \cdots t_r h) = r + p(h).$$

Thus, letting h be the identity (which has parity $n - n = 0$) in this expression, we see that $p(g) = r$, which means that we have $p(gh) = p(g) + p(h)$, and we have shown that p is a homomorphism.

By (a) of Theorem 4.11.7, the kernel of p, which is the set of all even permutations of S_n, forms a normal subgroup of S_n. Since $n \geq 2$ the image of p is $\mathbb{Z}_2$, and so by (c) of Theorem 4.11.7, the index of the subgroup is 2. □

Definition 4.12.6. Let $n \geq 2$ be an integer. The group consisting of the even permutations of S_n is called the **alternating group of degree** n and is denoted by A_n.

The homomorphism $p : S_n \to \mathbb{Z}_2$ defined in the proof of Theorem 4.12.5, namely parity, maps even permutations to 0 and odd permutations to 1. Thus, the composition of any two permutations of the same parity is an even permutation (because $p(gh) = p(g) + p(h)$, and because $0 + 0 = 1 + 1 = 0$ in $\mathbb{Z}_2$), and the composition of any two permutations of opposite parity is an odd permutation (because $p(gh) = p(g) + p(h)$, and because $0 + 1 = 1 + 0 = 1$ in $\mathbb{Z}_2$).

4.13 Simple Groups

Definition 4.13.1. A group is **simple** if it is non-trivial and its only normal subgroups are the trivial subgroup and the whole group.

Example 4.13.2. By Lagrange's Theorem (Theorem 4.10.5), any subgroup of a group of prime order has order 1 or p, and hence is the trivial group or the whole group. Thus, any group of prime order is a simple group. We know that if p is prime, then any group of order p is isomorphic to $\mathbb{Z}_p$. So the cyclic groups of prime order form an infinite family of simple groups.

Theorem 4.13.3. For $n = 3$ and for each integer $n \geq 5$, the group A_n is simple. The alternating group A_5 is the smallest non-abelian simple group.

Proof Omitted. □

Simple groups play a similar role in finite groups to the role played by prime numbers in the integers. Every finite group G has a *composition series*

$$1 = H_0 \triangleleft H_1 \triangleleft H_2 \cdots \triangleleft H_n = G$$

such that H_i/H_{i-1} is a simple group for $i = 1, 2, \ldots, n$. The simple groups

$$H_1/H_0, H_2/H_1, \ldots, H_n/H_{n-1}$$

are called the *composition factors* of G.

The *Jordan-Hölder Theorem* states that the composition factors of a finite group are unique up to order and isomorphism (compare to the Fundamental Theorem of Arithmetic which states that the prime factors of a positive integer are unique up to order). Finite simple groups may be thought of as the "basic building blocks" of finite groups, in a similar way as prime numbers may be thought of as the "basic building blocks" of the integers.

Example 4.13.4. Each of the following is a composition series for the group $G = \mathbb{Z}_{12}$ (recall that every subgroup of an abelian group is normal).

(a) $1 = \mathbb{Z}_1 \lhd \mathbb{Z}_2 \lhd \mathbb{Z}_4 \lhd \mathbb{Z}_{12} = G$.

(b) $1 = \mathbb{Z}_1 \lhd \mathbb{Z}_2 \lhd \mathbb{Z}_6 \lhd \mathbb{Z}_{12} = G$.

(c) $1 = \mathbb{Z}_1 \lhd \mathbb{Z}_3 \lhd \mathbb{Z}_6 \lhd \mathbb{Z}_{12} = G$.

In each case, G has two composition factors that are isomorphic to $\mathbb{Z}_2$ and one composition factor that is isomorphic to $\mathbb{Z}_3$, although the order in which these factor groups occur in the composition series is different in each of the cases. Compare with the fact that 12 can be written as a product of primes as $12 = 2 \times 2 \times 3$ or $12 = 2 \times 3 \times 2$ or $12 = 3 \times 2 \times 2$. In general, the orders of the composition factors of $\mathbb{Z}_n$ are the prime factors of n. In this way, the Jordan-Hölder Theorem generalises the Fundamental Theorem of Arithmetic.

In the second half of the twentieth century (and with some small corrections/omissions made later), a program to classify all the finite simple groups was successfully undertaken. Up to isomorphism, the finite simple groups are

(a) $\mathbb{Z}_p$ where p is prime.

(b) A_n where $n \geq 5$.

(c) The so-called "groups of Lie type", which form an infinite family.

(d) 26 "sporadic groups".

The smallest sporadic group, the "Mathieu group M_{11}", has order $7,920$ and the largest sporadic group, the "Monster group", has order

$$808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000.$$

4.14 Table of Small Groups

We give a complete list of groups of small order, up to isomorphism.

It may seem that we have omitted some groups. For example $\mathbb{Z}_2 \times \mathbb{Z}_3$ is a group of order 6. However by Theorem 4.9.3, $\mathbb{Z}_2 \times \mathbb{Z}_3 \simeq \mathbb{Z}_6$, so we have listed a group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_3$.

n	Abelian Groups	Non-abelian Groups
1	$\{1\}$	
2	$\mathbb{Z}_2$	
3	$\mathbb{Z}_3$	
4	$\mathbb{Z}_4$, $\mathbb{Z}_2 \times \mathbb{Z}_2$	
5	$\mathbb{Z}_5$	
6	$\mathbb{Z}_6$	$S_3 = D_3$
7	$\mathbb{Z}_7$	
8	$\mathbb{Z}_8$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$	D_4, Q_8
9	$\mathbb{Z}_9$, $\mathbb{Z}_3 \times \mathbb{Z}_3$	
10	$\mathbb{Z}_{10}$	D_5
11	$\mathbb{Z}_{11}$	
12	$\mathbb{Z}_{12}$, $\mathbb{Z}_6 \times \mathbb{Z}_2$	D_6, A_4, T
13	$\mathbb{Z}_{13}$	
14	$\mathbb{Z}_{14}$	D_7
15	$\mathbb{Z}_{15}$	

Here Q_8 is the group of *quaternions*, and T is the so-called dicyclic group of order 12. There are 14 groups of order 16, 51 of order 32, and 267 different groups of order 64 .... There are 49,910,529,484 (almost 50 billion) groups of order ≤ 2000. Of these, 49,487,365,422 (more than 99%) have order 2^{10}, with all the other orders accounting for 423,164,062 (less than half a billion).

4.15 Fundamental Theorem of Finite Abelian Groups

Examining the table in the previous section, the finite abelian groups that appear are all just products of $\mathbb{Z}_n$ for various n. In fact this is true for any finite abelian group.

Theorem 4.15.1 (Fundamental Theorem of Finite Abelian Groups)**.** Let A be a finite abelian group. Then

$$A \simeq \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_r}$$

for some $m_1, \ldots, m_r \in \mathbb{Z}$.

Proof Omitted. □

There is no similar statement for non-abelian groups—these can be very complicated. It is known that every finite group is isomorphic to a subgroup of a symmetric group S_n for some n, but it is not necessarily easy to extract usable information from this result.

Chapter 5

Number Theory 3: Euler's φ Function and Theorem

5.1 Euler φ Function

Recall that a is invertible in $\mathbb{Z}_n$ if and only if $\gcd(a, n) = 1$ (see Theorem 3.3.5), and that the set $\mathbb{Z}_n^*$ of all invertible elements in $\mathbb{Z}_n$ is a group under multiplication (see Theorem 3.3.9).

Definition 5.1.1. Define a function $\varphi : \mathbb{N} \to \mathbb{N}$ by

$$\varphi(n) = |\{a \in \mathbb{N} : a \leq n, \gcd(a, n) = 1\}|.$$

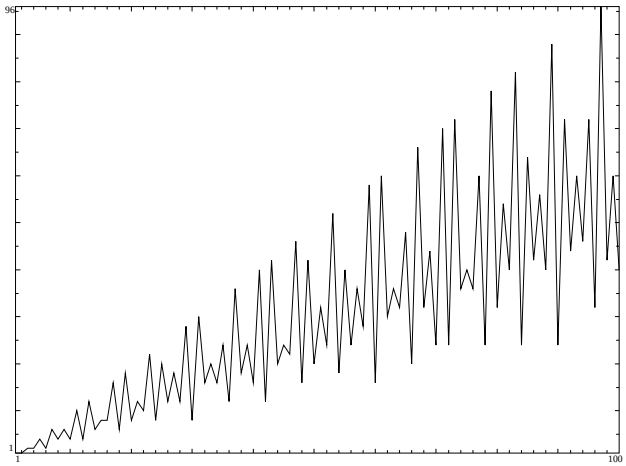
The function φ is called the **Euler φ function**. Since $\varphi(n)$ is the number of positive integers less than or equal to n and relatively prime with n, Theorem 3.3.5 tells us that $\varphi(n)$ is the number of invertible elements in $\mathbb{Z}_n$. That is, $\varphi(n)$ is the order of the group $\mathbb{Z}_n^*$.

Example 5.1.2. We saw in Example 3.3.7 that there are exactly 4 invertible elements in $\mathbb{Z}_{12}$, namely $1, 5, 7$ and 11, Thus, $\varphi(12) = 4$.

The table below gives the value of $\varphi(n)$ for $n = 2, 3, \ldots, 12$ and lists the invertible elements of $\mathbb{Z}_n$.

n	Invertible elements of $\mathbb{Z}_n$	$\varphi(n)$
2	1	1
3	1, 2	2
4	1, 3	2
5	1, 2, 3, 4	4
6	1, 5	2
7	1, 2, 3, 4, 5, 6	6
8	1, 3, 5, 7	4
9	1, 2, 4, 5, 7, 8	6
10	1, 3, 7, 9	4
11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10
12	1, 5, 7, 11	4

Below is a plot of $\varphi(n)$ for $1 \leq n \leq 100$.



Example 5.1.3. Calculate $\varphi(5^3)$.

Let $1 \leq a \leq 5^3 = 125$. If $5 \mid a$, then 5 is a common divisor of a and 5^3, and so $\gcd(a, 5^3) > 1$. On the other hand, the only positive divisors of 5^3 are $1, 5, 5^2$ and 5^3, and so if $\gcd(a, 5^3) > 1$, then $5 \mid a$. Thus, we have shown $\gcd(a, 5^3) > 1$ if and only if $5 \mid a$. There are $\frac{125}{5} = 25$ values of a such that a is a multiple of 5, namely $5 \cdot 1, 5 \cdot 2, \ldots, 5 \cdot 25$, and so 25 values of a such that $\gcd(a, 5^3) > 1$. This leaves $125 - 25 = 100$ values of a such that $\gcd(a, 5^3) = 1$. Thus, $\varphi(5^3) = 100$.

Example 5.1.3 generalises as follows.

Theorem 5.1.4. Let p be prime and $k \in \mathbb{N}$. Then

$$\varphi(p^k) = (p-1)p^{k-1}.$$

Proof Let $1 \leq a \leq p^k$. If $p \mid a$, then p is a common divisor of a and p^k, and so $\gcd(a, p^k) > 1$. On the other hand, the only positive divisors of p^k are $1, p, p^2, \ldots, p^{k-1}$ and p^k, and so if $\gcd(a, p^k) > 1$, then $p \mid a$. Thus, we have shown $\gcd(a, p^k) > 1$ if and only if $p \mid a$. There are $\frac{p^k}{p} = p^{k-1}$ values of a such that a is a multiple of p, namely $p \cdot 1, p \cdot 2, \ldots, p \cdot p^{k-1}$, and so p^{k-1} values of a such that $\gcd(a, p^k) > 1$. This leaves $p^k - p^{k-1}$ values of a such that $\gcd(a, p^k) = 1$. Thus, $\varphi(p^k) = p^k - p^{k-1} = (p-1)p^{k-1}$. $\square$

Theorem 5.1.5. If $\gcd(m, n) = 1$, then $\varphi(mn) = \varphi(m)\varphi(n)$.

Proof We know that for each $r \in \mathbb{N}$, we have $\varphi(n) = |\mathbb{Z}_n^*|$. Thus, to show that $\varphi(mn) = \varphi(m)\varphi(n)$ it suffices to show that there is a bijection from $\mathbb{Z}_{mn}^*$ to $\mathbb{Z}_m^* \times \mathbb{Z}_n^*$ (because $|\mathbb{Z}_{mn}^*| = \varphi(mn)$ and $|\mathbb{Z}_m^* \times \mathbb{Z}_n^*| = |\mathbb{Z}_m^*| \cdot |\mathbb{Z}_n^*| = \varphi(m)\varphi(n)$).

We saw in the proof of Theorem 4.9.3 that the function $f : \mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n$ given by

$$f([a]_{mn}) = ([a]_m, [a]_n)$$

is a bijection (in fact we showed that it is an isomorphism). If $[a]_{mn} \in \mathbb{Z}_{mn}^*$, then $\gcd(a, mn) = 1$ and so we have $\gcd(a, m) = 1$ and $\gcd(a, n) = 1$. Thus, $[a]_m \in \mathbb{Z}_m^*$ and $[a]_n \in \mathbb{Z}_n^*$, and so we have shown that if $[a]_{mn} \in \mathbb{Z}_{mn}^*$, then $f([a]_{mn}) = ([a]_m, [a]_n) \in \mathbb{Z}_m^* \times \mathbb{Z}_n^*$.

Thus, we can define a function $g : \mathbb{Z}_{mn}^* \to \mathbb{Z}_m^* \times \mathbb{Z}_n^*$ by

$$g([a]_{mn}) = f([a]_{mn}) = ([a]_m, [a]_n)$$

and G is injective (because f is injective).

To complete the proof we only need to show that g is surjective. Let $([a]_m, [b]_n) \in \mathbb{Z}_m^* \times \mathbb{Z}_n^*$. We know that there exists a $c \in \mathbb{Z}_{mn}$ such that $f([c]_{mn}) = ([a]_m, [b]_n)$ so if we can show that $[c]_{mn} \in \mathbb{Z}_{mn}^*$, then we have shown that g is surjective (because $g([c]_{mn}) = f([c]_{mn}) = ([a]_m, [b]_n)$). By the definition of f we have $c \equiv a \pmod{m}$ and $c \equiv b \pmod{n}$. Since $\gcd(a, m) = 1$ and $c \equiv a \pmod{m}$, we have $\gcd(c, m) = 1$ (see Theorem 3.3.6). Similarly, we have $\gcd(c, n) = 1$, and so we have $\gcd(c, mn) = 1$. Thus, $[c]_{mn} \in \mathbb{Z}_{mn}^*$ and we have shown that g is surjective. □

We now show that the bijection g in the proof of Theorem 5.1.5 is in fact an isomorphism.

Theorem 5.1.6. If $\gcd(m, n) = 1$, then $\mathbb{Z}_{mn}^* \simeq \mathbb{Z}_m^* \times \mathbb{Z}_n^*$.

Proof We saw in the proof of Theorem 5.1.5 that there is a bijection $g : \mathbb{Z}_{mn}^* \to \mathbb{Z}_m^* \times \mathbb{Z}_n^*$ given by $g([a]_{mn}) = ([a]_m, [a]_n)$. To show that g is an isomorphism we need to show that

$$g([a]_{mn} \odot_{mn} [b]_{mn}) = g([a]_{mn}) \odot_{m\times n} g([b]_{mn})$$

For each $r \in \mathbb{N}$, we use $\odot_r$ to denote the group operation of $\mathbb{Z}_r^*$, and we use $\odot_{m\times n}$ to denote the group operation of $\mathbb{Z}_m^* \times \mathbb{Z}_n^*$. We have

$$\begin{aligned} g([a]_{mn} \odot_{mn} [b]_{mn}) &= g([ab]_{mn}) \\ &= ([ab]_m, [ab]_n) \\ &= ([a]_m \odot_m [b]_m, [a]_n \odot_n [b]_n) \\ &= ([a]_m, [a]_n) \odot_{m\times n} ([b]_m, [b]_n) \\ &= g([a]_{mn}) \odot_{m\times n} g([b]_{mn}). \end{aligned}$$

Thus, g is an isomorphism and $\mathbb{Z}_{mn}^* \simeq \mathbb{Z}_m^* \times \mathbb{Z}_n^*$. □

Theorems 5.1.4 and 5.1.5 give us a formula for calculating $\varphi(n)$ for any $n \in \mathbb{N}$. If $n = p_1^{a_1} \cdots p_k^{a_k}$ where the p_i are distinct primes then

$$\begin{aligned} \varphi(n) &= \varphi(p_1^{a_1})\varphi(p_2^{a_2}) \cdots \varphi(p_k^{a_k}) \\ &= (p_1^{a_1-1})(p_1 - 1)(p_2^{a_2-1})(p_2 - 1) \cdots (p_k^{a_k-1})(p_k - 1) \end{aligned}$$

Example 5.1.7. Calculate $\varphi(540)$.

$$\begin{aligned} 540 &= 2^2 \cdot 3^3 \cdot 5 \\ \varphi(540) &= \varphi(2^2)\ \varphi(3^3)\ \varphi(5) \\ &= 2(2-1)3^2(3-1)(5-1) = 144 \end{aligned}$$

Definition 5.1.8. If the order of $a \in \mathbb{Z}_n^*$ is $\varphi(n)$, equivalently if $\langle a \rangle = \mathbb{Z}_n^*$, then a is called a **primitive root** modulo n.

Example 5.1.9.

(a) It can be easily checked that 2 is a primitive root modulo 5, modulo 9 and modulo 11, but 2 is not a primitive root modulo 7.

(b) There is no primitive root modulo 8. We have $\mathbb{Z}_8^* = \{1, 3, 5, 7\}$ but 1 has order 1 and every other element has order 2 (because $3^2 \equiv 5^2 \equiv 7^2 \equiv 1 \pmod 8$).

Theorem 5.1.10. There exists a primitive root modulo n if and only if $n = 2, 4, p^k$ or $2p^k$ where p is an odd prime and $k \in \mathbb{Z}$.

Proof Omitted. □

Theorem 5.1.11. The number of elements of order n in a cyclic group of order n is $\varphi(n)$.

Proof Let $G = \langle a \rangle$ be a cyclic group of order n and let $x \in \mathbb{N}$. By Theorem 4.7.3, $G = \{1, a, \ldots, a^{n-1}\}$, and by Theorem 4.7.4, a^x has order n if and only if $\gcd(x, n) = 1$. Thus, the number of elements of order n is $\varphi(n)$. □

Theorem 5.1.12. The number of elements of order d in a cyclic group of order n is $\varphi(d)$ if $d \mid n$ and 0 otherwise.

Proof Let G be a cyclic group of order n. Since the order of each element of G divides n (Theorem 4.10.7), if $d \nmid n$, then there are 0 elements of order d. Now suppose $d \mid n$. By Theorem 4.7.7 for each divisor d of n, there is a unique subgroup H_d of order d in G, and by Theorem 4.7.6, H_d is cyclic. Thus, the number of elements of order d in H_d is $\varphi(d)$ by Theorem 5.1.11. But by uniqueness, H_d contains every element of G that has order d (because each such element generates a subgroup of order d). Thus, we have that $\varphi(d)$ is the number of elements of order d in G. □

Theorem 5.1.13. For each positive integer n,

$$\sum_{d|n} \varphi(d) = n$$

where the sum is over all the positive divisors of n.

Proof Consider a cyclic group G of order n. By Theorem 5.1.12, the sum $\sum_{d|n} \varphi(d)$ counts the number of elements of order d in G as d ranges over the divisors of n. Since the order of any element divides n, it counts all the elements of G. □

5.2 Fermat's Little Theorem

Theorem 5.2.1 (Euler)**.** If $a, n \in \mathbb{N}$ and $\gcd(a, n) = 1$, then

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

Proof Let k be the order of a in $\mathbb{Z}_n^*$. Then $k \mid |\mathbb{Z}_n^*|$ by Theorem 4.10.7 (which is an immediate consequence of Lagrange's Theorem). Since $|\mathbb{Z}_n^*| = \varphi(n)$, we have $k \mid \varphi(n)$. Thus, if we let $\varphi(n) = kt$, then we have

$$a^{\varphi(n)} = a^{kt} = (a^k)^t.$$

But $a^k \equiv 1 \pmod{n}$ so $a^{\varphi(n)} \equiv 1 \pmod{n}$. □

Theorem 5.2.2 (Fermat's Little Theorem)**.** Let p be prime. Suppose $a \in \mathbb{N}$ is not divisible by p. Then

$$a^{p-1} \equiv 1 \pmod{p}.$$

Proof Take $n = p$ in Euler's Theorem. Then $\varphi(n) = p - 1$. □

Theorem 5.2.3. Let p be prime. Then every $a \in \mathbb{N}$ satisfies

$$a^p \equiv a \pmod{p}.$$

Proof If $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$, so the result follows on multiplying through by a. If $p \mid a$ then $a \equiv 0 \pmod{p}$ and the result is obvious. □

Example 5.2.4.

(a) Since 101 is prime, Fermat's Little Theorem tells us that $3^{100} \equiv 1 \pmod{101}$. Indeed $a^{100} \equiv 1 \pmod{101}$ for any $a \not\equiv 0 \pmod{101}$.

(b) Calculate the remainder when $5^{1,000,000}$ is divided by 18.

$\varphi(18) = \varphi(2)\varphi(3^2) = 1 \cdot 3(3-1) = 6$, so $5^6 \equiv 1 \pmod{18}$, by Euler's Theorem. Now $1{,}000{,}000 = 6 \cdot 166{,}666 + 4$, so

$$5^{1,000,000} \equiv (5^6)^{166,666} \cdot 5^4 \equiv 1^{166,666} \cdot 5^4 \equiv 25^2 \equiv 7^2 \equiv 13 \pmod{18}.$$

So the remainder is 13.

(c) Calculate $2^{322} \pmod{323}$. Is 323 prime?

$2^{322} = 2^{256} \cdot 2^{64} \cdot 2^2$. By repeated squaring we obtain $2^{64} \equiv 188 \pmod{323}$ and $2^{256} \equiv 35 \pmod{323}$. So $2^{322} \equiv 188 \cdot 35 \cdot 4 \equiv 157 \pmod{323}$. If 323 were prime then $2^{322} \equiv 1 \pmod{323}$ by Fermat's Little Theorem. Since this does not hold, 323 is not prime.

We have determined that 323 is not prime, without finding any factors or doing trial division. This idea underlies most primality tests. It is much easier to test if a number is prime, than it is to find explicit factors.

Note: $(323 = 17 \cdot 19)$.

Chapter 6

Abstract Algebra 3: Rings and Fields

6.1 Rings

So far we have studied algebraic systems with a single binary operation. However many systems have two operations: addition and multiplication. One such system is a *ring*. A ring is an algebraic generalisation of $\mathbb{Z}$, $M_n(\mathbb{R})$, $\mathbb{Z}_n$ etc.

Definition 6.1.1. A **ring** R is a triple $(R, +, \cdot)$ satisfying

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

(b) $(R, \cdot)$ is a semigroup,

(c) The distributive laws hold: for all a, b, $c \in R$

$$\begin{aligned} a \cdot (b + c) &= a \cdot b + a \cdot c \\ (a + b) \cdot c &= a \cdot c + b \cdot c. \end{aligned}$$

We call $+$ *addition* and $\cdot$ *multiplication.*

If we write this out in full detail, a ring is a non-empty set R, on which there are defined two binary operations $+$ and $\cdot$ satisfying for all a, b, $c \in R$

(a) $a + (b + c) = (a + b) + c$

(b) There exists element 0_R with $a + 0_R = a = 0_R + a$.

(c) For every a there exists $-a$ with $a + (-a) = 0_R = (-a) + a$

(d) $a + b = b + a$

(e) $a(bc) = (ab)c$

(f) $a(b + c) = ab + ac$

(g) $(a+b)c = ac + bc$.

By convention we give $\cdot$ higher precedence than $+$, so $a \cdot b + a \cdot c$ means $(a \cdot b) + (a \cdot c)$ (and not $a \cdot (b + a) \cdot c$).

Definition 6.1.2. A ring R is said to be **commutative** if multiplication is commutative; that is, if $ab = ba$ for all $a, b \in R$. It has an **identity** if there is a multiplicative identity, that is if there exists $1_R \in R$ with $1_R a = a = a 1_R$ for all $a \in R$.

Note that addition in a ring is always commutative, and there is always an additive identity. As usual, we will sometimes denote the additive and multiplicative identities by 0 and 1 instead of 0_R and 1_R if there is no risk of confusion.

Example 6.1.3.

(a) $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ are all commutative rings with identity.

(b) $\mathbb{Z}_n$ is a commutative ring with identity.

(c) $M_n(\mathbb{R})$ is a non-commutative ring, with identity I.

(d) The smallest possible ring is $\{0\}$, called the **zero ring,** often denoted 0 (instead of $\{0\}$). It satisfies the axioms for a commutative ring trivially.

Another important example of a ring is a polynomial ring.

Definition 6.1.4. Let R be a commutative ring. The **polynomial ring with coefficients in** R denoted $R[x]$ consists of all formal sums

$$\sum_{i=0}^{\infty} a_i x^i$$

such that $a_i = 0$ for all but finitely many values of i.

Addition and multiplication are defined on $R[x]$ as follows.

$$\sum_{i=0}^{\infty} a_i x^i + \sum_{i=0}^{\infty} b_i x^i = \sum_{i=0}^{\infty} (a_i + b_i) x^i$$

$$\Big(\sum_{i=0}^{\infty} a_i x^i\Big) \cdot \Big(\sum_{j=0}^{\infty} b_j x^j\Big) = \sum_{k=0}^{\infty} c_k x^k \quad \text{where } c_k = \sum_{i+j=k} a_i b_j.$$

Thus, addition and multiplication are defined as one would expect for polynomials. Here x is a variable, not an element of R. The ring $R[y]$ consists of polynomials in the variable y etc.

It is easy to check that $R[x]$ is a commutative ring. If R has an identity, so does $R[x]$ (namely the constant polynomial 1). We can view R as sitting inside $R[x]$, by viewing the elements of R as constant polynomials in $R[x]$.

Example 6.1.5. $\mathbb{Z}_3[x]$ consists of all polynomials with coefficients in $\mathbb{Z}_3$. So for example

$$p(x) = x^2 + 2 \in \mathbb{Z}_3[x] \text{ and } q(x) = x^2 + x + 1 \in \mathbb{Z}_3[x].$$

We have

$$p(x) + q(x) = 2x^2 + x$$

and

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

We have already noted that $M_n(\mathbb{R})$ is a ring. This does not rely on any special properties of the real numbers. We could also form the ring of matrices with entries in $\mathbb{C}$ or entries in $\mathbb{Z}$. All that we need to add and multiply matrices is to be able to add and multiply the corresponding entries, so we can take the entries to lie in any ring. The proofs that multiplication is associative, addition is associative and commutative etc are exactly the same as the proofs that these properties hold in $M_n(\mathbb{R})$.

Definition 6.1.6. Let R be a ring. Let $M_n(R)$ denote the set of $n \times n$ matrices with entries in R. Then $M_n(R)$ is a ring, under matrix multiplication and addition.

Note that for $n \geq 2$, $M_n(R)$ is not commutative, even though R may be commutative. It has an identity if R does, namely the identity matrix I.

A ring is a set with two binary operations, addition and multiplication. The operation of subtraction is not part of the definition, but it is easy to define a subtraction operation $\ominus$ as follows.

Definition 6.1.7. Define $a \ominus b$ to be $a + (-b)$ where $-b$ is the additive inverse of b.

Later we shall write $-$ instead of $\ominus$. However at first we need to distinguish between $-y$, the additive inverse of y, and $x \ominus y$, where $\ominus$ denotes the binary operation of subtraction. The notation works as one would expect.

Theorem 6.1.8. Let R be a ring, and let $a, b, c \in R$. Then

(a) If $a + b = a + c$ then $b = c$.

(b) $a \cdot 0 = 0 = 0 \cdot a$.

(c) $a \cdot (-b) = -(ab) = (-a) \cdot b$.

(d) $-(-a) = a$.

(e) $-(a + b) = (-a) \ominus b$.

(f) $-(a \ominus b) = (-a) + b$.

(g) $(-a)(-b) = ab$.

(h) If R has an identity 1, then $(-1)a = -a$.

Proof

(a) This follows from basic properties of the group $(R, +)$.

(b) We have $a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$ (using the distributive law) and then by cancellation in the additive group we obtain $a \cdot 0 = 0$. We can obtain $0 \cdot a = 0$ in a similar manner.

(c) We have to show that $a \cdot (-b)$ is the additive inverse of ab. That is, we must show $a \cdot (-b) + ab = 0$ (addition is commutative, so then $ab + a \cdot (-b) = 0$ also). But using the distributive property

$$a \cdot (-b) + ab = a\Big((-b) + b\Big) = a \cdot 0 = 0$$

by (b). Similarly $(-a) \cdot b + ab = ((-a) + a)b = 0 \cdot b = 0$.

(d) This follows from properties of the group $(R, +)$ (the inverse of the inverse of a is a).

(e) We must show that $(-a) \ominus b$ is the additive inverse of $(a + b)$, that is, we must show they add to 0. Here $(-a) \ominus b$ stands for $(-a) + (-b)$, so we need to check that $\Big(a + b\Big) + \Big((-a) + (-b)\Big) = 0$. This is clear, since $+$ is commutative and $a + (-a) = 0$ and $b + (-b) = 0$.

The rest are left as exercises. □

Ring homomorphisms and isomorphisms are defined in an analogous way to group homomorphisms and isomorphisms. However, they need to respect both the additive and multiplicative structure of the ring.

Definition 6.1.9. Let R and S be rings. A **ring homomorphism** $f : R \to S$ is a function satisfying

$$\begin{aligned} f(a + b) &= f(a) + f(b) \\ f(ab) &= f(a)f(b) \end{aligned}$$

for all a, $b \in R$. A ring **isomorphism** is a bijective ring homomorphism. We write $R \simeq S$ and say that R and S are **isomorphic** if there exists a ring isomorphism $R \to S$.

If $R \simeq S$ then R and S are structurally identical. The elements in S are just renamed versions of the elements in R.

Theorem 6.1.10. Let R, S and T be rings.

(a) $R \simeq R$.

(b) If $R \simeq S$, then $S \simeq R$.

(c) If $R \simeq S$ and $S \simeq T$, then $R \simeq T$.

Proof The proof is similar to that of Theorem 4.9.5 (which shows the corresponding properties for group isomorphism). □

6.2 Units and Fields

In a ring R, the invertible elements in $(R, \cdot)$ are called *units*.

Definition 6.2.1. Let R be a ring with identity. An element $u \in R$ is a **unit** if it has a multiplicative inverse. That is, u is a unit if and only if there exists $v \in R$ with $uv = 1 = vu$. We denote v by u^{-1} and call it the **inverse** of u. The set of units of R is denoted R^*.

Example 6.2.2.

(a) $\mathbb{Z}^* = \{1, -1\}$.

(b) $\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}$.

(c) $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$.

(d) $\mathbb{Z}_n^*$ consists of the congruence classes $[a]$ with $\gcd(a, n) = 1$. This is a set with $\varphi(n)$ elements.

(e) The units of $M_n(\mathbb{R})$ are the invertible $n \times n$ matrices (those with non-zero determinant).

Definition 6.2.3. A **field** is a set F together with two binary operations, addition denoted by $+$ and multiplication denoted by $\cdot$, such that for all $a, b, c \in F$

(a) $a + (b + c) = (a + b) + c$ and $a(bc) = (ab)c$ (associativity of addition and multiplication);

(b) $a + b = b + a$ and $ab = ba$ (commutativity of addition and multiplication);

(c) there exist distinct $0, 1 \in F$ such that $a + 0 = a$ and $a \cdot 1 = a$ (additive and multiplicative identities);

(d) there exists $-a$ such that $a + (-a) = 0$ (additive inverses for all elements);

(e) for $a \neq 0$ there exists a^{-1} such that $aa^{-1} = 1$ (multiplicative inverses for all non-zero elements);

(f) $a \cdot (b + c) = (ab) + (ac)$ (distributivity of multiplication over addition).

In the axioms for a ring (see Definition 6.1.1) there are two distributive laws, whereas there is only one distributive law in the above axioms for a field. However, as the following result shows, in a field the second distributive law can be deduced from the first by using commutativity of multiplication.

Lemma 6.2.4. If $(F, +, \cdot)$ is a field and $a, b, c \in F$, then $(a + b)c = ac + bc$.

Proof Let $(F, +, \cdot)$ is a field and $a, b, c \in F$. Then

$$\begin{aligned} (a+b)c &= c(a+b) && \text{(by commutativity of multiplication)} \\ &= ca + cb && \text{(by distributivity of multiplication over addition)} \\ &= ac + bc && \text{(by commutativity of multiplication).} \end{aligned}$$

□

The following result gives some basic properties of fields which follow easily from the axioms.

Lemma 6.2.5. If $(F, +, \cdot)$ is a field, then for all $a, b \in F$

(a) $0 \cdot a = 0$;

(b) $a \cdot b = 0$ implies $a = 0$ or $b = 0$; and

(c) $-a = (-1) \cdot a$.

Proof Let $(F, +, \cdot)$ be a field, and let $a, b \in F$.

(a)
$$\begin{aligned} 0 \cdot a &= (0 + 0) \cdot a && \text{(0 is the additive identity)} \\ &= 0 \cdot a + 0 \cdot a && \text{(by Lemma 6.2.4)} \end{aligned}$$
and this implies $0 \cdot a = 0$.

(b) Suppose $a \cdot b = 0$ and $b \neq 0$. Then b^{-1} exists and we have $(a \cdot b) \cdot b^{-1} = 0 \cdot b^{-1}$. Using associativity of multiplication and the result from (a) we thus obtain $a \cdot (b \cdot b^{-1}) = 0$, from which it follows that $a = 0$.

(c) We have
$$a + (-1) \cdot a = 1 \cdot a + (-1) \cdot a = (1 + (-1)) \cdot a = 0 \cdot a = 0$$
and (together with commutativity of addition) this implies that $(-1) \cdot a$ is the additive inverse of a.

□

It can be seen that a field satisfies the ring axioms, see Definition 6.1.1, with the second distributive law being proved in Lemma 6.2.4. Thus, a field is a non-trivial commutative ring with identity in which every non-zero element is a unit. The zero ring is not a field.

Note that if $(F, +, \cdot)$ is a field, then $(F, +)$ is an abelian group with identity 0, and $(F \setminus \{0\}, \cdot)$ is an abelian group with identity 1 (the fact that $(F \setminus \{0\}, \cdot)$ is closed follows from Lemma 6.2.5 (b)).

In a field we have division defined for all elements except that division by 0 is not defined. For $b \neq 0$ we have $a \div b = ab^{-1}$.

Example 6.2.6.

(a) $\mathbb{Q}$ and $\mathbb{R}$ are fields.

(b) $\mathbb{Z}$ is not a field.

(c) If p is prime, then $\mathbb{Z}_p$ is a field.

(d) If n is not prime, then $\mathbb{Z}_n$ is not a field.

6.3 Polynomial Rings

Definition 6.3.1. The **degree** of a polynomial $f(x) \in R[x]$, denoted $\deg(f(x))$, is the largest integer k such that the coefficient of x^k in $f(x)$ is not zero. In the special case of the zero polynomial (where all coefficients are zero) the degree is defined to be $-\infty$.

Note that the degree of a polynomial is in $\{-\infty, 0, 1, 2, \ldots\}$.

Example 6.3.2. The degree of $4x^3 - x + 1$ is 3, and the degree of the constant polynomial $3 = 3 \cdot x^0$ is 0.

Any field F is also a ring, and so it makes sense to talk about the ring $F[x]$ of polynomials over a field F.

Theorem 6.3.3. Let F be a field and let $f, g \in F[x]$. Then

$$\deg(f+g) \leq \max(\deg(f), \deg(g)); \qquad \textit{and}$$
$$\deg(f \cdot g) = \deg(f) + \deg(g).$$

Proof Both properties follow easily from the definitions of multiplication and addition of polynomials. If exactly one of f or g is the zero polynomial, then $\deg(f \cdot g) = \deg(0) = -\infty$ and $\deg(f) + \deg(g) = c + (-\infty) = -\infty$ where c is the degree of the nonzero polynomial. If both f and g are the zero polynomial, then $\deg(f \cdot g) = \deg(0) = -\infty$ and $\deg(f) + \deg(g) = (-\infty) + (-\infty) = -\infty$. □

We now discuss polynomial division.

Definition 6.3.4. Let F be a field. For $f(x), g(x) \in F[x]$ we say that $g(x)$ **divides** $f(x)$ and write $g(x) \mid f(x)$ if and only if there exists $q(x) \in F[x]$ such that $f(x) = q(x)g(x)$.

The following theorem is an analogue of the result that for integers a and b with $b \neq 0$, there exist unique integers q and r with $0 \leq r < |b|$ such that $a = qb + r$ (see Theorem 1.3.1).

Theorem 6.3.5. If F is a field and $f(x), g(x) \in F[x]$ with $g(x) \neq 0$, then there exist polynomials $q(x), r(x) \in F[x]$ with $\deg(r) < \deg(g)$ such that

$$f(x) = q(x)g(x) + r(x).$$

Proof Let

$$A = \{f(x) - g(x)q(x) : q(x) \in F[x]\}$$

and consider the set $D = \{\deg(a) : a(x) \in A\}$. The set D is nonempty because it contains $\deg(f)$, and is a subset of $\mathbb{N} \cup \{0, -\infty\}$. Thus, D contains a least element n. Let $r(x)$ be an element of A with $\deg(r) = n$. Thus, by the definition of A, we have

$$f(x) = q(x)g(x) + r(x)$$

for some $q(x) \in F[x]$. It remains to show that $\deg(r) < \deg(g)$.

If $r(x) = 0$, then we have $\deg(r) = -\infty < 0 \leq \deg(g)$ (because $g(x) \neq 0$). So we can assume $r(x) \neq 0$. Let $m = \deg(g)$, let $r(x) = r_n x^n + \cdots + r_1 x + r_0$ and let $g(x) = g_m x^m + \cdots + g_1 x + g_0$. So $r_n \neq 0$ and $g_m \neq 0$. Suppose for a contradiction that $n \geq m$.

Consider the polynomial

$$s(x) = r(x) - \tfrac{r_n}{g_m} x^{n-m} g(x).$$

Since $r(x) = f(x) - q(x)g(x)$, we have

$$s(x) = f(x) - q(x)g(x) - \tfrac{r_n}{g_m} x^{n-m} g(x) = f(x) - g(x)\Big(q(x) + \tfrac{r_n}{g_m} x^{n-m}\Big),$$

and so $s(x) \in A$ and $\deg(s) \in D$. But we also have

$$\begin{aligned} s(x) &= r(x) - \tfrac{r_n}{g_m} x^{n-m}(g_m x^m + \cdots + g_1 x + g_0) \\ &= r(x) - (r_n x^n + \cdots + \tfrac{r_n g_1}{g_m} x^{n-m+1} + \tfrac{r_n g_0}{g_m} x^{n-m}) \\ &= (r_n x^n + \cdots + r_1 x + r_0) - (r_n x^n + \cdots + \tfrac{r_n g_1}{g_m} x^{n-m+1} + \tfrac{r_n g_0}{g_m} x^{n-m}). \end{aligned}$$

In $s(x)$, the coefficients of x^n are r_n and $-r_n$, which means that $\deg(s) < n$ (recall that $r_n \neq 0$). This contradicts the fact that n is the least element of D. We conclude that $\deg(r) < \deg(g)$. □

For given $f(x)$ and $g(x)$, the following example illustrates a general method for finding $q(x)$ and $r(x)$ with $\deg(r) < \deg(g)$ such that $f(x) = q(x)g(x) + r(x)$.

Example 6.3.6. Let $f(x) = 2x^4 + 5x^3 + x + 3$ and $g(x) = x^2 + 4$ be polynomials in $\mathbb{Z}_7[x]$. Find polynomials $q(x)$ and $r(x)$ in $\mathbb{Z}_7[x]$ with $\deg(r) < \deg(g)$ such that $f(x) = q(x)g(x) + r(x)$.

We solve the problem using "long division":-

$$\begin{array}{rl|llllllll} & & & & & 2x^2 & + & 5x & + & 6 \\ \hline x^2+4 & | & 2x^4 & + & 5x^3 & + & & + & x & + & 3 \\ & & 2x^4 & & & + & x^2 & & & & \\ \cline{3-6} & & & & 5x^3 & + & 6x^2 & + & x & + & 3 \\ & & & & 5x^3 & & & + & 6x & & \\ \cline{4-7} & & & & & & 6x^2 & + & 2x & + & 3 \\ & & & & & & 6x^2 & & & + & 3 \\ \cline{5-8} & & & & & & & & 2x & & \end{array}$$

So we obtain $q(x) = 2x^2 + 5x + 6$ and $r(x) = 2x$. We can check the answer as follows.

$$q(x)g(x) + r(x) = (2x^2 + 5x + 6)(x^2 + 4) + 2x = 2x^4 + 5x^3 + 6x + 3 + 2x = 2x^4 + 5x^3 + x + 3 = f(x)$$

Although polynomials are formal sums, rather than functions, we have an obvious definition of the evaluation of a polynomial in $R[x]$ at an element $c \in R$.

Definition 6.3.7. Let R be a commutative ring with identity, let $c \in R$, and let $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \in R[x]$. The **evaluation** of $f(x)$ at c, denoted by $f(c)$, is the element of R given by

$$f(c) = a_n c^n + a_{n-1} c^{n-1} + \cdots + a_1 c + a_0.$$

Definition 6.3.8. Let R be a commutative ring with identity and let $f(x) \in R[x]$. An element $c \in R$ is a **root** of $f(x)$ if $f(c) = 0$.

Theorem 6.3.9. Let F be a field and let $f(x) \in F[x]$. Then c is a root of $f(x)$ if and only if $x - c$ divides $f(x)$.

Proof Suppose c is a root of $f(x)$. By Theorem 6.3.5 we can write $f(x) = q(x)(x-c)+r(x)$ where $\deg(r) < \deg(x-c) = 1$. Thus, $\deg(r) \in \{0, -\infty\}$. If $\deg(r) = 0$, then $r(x)$ is a nonzero constant and we have

$$0 = f(c) = q(c)(c-c) + r(c) = r(c),$$

which is a contradiction. We conclude that $\deg(r) = -\infty$. So $r(x)$ is the zero polynomial, and we have $f(x) = q(x)(x-c)$. That is, we have $x - c$ divides $f(x)$.

Now suppose $x - c$ divides $f(x)$. Then we have $f(x) = q(x)(x-c)$ for some $q(x) \in F[x]$, and so

$$f(c) = q(c)(c-c) = 0.$$

Thus, c is a root of $f(x)$. □

Theorem 6.3.10. Let F be a field. If the number of roots of a polynomial $f(x) \in F[x]$ is greater than $\deg(f)$, then $f(x)$ is the zero polynomial.

Proof The proof is by induction on $\deg(f)$. If $f(x)$ is the zero polynomial then we are done, so we can assume $\deg(f) \geq 0$. If $\deg(f) = 0$, then $f(x)$ is a nonzero constant and thus has 0 roots. So the result holds when $\deg(f) = 0$. Let $\deg(f) = n > 0$ and assume the result holds for polynomials of degree $n - 1$.

Now suppose $c_1, c_2, \ldots, c_{n+1}$ are distinct roots of $f(x)$. Then by Theorem 6.3.9 we can write

$$f(x) = (x - c_{n+1})q(x)$$

for some $q(x) \in F[x]$. It is clear that $\deg(q) \leq n - 1$. For $i = 1, 2, \ldots, n$, we have

$$0 = f(c_i) = (c_i - c_{n+1})q(c_i).$$

Since $c_i - c_{n+1} \neq 0$, this implies $q(c_i) = 0$ for $i = 1, 2, \ldots, n$. Thus, $c_1, c_2, \ldots, c_n$ are n distinct roots of $q(x)$ and so by the inductive hypothesis $q(x)$ is the zero polynomial. Thus, $f(x)$ is also the zero polynomial and the result holds by induction. □

In $F[x]$, the analogue of a prime number is an *irreducible* polynomial.

Definition 6.3.11. Let F be a field. A polynomial $f(x) \in F[x]$ of degree at least 1 is **irreducible** if there do not exist polynomials $g(x), h(x) \in F[x]$, each of degree at least 1, such that $f(x) = g(x)h(x)$.

Example 6.3.12. The polynomial $x^2 + 1$ is irreducible as a polynomial in $\mathbb{R}[x]$. However, it is not irreducible as a polynomial in $\mathbb{Z}_2[x]$ because then we have

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

Theorem 6.3.13. For any prime p and any positive integer n, there exists an irreducible polynomial of degree n in $\mathbb{Z}_p[x]$.

Proof Omitted. □

6.4 Finite Fields

Finite fields are extremely useful in cryptography, coding, combinatorics, etc.

We have seen that $\mathbb{Z}_p$ is a field when p is prime. Are there any other finite fields? The following theorem gives a complete answer to this question.

Theorem 6.4.1. If F is a finite field, then F has p^n elements where p is prime and n is a positive integer. Moreover, there exists exactly one finite field (up to isomorphism) of order p^n for each p and n.

Proof Omitted. □

Note that two fields are said to be isomorphic if and only if they are isomorphic as rings.

Definition 6.4.2. Let q be a prime power, that is, $q = p^n$ where p is prime and n is a positive integer. The unique (up to isomorphism) field of order q is denoted by $\mathbb{F}_q$.

We will now show how to construct the finite field $\mathbb{F}_q$ of order $q = p^n$ where p is prime and $n \geq 1$.

Let $f(x)$ be an irreducible polynomial of degree n in $\mathbb{F}_p[x]$, where p is prime and $n \geq 1$. Define an equivalence relation $\equiv_f$, or just $\equiv$ when f does not need to be specified, on $\mathbb{F}_p[x]$ as follows. For all $a(x), b(x) \in \mathbb{F}_p[x]$,

$$a(x) \equiv b(x) \qquad \text{if and only if} \qquad f(x) \mid (a(x) - b(x)).$$

Notice the similarity with the definition of equivalence in $\mathbb{Z}_p$ where we have $a \equiv b \pmod{p}$ if and only if $p \mid (a - b)$. It is easy to check that $\equiv$ is indeed an equivalence relation on $\mathbb{F}_p[x]$. If $a(x) \equiv_f b(x)$ then we say $a(x)$ and $b(x)$ are equivalent modulo $f(x)$, and write

$$a(x) \equiv b(x) \pmod{f(x)}.$$

For each $a(x) \in \mathbb{F}_p[x]$ the set $\{b(x) \in \mathbb{F}_p[x] : b(x) \equiv a(x)\}$ is called the equivalence class of $a(x)$ modulo $f(x)$ and is denoted by $[a(x)]_f$, or just $[a(x)]$ if we do not need to specify the irreducible polynomial $f(x)$. The set of all equivalence classes of $\mathbb{F}_p[x]$ modulo $f(x)$ is denoted by $\mathbb{F}_p[x]_f$ (just like the set of all congruence classes of $\mathbb{Z}$ modulo p is denoted by $\mathbb{Z}_p$).

Define binary operations of addition $+$ and multiplication $\cdot$ on $\mathbb{F}_p[x]_f$ by

$$[a(x)] + [b(x)] = [a(x) + b(x)] \qquad \text{and} \qquad [a(x)] \cdot [b(x)] = [a(x)b(x)].$$

It can be verified that these binary operations are well defined and it turns out that $(\mathbb{F}_p[x]_f, +, \cdot)$ is the finite field $\mathbb{F}_q$ of order $q = p^n$ where n is the degree of the irreducible polynomial $f(x)$.

Any polynomial in $\mathbb{F}_p[x]$ is equivalent modulo $f(x)$ to a polynomial of degree at most $n-1$. To see this observe that by Theorem 6.3.5 there exist polynomials $q(x), r(x) \in \mathbb{F}_p[x]$ with $0 \leq \deg(r(x)) < n$ or $r(x) = 0$ such that

$$a(x) = q(x)f(x) + r(x).$$

Thus, $f(x) \mid (a(x) - r(x))$ and so $a(x) \equiv r(x) \pmod{f(x)}$.

The polynomials of degree at most $n-1$ in $\mathbb{F}_p[x]$ are of the form

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

where each $a_i \in \mathbb{F}_p$. The difference of any two distinct such polynomials is never divisible by $f(x)$ (because $f(x)$ has degree n). Thus, the set of polynomials of degree at most $n-1$ in $\mathbb{F}_p[x]$ forms a set of representatives for the equivalence classes of $\mathbb{F}_p[x]$ modulo $f(x)$. So we can take these polynomials as the elements of our field $\mathbb{F}_q$. The number of polynomials of degree at most $n-1$ in $\mathbb{F}_p[x]$ is p^n because there are p choices for each of the n coefficients $a_0, a_1, \ldots, a_{n-1}$.

Example 6.4.3. Construction of $\mathbb{F}_9$. Let $p = 3$ and $n = 2$ so that $p^n = 9$. The elements of the field $\mathbb{F}_9$ are

$$0, 1, 2, x, x+1, x+2, 2x, 2x+1, 2x+2.$$

In this example we take $f(x) = x^2 + x + 2 \in \mathbb{F}_3[x]$ as our irreducible polynomial. The addition and multiplication tables for the field are shown below.

$+$	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
0	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
1	1	2	0	$x+1$	$x+2$	x	$2x+1$	$2x+2$	$2x$
2	2	0	1	$x+2$	x	$x+1$	$2x+2$	$2x$	$2x+1$
x	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$	0	1	2
$x+1$	$x+1$	$x+2$	x	$2x+1$	$2x+2$	$2x$	1	2	0
$x+2$	$x+2$	x	$x+1$	$2x+2$	$2x$	$2x+1$	2	0	1
$2x$	$2x$	$2x+1$	$2x+2$	0	1	2	x	$x+1$	$x+2$
$2x+1$	$2x+1$	$2x+2$	$2x$	1	2	0	$x+1$	$x+2$	x
$2x+2$	$2x+2$	$2x$	$2x+1$	2	0	1	$x+2$	x	$x+1$

$\cdot$	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
2	0	2	1	$2x$	$2x+2$	$2x+1$	x	$x+2$	$x+1$
x	0	x	$2x$	$2x+1$	1	$x+1$	$x+2$	$2x+2$	2
$x+1$	0	$x+1$	$2x+2$	1	$x+2$	$2x$	2	x	$2x+1$
$x+2$	0	$x+2$	$2x+2$	$x+1$	$2x$	2	$2x+2$	1	x
$2x$	0	$2x$	x	$x+2$	2	$2x+2$	$2x+1$	$x+1$	1
$2x+1$	0	$2x+1$	$x+2$	$2x+2$	x	1	$x+1$	2	$2x$
$2x+2$	0	$2x+2$	$x+1$	2	$2x+1$	x	1	$2x$	$x+2$

We illustrate how this multiplication table is constructed with the example of $(x+2)(2x+1)$. Using polynomial multiplication and then reducing the coefficients modulo 3 we have

$$(x+2)(2x+1) = 2x^2 + 5x + 2 = 2x^2 + 2x + 2.$$

Dividing the polynomial $2x^2 + 2x + 2$ by our irreducible polynomial $f(x) = x^2 + x + 2$ (see Example 6.3.6) we see that

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

So $2x^2 + 2x + 2 \equiv 1 \pmod{f(x)}$. This is consistent with the entry in the multiplication table above.

6.5 More on Finite Fields

Theorem 6.5.1. Let $q = p^n$ where p is prime and $n \in \mathbb{N}$. Then $(\mathbb{F}_q, +) \simeq \mathbb{Z}_p^n$.

Proof We show that the function $\theta : \mathbb{Z}_p^n \to \mathbb{F}_q$ given by

$$\theta((a_0, a_1, a_2, \ldots, a_{n-1})) = [a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}]$$

is an isomorphism. We saw earlier that the polynomials in $\mathbb{F}_p[x]$ that have degree at most $n-1$ form a set of representatives for the p^n equivalence classes of $\mathbb{F}_p[x]$ modulo an irreducible polynomial of degree n. Thus, θ is a bijection from $\mathbb{Z}_p^n$ to $\mathbb{F}_q$. Also,

$$\begin{aligned}
& \theta((a_0, a_1, a_2, \ldots, a_{n-1})) + \theta((b_0, b_1, b_2, \ldots, b_{n-1})) \\
= \; & [a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}] + [b_0 + b_1 x + b_2 x^2 + \cdots + b_{n-1} x^{n-1}] \\
= \; & [(a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}) + (b_0 + b_1 x + b_2 x^2 + \cdots + b_{n-1} x^{n-1})] \\
= \; & [(a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_{n-1} + b_{n-1})x^{n-1}] \\
= \; & \theta((a_0 + b_0, a_1 + b_1, a_2 + b_2, \ldots, a_{n-1} + b_{n-1})) \\
= \; & \theta((a_0, a_1, a_2, \ldots, a_{n-1}) + (b_0, b_1, b_2, \ldots, b_{n-1}))
\end{aligned}$$

so θ is an isomorphism. □

Theorem 6.5.2. Let $q = p^n$ where p is prime and $n \in \mathbb{N}$. Then $\mathbb{F}_q^*$ is cyclic.

Proof Let $d \in \mathbb{N}$ and consider the polynomial $x^d - 1 \in \mathbb{F}_q[x]$. If the (multiplicative) order of an element $a \in \mathbb{F}_q$ divides d, then by Theorem 4.5.3 we have $a^d = 1$, which implies that a is a root of $x^d - 1$. By Theorem 6.3.10, $x^d - 1$ has at most d roots. Thus, the number of elements of $\mathbb{F}_q$ whose order divides d is at most d.

Now suppose there is an element α of order d in $\mathbb{F}_q^*$. Then by Lagrange's Theorem (Theorem 4.10.5), the order of each of the d elements in the subgroup $\langle \alpha \rangle$ of $\mathbb{F}_q^*$ divides d. Thus, an element of $\mathbb{F}_q^*$ has order dividing d if and only if it belongs to $\langle \alpha \rangle$. In particular, any element of $\mathbb{F}_q^*$ having order d is in $\langle \alpha \rangle$.

Now, the number of elements of order equal to d in $\langle \alpha \rangle$ is $\varphi(d)$. Thus, the number of elements of order d in $\mathbb{F}_q^*$ is either 0 (if there is no element of order d in $\mathbb{F}_q^*$) or $\varphi(d)$ (if there is an element of order d in $\mathbb{F}_q^*$). Let $\psi(d)$ denote the number of elements of order d in $\mathbb{F}_q^*$, and note that $\psi(d) \leq \varphi(d)$.

For a contradiction, suppose that $\mathbb{F}_q^*$ is not cyclic. Then there is no element of order $q-1$ in $\mathbb{F}_q^*$ and so we have $\psi(q-1) = 0 < \varphi(q-1)$. Thus, we have

$$\sum_{d|q-1} \psi(d) < \sum_{d|q-1} \varphi(d).$$

But Theorem 5.1.13 tells us that

$$\sum_{d|q-1} \varphi(d) = q - 1$$

and so we have

$$\sum_{d|q-1} \psi(d) < q - 1.$$

Since the order of each element of $\mathbb{F}_q^*$ divides $q-1$, the left side of the above inequality is the number of elements in $\mathbb{F}_q^*$. Thus we have less than $q-1$ elements in $\mathbb{F}_q^*$. This is a contradiction and we conclude that $\mathbb{F}_q^*$ is cyclic. □

If p is prime, then we have $\mathbb{F}_p^* = \mathbb{Z}_p^*$, and an element of order $p-1$ in $\mathbb{Z}_p^*$ is called a primitive root. Theorem 6.5.2 guarantees that $\mathbb{Z}_p^*$ has a primitive root for every prime p. We saw in Example 5.1.9 (a) that 2 is a primitive root in $\mathbb{Z}_5$ and $\mathbb{Z}_{11}$, but not in $\mathbb{Z}_7$. In $\mathbb{Z}_7$ the element 3 is a primitive root. It is an unsolved problem whether there are infinitely many primes p such that 2 is a primitive root in $\mathbb{Z}_p^*$.

Theorem 6.3.13 tells us that there exists an irreducible polynomial of degree n in $\mathbb{Z}_p[x]$ for any prime p and any positive integer n. As we noted above, such a polynomial can be used to construct $\mathbb{F}_q$ where $q = p^n$ and $n \geq 2$. We now define a special type of irreducible polynomial that gives a much easier construction of $\mathbb{F}_q$.

Definition 6.5.3. Let p be prime, let $f(x) \in \mathbb{F}_p[x]$ be an irreducible polynomial of degree n, and let F be the field consisting of the equivalence classes of $\mathbb{F}_p[x]$ modulo $f(x)$. If the polynomial x generates the multiplicative group $(F \setminus \{0\}, \cdot)$, then $f(x)$ is called a **primitive polynomial**.

Theorem 6.5.4. For any prime p and any positive integer n, there exists a primitive polynomial of degree n in $\mathbb{Z}_p[x]$.

Proof Omitted. □

The following tables give a primitive polynomial of degree n over $\mathbb{F}_p$ for various small values of p and n.

n	Primitive polynomial of degree n over $\mathbb{F}_2$
1	$x+1$
2	x^2+x+1
3	x^3+x+1
4	x^4+x+1
5	x^5+x^2+1
6	x^6+x+1

n	Primitive polynomial of degree n over $\mathbb{F}_3$
1	$x+1$
2	x^2+x+2
3	x^3+2x+1
4	x^4+x+2
5	x^5+2x+1
6	x^6+x+2

n	Primitive polynomial of degree n over $\mathbb{F}_5$
1	$x+2$
2	x^2+x+2
3	x^3+3x+2
4	x^4+x^2+2x+2

n	Primitive polynomial of degree n over $\mathbb{F}_7$
1	$x+2$
2	x^2+x+3
3	x^3+3x+2
4	x^4+x^2+3x+5

Example 6.5.5. Construction of $\mathbb{F}_9$ using the primitive polynomial $x^2 + x + 2$ of degree 2 over $\mathbb{F}_3$.

We work modulo the primitive polynomial $x^2 + x + 2$. Using polynomial division and writing just " =" rather than " $\equiv$" we have

$$\begin{array}{rcll} x^0 & = & 1 & \\ x^1 & = & x & \\ x^2 & = & 2x + 1 & \\ x^3 & = & 2x^2 + x & = \quad 2x + 2 \\ x^4 & = & 2x^2 + 2x & = \quad 2 \\ x^5 & = & 2x & \\ x^6 & = & 2x^2 & = \quad x + 2 \\ x^7 & = & x^2 + 2x & = \quad x + 1 \\ (x^8 & = & x^2 + x & = \quad 1). \end{array}$$

As expected, x generates all the nonzero elements of $\mathbb{F}_9$. This allows us to calculate products very easily. For example, we have

$$(x + 2)(x + 1) = x^6 x^7 = x^{13} = x^8 x^5 = x^5 = 2x.$$