On Beyond $\mathbb{Z}$:
Groups, Rings, and Fields

Michael E. O'Sullivan
mosullivan@sdsu.edu

August 25, 2024

Contents

Bibliography

[Hun12] T.W. Hungerford, *Abstract algebra: An introduction*, Cengage Learning, 2012.

[Men15] E. Mendelson, *Introduction to mathematical logic*, Discrete Mathematics and Its Applications, CRC Press, 2015.

[Ros11] K.H. Rosen, *Elementary number theory and its applications*, Addison-Wesley, 2011.

Chapter 1

Getting Started

1.1 The Integers

What are the integers? This is not a simple question, if you want to be rigorous about defining the integers. Formally doing so would distract from developing our core topics, so we will take as our foundation the following. The ring of integers $\mathbb{Z}$ is

- the set of natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$ along with the number 0 and the additive inverses of the natural numbers $\{-1, -2, -3. \dots\}$;
- the operation of addition (and the properties of addition we know from elementary school);
- the operation of multiplication (and the properties of multiplication we know from elementary school);
- the ordering defined by positive numbers being greater than 0 and $a > b$ if and only if $a - b > 0$;
- properties of order related to addition and multiplication such as $a > b$ implies $a + c > b + c$;
- the well-ordering principle—any non-empty subset of the natural numbers has a least element.

We may think of subtraction as either $a - b = a + (-b)$ or equivalently (after some argument) $a - b$ is the the number s (which we should show is unique) such that $s + b = a$. Division will be dealt with below. While the integers are familiar from elementary school, the well-ordering principle is not (unless you attended a very special elementary school!). It is actually key to the formal definition of the integers (see the Peano axioms and [Men15]) and to mathematical induction stripped to its essentials:

> **The Principle of Mathematical Induction**
> Let K be a subset of $\mathbb{N}$ satisfying these two properties
>
> - $1 \in K$;
> - whenever $a \in K$ it is also true that $a + 1 \in K$.
>
> Then $K = \mathbb{N}$.

The well-ordering principle is easily extended to say that any non-empty subset of the integers that is bounded below has a least element. Similarly, the principle of induction can use any integer (instead of 1) as an initial element in the set K.

Beyond the basic properties above there are five main results for the integers that are fundamental. For the purposes of easy reference I will call them the Quotient-Remainder (QR) Theorem, the Greatest Common Divisor (GCD) Theorem, the Euclidean Algorithm, the Prime-Irreducible Theorem (Euclid's lemma), and the Unique Factorization Theorem (the Fundamental Theorem of Arithmetic). The Quotient-Remainder Theorem and Unique Factorization will be familiar; the other results, perhaps less so. The proofs here will be concise, and just a few exercises are included because this material is treated very well in other resources [Hun12].

In addition to these key results about the integers we introduce modular arithmetic in this section. Modular arithmetic creates a new algebraic structure known as the integers modulo n (for some $n > 1$), which we write $\mathbb{Z}/n$.

The Quotient-Remainder Theorem and Divisibility

We have implicitly used the Quotient-Remainder Theorem since elementary school, when we computed (the unique!) quotient and remainder of two integers. The proof relies on something sophisticated: the well-ordering principle.

Theorem 1.1.1 (Quotient-Remainder)**.** *Let a and b be integers with $b \neq 0$. There exist unique integers q, r such that*

(1) $a = bq + r$, *and*

(2) $0 \leq r < |b|$.

If the remainder of a divided by b is 0, we say b **divides** *a and a is a* **multiple of** *b.*

Proof. We prove this for $b > 0$ and leave modifications necessary for the case $b < 0$ as an exercise. Consider the set $S = \{a - bc : c \in \mathbb{Z}\} \cap \mathbb{N}_{\geq 0}$. As a nonempty subset of the nonnegative integers, it has a least element. Let r be the minimal element of S, and let q be the integer such that $r = a - bq$. If $r \geq b$ we would have a contradiction because then $r - b \geq 0$ and $r - b = a - b(q + 1)$ would put $r - b \in S$. Thus we must have $0 \leq r < b$. This establishes existence of q, r as claimed.

To prove uniqueness, suppose another r', q' satisfy (1) and (2) and suppose without loss of generality that $r \geq r'$. (We want to show that $r' = r$ and $q' = q$.) Then $a = bq + r = bq' + r'$ so $r - r' = b(q' - q)$. Now $b > r - r' \geq 0$ but $b(q' - q)$ is a multiple of b. The only multiple of b in the interval $[0, b)$ is 0, so the only possibility is $r - r' = 0 = b(q - q)'$, and therefore $r = r'$ and $q = q'$. □

It is fairly common in programming languages (in particular in Python and Sage) to write the integer quotient as $a//b$ and the remainder as $a\%b$. We will use this in the exercises and the discussion of the Euclidean algorithm in this chapter.

Exercises 1.1.2. More on the Quotient-Remainder Theorem.

(a) Prove the QR Theorem for negative integers: Only minor changes are needed.

(b) For $b > 0$, show that $a//(-b) = -(a//b)$ and $a\%(-b) = a\%b$. [Don't let the notation make this hard!]

(c) Prove this alternative version of the QR Theorem. Let a and b be integers with $b \neq 0$. There exist unique integers q, r such that

(1) $a = bq + r$, and

(2) $|b|/2 < r \leq |b|/2$

[There are two approaches: use the existing QR Theorem to prove the alternative, or prove it from scratch by redefining S and modifying the proof of the QR Theorem.]

Let a and b be integers, at least one of which is not 0. The **common divisors** of a and b are the integers that divide both a and b. The **greatest common divisor** (gcd) is the largest positive integer dividing both a and b. The **common multiples** of a and b are the integers that are multiples of both a and b. The **least common multiple** (lcm) is the smallest positive integer that is a multiple of both a and b.

A **linear combination** of a and b is an integer that can be expressed as $au + bv$ for some integers u and v.

Exercises 1.1.3. Properties of divisibility.

(a) Show that if b divides a and d divides b that d also divides a.

(b) Show that if d divides a and d divides b that d also divides any linear combination of a and b. (In particular, this proves Lemma 1.1.5 below.)

The following result is an important property of the integers, and not an obvious one. It is an important tool in the study of groups. We will see echos of this result and the proof when we study ideals in rings (Section 4.4).

Theorem 1.1.4 (GCD). *Let a and b be integers, at least one of them nonzero. The gcd of a and b is the smallest positive linear combination of a and b. In particular,* $\gcd(a, b) = au + bv$ *for some integers u and v.*

The set of all linear combinations of a and b equals the set of multiples of $\gcd(a, b)$.

Proof. Let $S = \{ar + bs : r, s \in \mathbb{Z}\}$ be the set of all linear combinations of a and b. Let d be the smallest *positive* element of S and let u, v be such that $d = au + bv$. I claim d divides a and b.

By the QR Theorem applied to a and d, $a = dq + r$ for some integer q and nonnegative integer $r < d$. Then

$$r = a - dq = a - (au + bv)q = a(1 - uq) - bvq$$

This shows that r is also in S. But, d is the smallest positive element of S, and $0 \leq r < d$. Consequently, $r = 0$, so d divides a.

Similarly, one shows d divides b, so d is a common divisor of a and b. To show it is the greatest common divisor, let c be any other common divisor of a and b. Then c divides $au + bv = d$ (by divisibility properties). Since d is positive $c \leq d$.

Since d divides a and b the elements of S are all multiples of d by Exercise 1.1.3. On the other hand, any multiple of d is a linear combination of a and b since d is a linear combination of a and b. This establishes the last sentence of the theorem. □

We say two integers are **coprime** (or **relatively prime**) when their gcd is 1.

Given a and b, how do we find their gcd? The answer (for arbitrary large integers) is not to factor each and look for common factors. Rather, use the Euclidean algorithm.

Let's assume $a \geq b \geq 0$. Recall that $a//b$ is the integer quotient and $a\%b$ the remainder as determined by the QR Theorem. Set $r_{-1} = a$ and $r_0 = b$, and define inductively (while $r_k \neq 0$)

$$\begin{aligned} q_k &= r_{k-1}//r_k \\ r_{k+1} &= r_{k-1}\%r_k, \qquad \text{so that} \\ r_{k-1} &= r_k q_k + r_{k+1}. \end{aligned}$$

Rearranging these equalites by solving for r_{k+1}, we get a sequence

$$\begin{aligned}
r_1 &= r_{-1} - r_0 q_0 = a - b q_0 \\
r_2 &= r_0 - r_1 q_1 \\
r_3 &= r_1 - r_2 q_2 \\
&\vdots \\
r_{k+1} &= r_{k-1} - q_k r_k \\
&\vdots \\
r_{n+1} &= r_{n-1} - q_n r_n = 0.
\end{aligned}$$

The r_k are a strictly decreasing sequence of nonnegative integers, so the process must terminate: for some n, $r_{n+1} = 0$ so r_n divides r_{n-1}. Now we make use of the following lemma, proved using basic divisibility properties (see Exercise 1.1.3).

Lemma 1.1.5. *For integers a, b, c, s such that $a = bs + c$, we have* $\gcd(a, b) = \gcd(b, c)$.

Let's apply this to the sequence r_k, letting n be minimal such that $r_{n+1} = 0$. We have (since $r_{n+1} = 0$)

$$\gcd(a, b) = \gcd(b, r_1) = \cdots = \gcd(r_n, r_{n+1}) = \gcd(r_n, 0) = r_n$$

This argument shows that the Euclidean algorithm produces the gcd of a and b. In the following Sage code we only keep two of the remainders at any time, not the whole sequence: after the kth pass through the while loop, r in the algorithm is r_{k-1} and s is r_k.

```
def euclid_alg(a,b):
   if b == 0:
      print ("division by zero")
      return false
   else:
      r = a
      s = b
      while s != 0:
         rem =  r %s
         r = s
         s = rem
      return r
```

There is a heftier Euclidean algorithm—often called the Extended Euclidean algorithm—which produces two integers u, v such that $au + bv = \gcd(a, b)$. I like the following matrix version of the algorithm. Let

$$Q_k = \begin{bmatrix} -q_k & 1 \\ 1 & 0 \end{bmatrix} \qquad \text{and} \qquad R_k = \begin{bmatrix} r_k \\ r_{k-1} \end{bmatrix}$$

where the sequence r_k and q_k are the same as used above in the Euclidean algorithm. Verify that $R_{k+1} = Q_k R_k$. Consequently,

$$R_{n+1} = \begin{bmatrix} 0 \\ r_n \end{bmatrix} = Q_n Q_{n-1} \cdots Q_0 R_0 \tag{1.1}$$

where $R_0 = \begin{bmatrix} b \\ a \end{bmatrix}$. Let $M = \begin{bmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{bmatrix} = Q_n Q_{n-1} \cdots Q_0$. Then, after the algorithm terminates, we have $m_{2,1} b + m_{2,2} a = r_n = \gcd(a, b)$.

Here is Sage code for the extended Euclidean algorithm. (Note that Sage indexes rows and columns of matrices starting from 0 not 1.) Initially M is the 2×2 identity matrix and R is the matrix $\begin{bmatrix} a \\ b \end{bmatrix}$. The algorithm iteratively computes q (which, at the kth iteration is q_{k-1}), the quotient of $m_{1,0}$ by $m_{0,0}$. It forms the matrix $Q = \begin{bmatrix} -q & 1 \\ 1 & 0 \end{bmatrix}$ and multiplies both R and M by Q. The result after iteration k (for $k = 1, \dots$), is that R is R_k and M is the product $Q_{k-1} Q_{k-2} \cdots Q_1 Q_0$ in (1.1).

```
def ext_euclid_alg(a,b):
   if b == 0:
      print ("division by zero")
      return false
   else:
      M = matrix.identity(2)
      R = matrix(2,1, [b,a])
      while R[0,0] != 0:
         q =  R[1,0]//R[0,0]
         Q = matrix(2,2,[ -q , 1, 1, 0])
         M = Q * M
         R = Q * R
      return M
```

We have proven that the Euclidean algorithm terminates with the greatest common divisor of the input integers. A bit closer analysis of the algorithm reveals a connection with the Fibonacci numbers and allows one to bound the number of steps for the Euclidean algorithm. See [Ros11] for details.

Theorem 1.1.6 (Euclidean Algorithm)**.** *Let a, b be integers with $b \neq 0$. The Euclidean Algorithm outputs* $\gcd(a, b)$ *in at most* $1 + \log_2 b/\log \alpha$ *steps, where* $\alpha = (1 + \sqrt{5})/2$ *is the golden ratio. The Extended Euclidean Algorithm outputs integers* u, v *with* $|u| < b$ *and* $|v| < a$ *such that* $au + bv = \gcd(a, b)$*.*

The greatest common divisor of a finite set of integers (that contains a nonzero integer) is simply the largest integer that divides each element of the set. A simple induction argument shows that the set of common divisors of $\{a_1, \ldots, a_n\}$ is equal to the set of common divisors of $\{\gcd(a_1, \ldots, a_{n-1}), a_n\}$. To compute the greatest common divisor of this set efficiently, one computes iteratively $\gcd(a_1, \ldots, a_k) = \gcd(\gcd(a_1, \ldots, a_{k-1}), a_k)$. (There are moreoptimized algorithms, but understanding this approach is sufficient here.)

Exercises 1.1.7. Using the Euclidean algorithm

Express the greatest common divisor as a linear combination of the given integers.

(a) 89, 24

(b) 24, 40, 30

Exercises 1.1.8. Extensions of the GCD theorem

(a) Given any set of integers $\{a_1, \ldots, a_n\}$ their greatest common divisor may be expressed as a linear combination $d = u_1a_1 + \ldots u_na_n$ for $u_i \in \mathbb{Z}$. Prove this using induction.

(b) Given $a, b \in \mathbb{Z}$, characterize the set of $\{(u, v) \in \mathbb{Z} \times \mathbb{Z} : ua + vb = \gcd(a, b)\}$. That is, if $(\bar{u}, \bar{v})$ is one such pair, what are all the others? [It is easiest to do this first for a, b coprime. Then extend the result.]

(c) For the integers $a = 6$, $b = 10$, and $c = 15$ identify the set of all triples u, v, w such that $au + bv + cw = \gcd(a, b, c)$.

(d) (Much more challenging, and just to tantalize you.) Characterize the set of integers $u_1, \ldots, u_n$ such that $a_1u_1 + a_2u_2 + \cdots + a_nu_n = \gcd(a_1, \ldots, a_n)$.

Primes, Irreducibles, and Unique Factorization

The next main result is the Fundamental Theorem of Arithmetic, which says (roughly) that every nonzero integer has a unique factorization as a product of primes. We will now define what it means for an integer to be prime, but it will not be the school definition. We also define the term irreducible, which is what we customarily use for primality. The definitions given here are the accepted ones in more general contexts. Fortunately, the following theorem (Euclid's Lemma), which I will refer to as the Prime-Irreducible Theorem shows that for integers the notions are equivalent.

Definition 1.1.9. Let r be an integer with $|r| > 1$. We say r is **irreducible** when $r = ab$ implies that either $a = \pm 1$ or $b = \pm 1$ (and the other is $\pm r$). We say r is **prime** when $r|ab$ implies $r|a$ or $r|b$.

Theorem 1.1.10 (Prime-Irreducible)**.** *An integer is irreducible if and only if it is prime.*

Proof. We prove this for positive integers; minor adjustments can be made for a negative number.

Let $p > 1$ be an irreducible; let us show it is prime. Suppose that $p|ab$ for some integers a and b. We need to show $p|a$ or $p|b$. If p divides a we are done, so suppose it does not divide a. Since p is irreducible, its only positive divisors are 1 and p, so the GCD of a and p is 1. By the GCD Theorem, there are integers u, v such that

$$1 = au + pv$$

Multiplying by b

$$b = abu + pbv$$

Since $p|ab$ we have that p divides the right hand side. Thus p divides b, as was to be shown.

Suppose now that $p > 1$ is prime, we will show it is irreducible. Let $p = ab$ be a factorization of p. We must show one of a or b is ± 1. Since p is prime and it divides (in fact equals) the product ab it must divide one of the factors. Without loss of generality, say $p|a$, so $a = px$ for some integer x. Then $p = ab = pxb$, so $p(1 - xb) = 0$. Since $p \neq 0$, we have $1 - xb = 0$, so $x = b = \pm 1$. □

The previous theorem is the key ingredient to establishing unique factorization.

Theorem 1.1.11 (Unique Factorization)**.** *Let a be a positive integer $a > 1$. There is a nonnegative integer t, there are positive prime numbers $p_1 < p_2 < \cdots < p_t$, and there are positive integers $e_1, \ldots, e_t$, such that*

$$a = p_1^{e_1} \cdots p_t^{e_t}$$

This factorization of a is uniquely determined.

Proof. We prove existence first, then uniqueness, both by induction.

Clearly, any prime integer, such as 2 can be factored as claimed using $t = 1$ and $e_1 = 1$. Let us assume that all integers less than n can be factored as a product of primes, and prove that n can also be factored as a product of primes. If n is prime, we are done (again using $t = 1$, $e_t = 1$). Otherwise $n = ab$ with $a < n$ and $b < n$. By the induction hypothesis, each of a and b can be factored as a product of primes, let's say there are t distinct primes involved altogether in the factorizations of a and b. Allowing some of the exponents to be 0 we have

$$\begin{aligned} a &= p_1^{d_1} \cdots p_t^{d_t} \qquad \text{and} \\ b &= p_1^{e_1} \cdots p_t^{e_t} \qquad \text{so} \\ n = ab &= \left(p_1^{d_1} \cdots p_t^{d_t}\right)\left(p_1^{e_1} \cdots p_t^{e_t}\right) \\ n = ab &= p_1^{d_1+e_1} \cdots p_t^{d_t+e_t} \end{aligned}$$

To prove uniqueness, we again note that there is only one way to factor a prime number, since it is only divisible by itself and 1. This gives us our base step, $n = 2$. Assume uniqueness of the prime factorization for all integers less than n. Suppose n has two distinct factorizations.

$$n = p_1^{d_1} \cdots p_s^{d_s} = q_1^{e_1} \cdots q_t^{e_t}$$

with d_i, e_i positive. Since p_1 is prime and is a factor of n it must divide one of the factors of n in the q factorization; say p_1 divides q_j. But q_1 is the smallest prime among the q_i, so $q_1 \leq p_1$. An analogous argument shows $p_1 \leq q_1$, and consequently, $p_1 = q_i$. Dividing both sides by p_1 we get n/p_1 is an integer smaller than n. It therefore has a unique factorization. Consequently, $s = t$ and for all i, $p_i = q_i$ and $d_i = e_i$. □

Exercises 1.1.12. CONSEQUENCES OF UNIQUE FACTORIZATION

Prove these statements using the Unique Factorization theorem.

(a) Every nonzero integer can be uniquely expressed in the form $u2^eb$ in which $u = \pm 1$, $e \in \mathbb{N}_0$ and b is odd.

(b) Every nonzero rational number a can be uniquely expressed in the form

$$a = up_1^{e_1} \cdots p_t^{e_t}$$

for some $u = \pm 1$, $t \in \mathbb{N}_0$, prime numbers $p_1 < p_2 < \cdots < p_t$, and nonzero integers $e_1, \ldots, e_t$.

(c) Any rational number can be expressed in a unique way in the form a/b with $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ with $\gcd(a, b) = 1$. We call this expressing the rational number in **lowest terms**.

Exercises 1.1.13. EXTENDING GREATEST COMMON DIVISOR TO RATIONAL NUMBERS.

This will be easier to do using unique factorization than just the GCD theorem. Let $\{a_1/b_1, \ldots, a_n/b_n\}$ be a set of rational numbers in lowest terms. Let $B = \operatorname{lcm}\{b_1, \ldots, b_n\}$ and $A = \gcd\{a_1, \ldots, a_n\}$.

(a) Show that $Ba_i/(Ab_i)$ is an integer for all $i = 1, \ldots, n$.

(b) Show that $\gcd\{Ba_1/(Ab_1), \ldots, Ba_n/(Ab_n)\} = 1$. [Suppose p is a prime dividing B, show there is some i such that p is not a factor of $Ba_i/(Ab_i)$.]

(c) Observe that the rational number A/B is similar to the GCD. If we factor it out of the rational numbers a_i/b_i we are left with a set of integers that have no common factor.

Modular Arithmetic

Our discussion of the integers culminates with a quick summary of arithmetic modulo an integer n. This is a model for the construction of quotient groups and quotient rings that will be taken up later.

Definition 1.1.14. Let n be a nonzero integer. For integers a and b we say a is **congruent modulo** n to b when n divides $a - b$.

Theorem 1.1.15. *Congruence modulo n is an eqivalence relation. Furthermore, the set $\{0, 1, \ldots, n-1\}$ is a system of representatives for congruence modulo n in the sense that each integer a is congruent modulo n to exactly one element of $\{0, 1, \ldots, n-1\}$.*

Proof. The relation of being congruent modulo n is clearly reflexive, since for any $a \in \mathbb{Z}$, $n|(a-a)$. It is symmetric because if $n|(a-b)$ then also $n|(b-a)$. It is transitive because if a is congruent to b and b is congruent to c modulo n then $n|(a-b)$ and $n|(b-c)$. This implies that n divides the sum $(a-b)+(b-c) = a-c$, by Exercise 1.1.3, so a is congruent to c modulo n.

From the Quotient-Remainder Theorem, an integer a is congruent to its remainder when divided by n, since there is an integer q such that $a = nq + r$. This remainder is one of the elements of $\{0, 1, \ldots, n-1\}$. No two of these numbers differ by a multiple of n so they are distinct modulo n. □

The integers n and $-n$ give the same equivalence relation, so we always use positive integers for the modulus. It is common to write $[a]_n$ for the congruence class of a modulo n, whenever we need to be careful to distinguish between the integer a and the congruence class, or when we have more than one modulus to worry about. If there is a unique modulus the subscript n may be omitted. If it is clear from context that we are working modulo n, we may simply write a.

Finally we have:

Theorem 1.1.16 (Arithmetic modulo n)**.** *Suppose that $a \equiv b \bmod n$ and $r \equiv s \bmod n$. Then $a + r \equiv b + s \bmod n$ and $ar \equiv bs \bmod n$. Thus, arithmetic on congruence classes modulo n is well-defined as follows.*

- $[a] + [r] = [a + r]$
- $[a] * [r] = [ar]$

Proof. Suppose that $a \equiv b \bmod n$ and $r \equiv s \bmod n$. We have $a = b + jn$ for some integer j and $r = s + kn$ for some integer k. Then $a + r = b + s + (j + k)n$ so $a + r \equiv b + s \bmod n$. We also have $ar = bs + (ak + bj + jkn)n$ so $ar \equiv bs \bmod n$.

This shows that no matter what element of an congruence class is used to represent the class, arithmetic operations modulo n will give the same result. □

We will write $\mathbb{Z}/n$ for the set of congruence classes modulo n, with the operations $+$ and $*$. (It is also common to use $\mathbb{Z}_n$, but $\mathbb{Z}/n$ is consistent with notation we will use later.) When there is no chance of ambiguity, we write the congruence classes as $0, 1, \ldots, n - 1$ (without the brackets and using the least nonnegative representatives for each class). But, sometimes it is handy to be a bit flexible. For example it is good to remember that $n - b$ is equal to $-b$ in $\mathbb{Z}/n$. So (in $\mathbb{Z}/n$)

$$b(n-1) = (-1)b = n - b \text{ (computing in } \mathbb{Z}/n).$$

Exercises 1.1.17. Alternative representatives for elements of $\mathbb{Z}/n$

(a) Use the Quotient-Remainder Theorem from Exercises 1.1.2 to show alternative sets of representatives for the integers modulo n are:

$$-\frac{n-1}{2}, -\frac{n-3}{2}, \ldots, \frac{n-3}{2}, \frac{n-1}{2} \qquad \text{for } n \text{ odd, and,}$$
$$-\frac{n-2}{2}, -\frac{n-4}{2}, \ldots, \frac{n-4}{2}, \frac{n-2}{2}, \frac{n}{2} \qquad \text{for } n \text{ even.}$$

Exercises 1.1.18. Invertible elements in $\mathbb{Z}/n$

(a) Let p be a prime number. Let $[a] \in \mathbb{Z}/p$ with $[a] \neq [0]$ (so a is not divisible by p). Use the GCD Theorem 1.1.4 to show there is some $r \in \mathbb{Z}/p$ such that $[a][r] = [1]$.Consequently, each nonzero element of $\mathbb{Z}/p$ has a multiplicative inverse.

(b) Extend this result, partially, to $\mathbb{Z}/n$ for composite n. If $[a] \in \mathbb{Z}/n$ is such that the integer a is coprime to n, then there is some $[r] \in \mathbb{Z}/n$ such that $[a][r] = [1]$.

1.2 Groups, Rings, and Fields

Let us now introduce our three objects of study: groups, rings, and fields. This section will discuss some familiar number systems in the context of groups, rings, and fields. We include also some perspective on the history of the number systems as well as their appearance in our system of education.

Definition 1.2.1. A **binary operation** on a set S is a function from $S \times S$ to S. A **unary operation** on S is a function from S to S. An "operation" on S is usually assumed to be binary if not stated otherwise[1].

A binary operation $*$ on S is **associative** when $(a * b) * c = a * (b * c)$. It is **commutative** when $a * b = b * a$. It has an **identity element** when there is some element $e \in S$ such that $a * e = e * a = a$ for all $a \in S$.

A group has one binary operation, generally denoted $*$, while rings and fields have two binary operations, generally denoted $+$ and $*$.

Definition 1.2.2. A **group** is a set G with a binary operation $*$ and a unary operation, denoted $a \longmapsto a^{-1}$, satisfying the following properties.

(1) Associativity of $*$.

(2) Identity for $*$: There is an element, generally denoted e, such that $e * a = a = a * e$ for all $a \in G$.

(3) Inverses for $*$: For each $a \in G$ the unary operation $a \longmapsto a^{-1}$ gives the inverse for a. That is, $a * a^{-1} = e = a^{-1} * a$.

A group which also satisfies $a * b = b * a$ is called **commutative** or **abelian** (after the mathematician Abel).

Definition 1.2.3. A **ring** is a set R, with two operations $+$ and $*$ that satisfy the following properties.

(1) Associativity for both $+$ and $*$.

(2) Commutativity for both $+$ and $*$.

(3) Identity elements for both $+$ and $*$. There is some element in R, that we call 0, such that $a + 0 = a$ and there is an element, that we call 1, such that $a * 1 = a$.

[1] One can define ternary ($S \times S \times S \longrightarrow S$) and, more generally, n-ary operations, but we will have no use for these.

(4) Inverses for $+$. For each $a \in R$ there is some other element, which we write $-a$, such that $a + (-a) = 0$.

(5) Distributivity of $*$ over $+$. That is, $a * (b + c) = a * b + a * c$.

A **field** is a ring with one additional property,

(6) Inverses for $*$. For each nonzero $a \in R$ there is some other element, that we write a^{-1}, such that $a * a^{-1} = 1$.

Comparing these definitions, one sees that a ring R under the operation $+$ is an abelian (commutative) group with identity element 0 and additive inverse operation $a \longmapsto -a$. Under the operation of multiplication, $*$, a ring may lack the property of inversion. A field F is an abelian group under $+$ and the set of nonzero elements, $F^* = F \setminus \{0\}$, is an abelian group under $*$. The interaction between the two operations of $+$ and $*$ for both rings and fields is given by the distributive property.

Now to the question: what examples do we have of groups, rings, and fields?

Familiar Rings and Fields

The first number system that a child learns in school is the natural numbers $\mathbb{N} = \{1, 2, 3, 4, \dots\}$, and eventually this is expanded to the integers by including 0 and the additive inverse of each positive integer. The integers, denoted $\mathbb{Z}$, are an abelian group under addition. Once the operation of multiplication as repeated addition is introduced, we have the first example of a ring. The integers in fact form the prototypical ring, as we shall see in Theorem 4.2.11.

Students in elementary school—the lucky ones—may also learn "clock arithmetic" in which addition is done on a clock, so 8:00 plus 7 hours is 3:00. This is essentially modular arithmetic with modulus 12 (although we usually use representatives 1:00, 2:00, . . . , 12:00 rather than using 0:00, 1:00, . . . , 11:00). We saw in Section 1.1 that multiplication is also well defined modulo n, and one can check that the properties of a ring are satisfied. We will denote this number system $\mathbb{Z}/n$ (although $\mathbb{Z}_n$ is also commonly used).

The next step in mathematics education is to expand this integer number system. The integers do not form a field since the only numbers with a multiplicative inverse in $\mathbb{Z}$ are ± 1. There is a complicated process that enlarges the set of integers by adding fractions to create the rationals, $\mathbb{Q}$. I say the process is complicated because lots of people have trouble understanding fractions well, and a key part of the problem is that a given number has an infinite number of different names: $1/2 = 2/4 = 3/6 = \cdots$. The process of forming fractions can be generalized to other rings, but it has delicate and subtle steps involving equivalence relations.

When you see the construction in Section 4.7 you may appreciate that these subtleties are closely tied to the difficulties people have with fractions.

There are two other fields that are introduced in secondary school education, although they are challenging to understand fully: the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$. Formally defining the real numbers is a sophisticated process, but treating $\mathbb{R}$ as the set of all decimal numbers (including infinite non-repeating ones) and imagining numbers as points on the number line is a way to work with them effectively enough to do most college level mathematics.

We won't have much need for the real numbers, but the relationship between the reals and the complex numbers is something that is key to studying fields in general. The complex numbers are lightly treated in secondary school by "imagining" a number i whose square is -1. The complex numbers are those of the form $a + bi$ in which $a, b \in \mathbb{R}$. Addition is "componentwise," $(a+bi)+(c+di) = (a+c)+(b+d)i$, and the additive inverse of $a+bi$ is $-a-bi$. Multiplication is based on $i^2 = -1$: $(a+bi)*(c+di) = (ac-bd)+(ad+bc)i$. One can check that $(a-bi)/(a^2+b^2)$ is the multiplicative inverse of $a+bi$. With these operations, $\mathbb{C}$ is a field.

In the discussion of the complex numbers above, there was actually no need to use real numbers for a, b, c, d. We could have restricted them to be rational numbers and the statements about addition, multiplication and inverses would still hold true. Thus we can introduce a field derived from $\mathbb{Q}$ that includes i and uses the rules above for addition, multiplication, and the inverses for each. We call this field the Gaussian rationals and denote it $\mathbb{Q}(i)$.

There is one other field that is accessible to those who have learned "clock arithmetic," (essentially arithmetic modulo 12 as noted above). If our clock had a prime number p of positions, the arithmetic would be in $\mathbb{Z}/p$. It was shown in Exercise 1.1.18 that every nonzero element in $\mathbb{Z}/p$ has a multiplicative inverse, thus $\mathbb{Z}/p$ is a field. When we focus on this modular ring as a field we will write it as $\mathbb{F}_p$, instead of $\mathbb{Z}/p$. Thus we have our slate of elementary fields, $\mathbb{Q}$, $\mathbb{Q}(i)$ and $\mathbb{F}_p$, supplemented if we want by $\mathbb{R}$ and $\mathbb{C}$.

Each of these fields is of course a ring. Our collection of rings that are not fields includes $\mathbb{Z}$ and $\mathbb{Z}/n$ for n not a prime. We may supplement it by the Gaussian integers

$$\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$$

It is routine to check that this is a ring. Notice that there are 4 elements that have multiplicative inverses, $\pm 1, \pm i$.

All of our rings are groups under addition and we can study them as groups by "forgetting" the multiplicative structure ("forget" is actually a term used by mathematicians in this context!) We also obtain a few other examples of groups

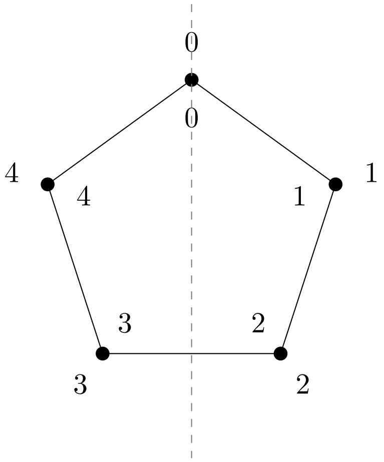


Figure 1.1: The Pentagon

by looking at the nonzero elements of a field under multiplication. These are the nonzero rationals, denoted $\mathbb{Q}^*$, the nonzero elements of $\mathbb{Q}(i)$, and the nonzero elements of $\mathbb{F}_p$, denoted $\mathbb{F}_p^*$ (and similarly for $\mathbb{R}$ and $\mathbb{C}$).

But what about a group that doesn't involve ignoring one of the operations on a ring? In particular, so far, we have no examples of groups in which the operation is not commutative.

The Dihedral Groups, D_n

Group theory actually arose from the study of transformations that preserve the structure of a mathematical objects. The symmetry of physical objects is perhaps the easiest entry point.

Consider a regular pentagon, as in Figure 1.1. Imagine a table with a pentagon carved into it and a clear pentagon that fits neatly into the enclosure. Enumerate the "base points" on the table and the vertices of the pentagon as shown.

Rotation counter-clockwise about the center by 72° takes the pentagon to itself. Only the enumeration of the vertices would indicate that a change occured. Calling this rotation r, we can see there are 5 rotational symmetries, which are rotation by $72°, 144°, 216°, 288°$ and $360°$. The latter has the same result as not moving the pentagon at all. This set of rotations is a group where the operation is just doing one rotation followed by another. Thinking of rotation as a function, we are composing functions. We may write these rotations as r, r^2, r^3, r^4 and r^5; the latter having the same effect as not moving pentagon at all, so $r^5 = r^0$ is the identity element. It should be clear that $r^i * r^j = r^{i+j}$. This group has 5 elements and "looks a lot like" $\mathbb{Z}/5$ under addition. (It is *isomorphic* to $\mathbb{Z}/5$, see Section 2.3.)

There is another type of symmetry indicated by the dashed line. For each vertex of the pentagon $0, \ldots, 4$ there is a line through that vertex and the midpoint of the opposite side that is an axis of symmetry for reflection. Let t_i be the reflection across the line at vertex i. We now have 10 symmetries of the pentagon: the identity, 4 non-trivial rotations, and 5 reflections. I claim these are the only symmetries. There are 5 possible places to put vertex 1; but then vertex 2 must be one notch away, either clockwise or counterclockwise. The positions of the other vertices are then determined by the rigidity of the pentagon. So there are only 10 possibilities. Notice also that after a rotation, the ordering of the pentagon vertices increases clockwise, but after a reflection the numbers of the pentagon vertices increase counterclockwise (and the numbers on the pentagon would be reversed as in a mirror).

Now let's consider the group operation: what happens when we follow one symmetry by another? To get started we must address some ambiguities in how we define rotation and reflection when the pentagon is not in the original position as in Figure 1.1. We adopt the following conventions:

- The rotation r is rotation of vertex i of the pentagon to the position of vertex $i-1$, so it is counter-clockwise when the enumeration on the pentagon increases clockwise and clockwise when the enumeration on the pentagon increases counter-clockwise (as it is after a reflection).

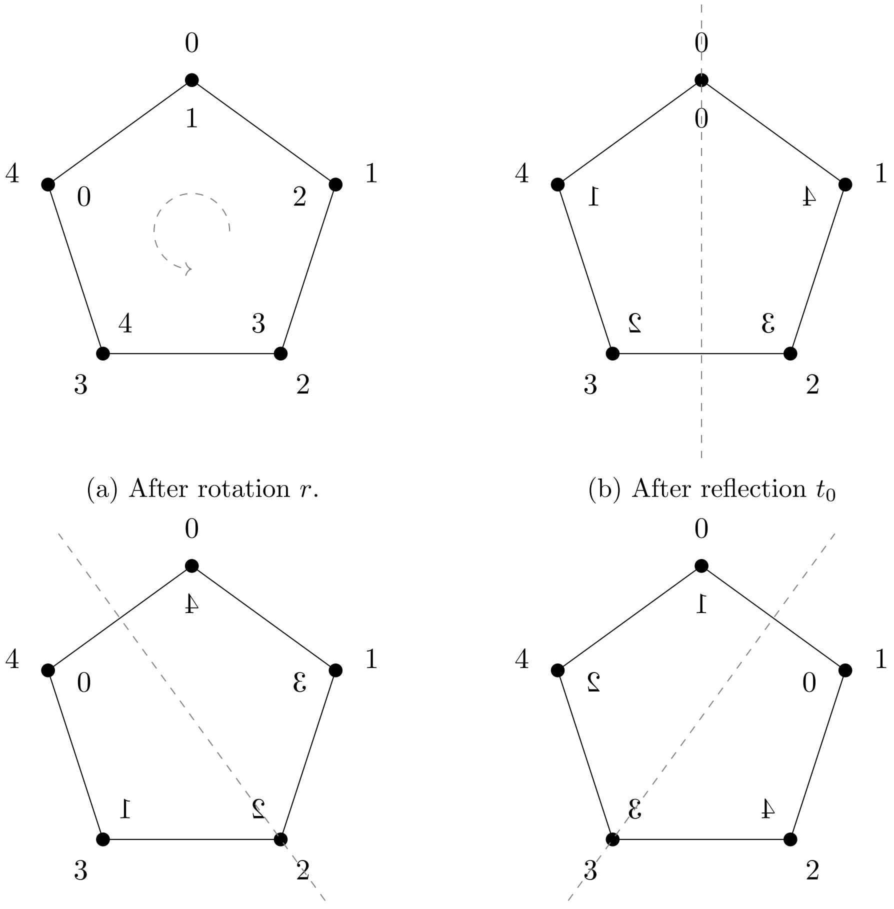
- A reflection t_i is reflection across the line through the ith vertex of the pentagon, not the label i on the table.

- The product $r*t_i$ means reflect then rotate. As is customary using functional notation, we apply the function on the right first.

Figure 1.2 shows the two computations, $t_0 * r = t_2$ and $r * t_0 = t_3$. These are unequal, so these computations show that we have our first example of a nonabelian group. It is called the dihedral group of order 5, and written D_5.

Definition 1.2.4. For $n \geq 3$, the **dihedral group** of order n, written D_n, is the group of symmetries of a regular n-gon. In addition to the identity, there are $n-1$ rotations and n reflections. The group operation is composition of the transformations.

One can verify, in a similar manner to that above, that D_n is nonabelian for all n. Our initial example could have been D_3 or D_4 rather than D_5. We did not choose D_3 because it is equal to the symmetric group, S_3 (see below), and we



(a) After rotation r.

(b) After reflection t_0

(c) After $t_0 r = t_2$.

(d) After $r t_0 = t_3$

Figure 1.2: The pentagon after various transformations

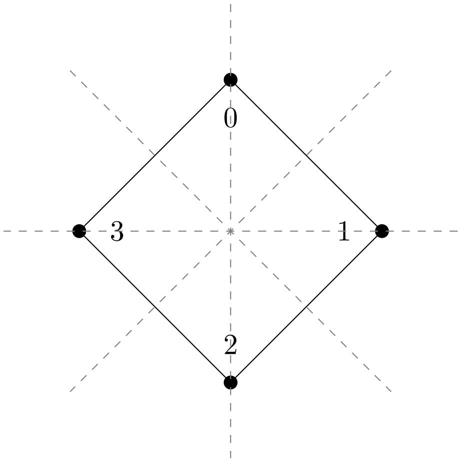


Figure 1.3: The Four Reflectional Symmetries of the Square

wanted the first example to distiguish between the symmetric and dihedral groups. We did not choose D_4 because there is a subtle complication with the notation when n is even. The reflections on D_4 (and for other even n) are not nicely indexed by the vertices. Some reflections go through a pair of vertices, the other reflections go through midpoints of opposite sides, as illustrated in Figure 1.3

We can consider each of the symmetries of the pentagon as a function on $\mathbb{Z}/5$ that assigns to the base point a on the table the index of the pentagon vertex at position a after applying the symmetry. So $r(a) = a + 1$, and $r^i(a) = a + i$ (computing modulo 5). The following exercise develops this example in more detail.

Exercises 1.2.5. Formulas for the products in D_5

We will use arithmetic in $\mathbb{Z}/5$ with the system of representatives $0, 1, 2, 3, 4$.

(a) Observe that the reflection t_i applied to the original position of the pentagon (in Figure 1.1) switches $i+1$ with $i-1$ and $i+2$ with $i-2$ where computations are modulo 5. Show that when t_i is applied to the original position of the pentagon the vertex at base point a is $2i - a$. We can write this as $t_i(a) = 2i - a$.

(b) Explain why the product of two reflections is a rotation, and find a formula for $t_i * t_j(a)$.

(c) Show that $r * t_i = t_{i+3}$ by arguing that rt_i is a reflection and that, applied to the pentagon in the original position, it takes $i + 3$ to itself.

(d) Find a formula for $r^j * t_i$; that is, give a function of $\mathbb{Z}/5$ [Hint: linear] for

$r^j * t_i(a)$. Do the same for $t_i * r^j$.

The Symmetric Groups, S_n

Recall that a function from one set to another is a bijection when it is both injective (one-to-one) and surjective (onto). If $f : A \longrightarrow B$ is a bijection, then there is a well-defined inverse function, f^{-1}, since each element of B has exactly one preimage. If $f : A \longrightarrow A$ is a bijection from A to itself then we say f **permutes** the elements of A; it rearranges them in a sense. We are particularly interested in the case when A is a finite set, even more specifically the set $\{1, 2, 3, \ldots, n\}$. In this case it is convenient to write a permutation as a table with the columns $i, f(i)$. Here are two examples (it is common to use Greek letters to denote permutations).

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \qquad\qquad \pi = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

Here $\sigma(1) = 1$, $\sigma(2) = 3$ and $\sigma(3) = 2$.

This tabular format makes it evident that there are $n!$ permutations of a set with n elements: There are n choices for the image of 1, call it $a_1 \in \{1, \ldots, n\}$, then there are $n-1$ possible images for 2, since it must be in $\{1, \ldots, n\} \setminus \{a_1\}$ and so on. The table would then be

$$\begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ a_1 & a_2 & a_3 & \ldots & a_n \end{pmatrix}$$

The tabular form indicates the sense in which a permutation is a rearrangement, with a_1 now being in the first position, a_2 in the second, and so on.

Since the composition of two bijections from A to itself is itself a bijection from A to itself, composition is an operation on the set of permutations. The inverse of a permutation is also a permutation. Thus we can make the following definition.

Definition 1.2.6. Let n be a positive integer. The set of all permutations of $\{1, 2, 3, \ldots, n\}$ along with composition and the unary operation that takes a permutation to its inverse function is called the **symmetric group** on $\{1, \ldots, n\}$ and is denoted S_n. We will write composition of permutations using $\circ$ when emphasizing that permutations are functions, but generally we use $*$, which is the usual notation for a product in groups.

For $n = 1$ the symmetric group has just 1 element, and for $n = 2$ it has two. The group S_3 has 6 elements. Three of these elements fix exactly one element, as σ does, and are called transpositions. The other two are π and $\pi^{-1} = \pi \circ \pi$. These are called 3-cycles. The symmetric groups S_n for $n \geq 3$ are nonabelian. One can compute $\pi \circ \sigma$ and $\sigma \circ \pi$ and see that they are unequal (applying the right hand function first is our convention).

$$\pi \circ \sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \qquad \qquad \sigma \circ \pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

The symmetric groups will be discussed in detail in Section 2.5.

An astute reader has perhaps noticed that the discussion of D_5 yielded a permutation of $\{0, 1, 2, 3, 4\}$ for each element of D_5. Allowing ourselves to let S_5 be the permutation group of $\{0, 1, \ldots, 4\}$ for the moment, we have essentially given a function of D_5 into S_5. Using a bit of Exercise 1.2.5 (and computing in $\mathbb{Z}/5$) we have;

$$\begin{aligned} r^j &\longmapsto \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ j & 1+j & 2+j & 3+j & 4+j \end{pmatrix} \\ t_i &\longmapsto \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 2i & 2i-1 & 2i-2 & 2i-3 & 2i-4 \end{pmatrix}. \end{aligned}$$

One can check that composition of the linear functions from Exercise 1.2.5 agree with the composition of the permutations; we are just composing functions.

In the terminology of the next chapter, we have given a homomorphism (Section 2.3) from D_5 to S_5 and the image is a subgroup of S_5 (Section 2.1).

Symmetry of Other Objects

The aesthetic appeal of symmetrical objects seems to be universal in human culture. The following images, some purely geometric, and some from artwork of various civilizations, show how rich the notion of symmetry can be. How does one describe the symmetry group of each of these objects? How does one understand the structure of these groups? These are questions that I hope will motivate the next two chapters.

A discussion of frieze patterns

Several examples of frieze patterns

Wikipedia article on symmetric tilings of the plane.

Wikipedia article on symmetric tilings of other the sphere and hyperbolic plane

Alhambra mosaics

A source for lots of art work with symmetry from ancient and modern times.

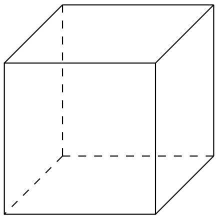
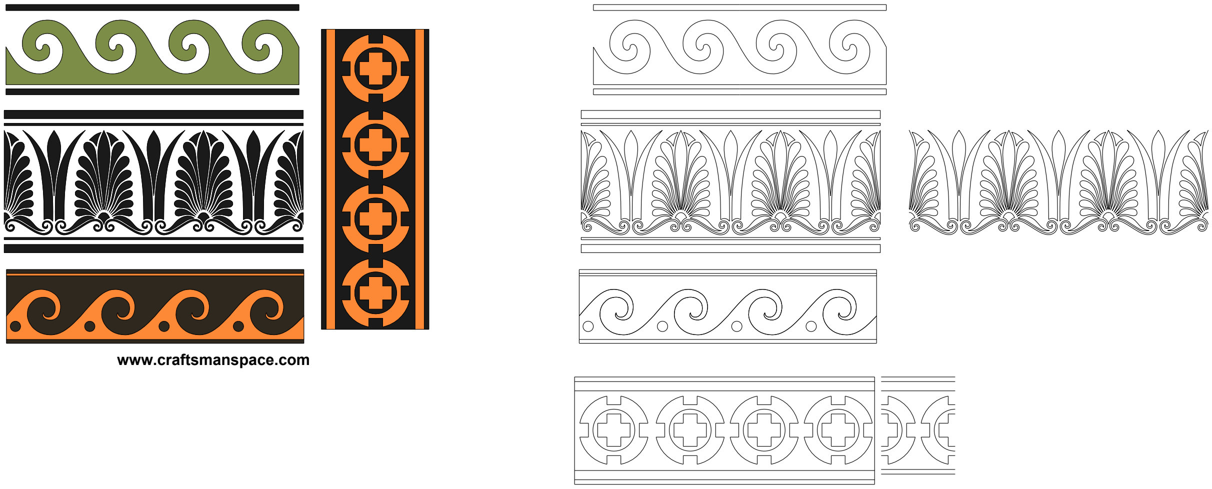
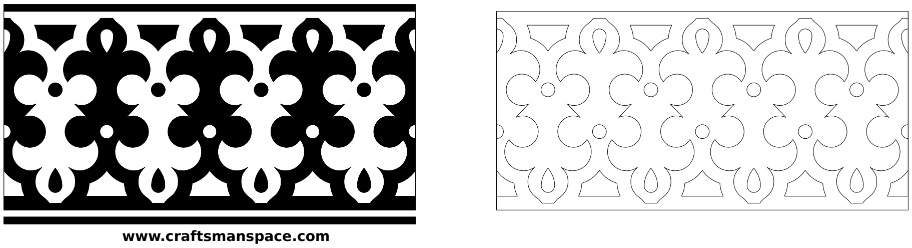


Figure 1.4: The cube has rotational symmetry along three axes: each going through the center of opposite faces.



www.craftsmanspace.com

Figure 1.5: Some Frieze Drawings: Imagine these extending infinitely in both directions



www.craftsmanspace.com

Figure 1.6: More Frieze Drawings: Imagine these extending infinitely in both directions

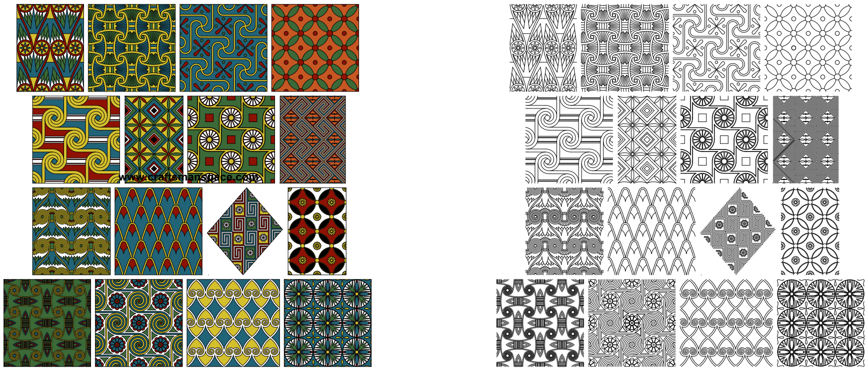


Figure 1.7: Some Mosaic Drawings: Imagine these extending infinitely to cover the plane

1.3 The Univariate Polynomial Ring over a Field

Let F be a field, in particular, we may consider F to be $\mathbb{Q}$, $\mathbb{F}_p$, or any of the other fields discussed in the Section 1.1. By $F[x]$ we mean the polynomial ring in the indeterminate x. The key theme of this section is that everything is that all the theorems we discussed for integers in Section 1.1 also hold—with appropriate modifications—for the polynomial ring over F. These are the Quotient-Remainder (QR) Theorem, the Greatest Common Divisor (GCD) Theorem, the Euclidean Algorithm, the Prime-Irreducible Theorem, and the Unique Factorization Theorem. This close relationship between $\mathbb{Z}$ and $F[x]$ is such an important theme in algebra and number theory that I want to lay out the fundamentals in detail in this section, which parallels substantially Section 1.1.

We can think of the polynomial ring consisting of two binary operations.

- It is a vector space over F with an infinite basis $1, x, x^2, \ldots$, and componentwise addition;

- It has a multiplicative structure defined by $x^i * x^j = x^{i+j}$ and the properties of commutativity, associativity and distributivity of multiplication over addition.

The result is thus a ring.

Sometimes it is useful to write a polynomial $b(x) \in F[x]$ as a sum $b(x) = \sum_{i=0}^{\infty} b_i x^i$ with the understanding that only a finite number of the b_i are nonzero. When all of the $b_i = 0$ we get the **zero polynomial**. Suppose $b(x) \neq 0$ and let δ be the largest integer such that $b_\delta \neq 0$. We call δ the **degree** of $b(x)$; $b_\delta x^\delta$ is the **leading term** of $b(x)$; x^δ is the **leading monomial** of $b(x)$; and b_δ is the

leading coefficient of $b(x)$. If $b_\delta = 1$ we say $b(x)$ is **monic**. When $\delta = 0$ we say $b(x)$ is a **constant** polynomial. The zero polynomial is also considered a constant polynomial and the degree is sometimes defined to be $-\infty$.

The product of a polynomial $a(x)$ of degree γ and $b(x)$ of degree δ has degree $\gamma + \delta$. Rules for divisibility of polynomials are similar to those for the integers. In particular if $a(x)$ divides $b(x)$ and $b(x)$ divides $c(x)$ then $a(x)$ divides $c(x)$. Furthermore if $d(x)$ divides both $a(x)$ and $b(x)$ then it divides their sum (and also any multiple of either $a(x)$ or $b(x)$).

The Quotient-Remainder Theorem and Divisibility

The following lemma simplifies the proof of the Quotient-Remainder Theorem. It is worth remarking that we are using the properties of a field when we divide by b_γ.

Lemma 1.3.1 (Division)**.** *Let $a(x)$ and $b(x)$ be in $F[x]$ with degrees γ and δ respectively and $\gamma > \delta$. Then the degree of $a(x) - \frac{a_\gamma}{b_\delta} x^{\gamma-\delta} b(x)$ is less than γ.*

Proof. This is a straightforward computation.

$$\begin{aligned}
a(x) &- \frac{a_\gamma}{b_\delta} x^{\gamma-\delta} b(x) \\
&= a_\gamma x^\gamma + a_{\gamma-1} x^{\gamma-1} + \cdots + a_1 x + a_0 \\
&\qquad - \frac{a_\gamma}{b_\delta} x^{\gamma-\delta} \Big(b_\delta x^\delta + b_{\delta-1} x^{\delta-1} + \cdots + b_1 x + b_0 \Big) \\
&= a_\gamma x^\gamma + a_{\gamma-1} x^{\gamma-1} + \cdots + a_1 x + a_0 \\
&\qquad - \Big(a_\gamma x^\gamma + \frac{a_\gamma b_{\delta-1}}{b_\delta} x^{\gamma-1} + \cdots + \frac{a_\gamma b_1}{b_\delta} x^{\gamma-\delta+1} + \frac{a_\gamma b_0}{b_\delta} x^{\gamma-\delta} \Big) \\
&= \Big(a_{\gamma-1} - \frac{a_\gamma b_{\delta-1}}{b_\delta} \Big) x^{\gamma-1} + \Big(a_{\gamma-2} - \frac{a_\gamma b_{\delta-2}}{b_\delta} \Big) x^{\gamma-2} + \cdots
\end{aligned}$$

The leading terms of $a(x)$ and $\frac{a_\gamma}{b_\delta} x^{\gamma-\delta} b(x)$ cancel and the result has degree less than γ. □

Theorem 1.3.2 (Quotient-Remainder)**.** *Let $a(x)$ and $b(x)$ be elements of $F[x]$ with $b(x) \neq 0$. There exist unique $q(x)$, $r(x)$ such that*

(1) $a(x) = b(x)q(x) + r(x)$, *and*

(2) $\deg r(x) < \deg b(x)$.

Proof. Consider the set $S = \{a(x) - b(x)c(x) : c(x) \in F[x]\}$. The set of degrees of the elements of S is a nonempty subset of the nonnegative integers, so it has a least element, δ. There is some polynomial of degree δ in S, call it $r(x)$, and suppose $r(x) = a(x) - b(x)q(x)$. I claim $\deg(r(x)) < \deg(b(x))$. Suppose not. Let $\gamma = \deg(r(x))$ and $\delta = \deg(b(x))$. Apply Lemma 1.3.1 to $r(x)$ and $b(x)$ to get

$$\begin{aligned} r(x) - \frac{r_\delta}{b_\gamma} x^{\gamma-\delta} b(x) &= a(x) - b(x)q(x) - \frac{r_\delta}{b_\gamma} x^{\gamma-\delta} b(x) \\ &= a(x) - b(x)(q(x) + \frac{r_\delta}{b_\gamma} x^{\gamma-\delta}). \end{aligned}$$

This is also in S and by Lemma 1.3.1 has lower degree than $r(x)$. This contradicts our choice of δ as the lowest degree of elements in S. Consequently, we must have $\deg(r(x)) < \deg(b(x))$. This establishes existence of $q(x)$ and $r(x)$ as claimed.

To prove uniqueness, suppose another $r'(x), q'(x)$ satisfy (1) and (2). (We want to show they are equal to $r(x)$ and $q(x)$!) Then

$$\begin{aligned} a(x) = b(x)q(x) + r(x) &= b(x)q'(x) + r'(x) \quad \text{so} \\ r(x) - r'(x) &= b(x)\big(q'(x) - q(x)\big). \end{aligned}$$

The degree on the left hand side is strictly less than the degree of $b(x)$. Since the right hand side is a multiple of $b(x)$, it must in fact be 0. Thus $r(x) = r'(x)$ and $q(x) = q'(x)$. □

The greatest common divisor of two integers was easy to define since the integers are well ordered. It is not obvious that, among the common divisors of $a(x), b(x) \in F[x]$, there is just one monic divisor of maximal degree. The following theorem shows the gcd can be uniquely defined and extends the GCD Theorem to polynomials. A **polynomial combination** of $a(x)$ and $b(x)$ is a polynomial that can be expressed as $a(x)u(x) + b(x)v(x)$ for some $u(x), v(x) \in F[x]$.

Theorem 1.3.3 (GCD)**.** *Let $a(x), b(x) \in F[x]$ with at least one of them nonzero. There is a unique polynomial $d(x)$ satisfying*

(1) $d(x)$ is a common divisor of $a(x)$ and $b(x)$,

(2) $d(x)$ is monic,

(3) $d(x)$ is divisible by all other common divisors of $a(x)$ and $b(x)$ (so it is the greatest common divisor).

Furthermore, there exist $u(x), v(x) \in F[x]$ such that $d(x) = a(x)u(x) + b(x)v(x)$. The set of all polynomial combinations of $a(x)$ and $b(x)$ equals the set of multiples of $d(x)$.

Proof. Let $S = \{a(x)s(x) + b(x)t(x) : s(x), t(x) \in F[x]\}$ be the set of all polynomial combinations of $a(x)$ and $b(x)$. Let $d(x)$ be a nonzero polynomial of minimal degree in S and let $u(x), v(x)$ be such that $d(x) = a(x)u(x) + b(x)v(x)$. We may assume that $d(x)$ is monic, since any constant multiple of a polynomial in S is also in S. I claim $d(x)$ divides $a(x)$ and $b(x)$.

By the QR Theorem applied to $a(x)$ and $d(x)$, $a(x) = d(x)q(x) + r(x)$ for some $q(x)$ and $r(x)$ in $F[x]$ with $\deg(r(x)) < \deg(d(x))$. Then

$$\begin{aligned} r(x) &= a(x) - d(x)q(x) \\ &= a(x) - \big(a(x)u(x) + b(x)v(x)\big)q(x) \\ &= a(x)\big(1 - u(x)q(x)\big) - b(x)v(x)q(x) \end{aligned}$$

This shows that $r(x)$ is also in S. If it were nonzero, it could be multiplied by a constant to get a monic element of S with lower degree than $d(x)$, which contradicts the choice of $d(x)$. We can thus conclude that $r(x) = 0$, and consequently $d(x)$ divides $a(x)$.

Similarly, one shows $d(x)$ divides $b(x)$. Thus $d(x)$ is a common divisor of $a(x)$ and $b(x)$. To show it is the greatest common divisor, let $c(x)$ be any other common divisor of $a(x)$ and $b(x)$. Then $c(x)$ divides $a(x)u(x) + b(x)v(x) = d(x)$ (by divisibility properties) as claimed.

Finally, we note that the set S (which we defined to be the set of polynomial combinations of $a(x)$) is also the set of multiples of $d(x)$. Since $d(x)$ divides each of $a(x)$ and $b(x)$ it will divide any polynomial combination of $a(x)$ and $b(x)$ by divisibility properties. On the other hand, since $d(x)$ is a polynomial combination of $a(x)$ and $b(x)$, any multiple of $d(x)$ is also a polynomial combination of $a(x)$ and $b(x)$ and therefore in S. □

The proof of the theorem can be adapted for any set of polynomials $P \subseteq F[x]$. One can show that the smallest degree monic polynomial that can be expressed as a combination of the elements of P actually divides all the elements of P.

The discussion of the Euclidean algorithm for integers carries over almost verbatim to $F[x]$. We use $a(x)//b(x)$ for the polynomial quotient and $a(x)\%b(x)$ for the remainder of division of $a(x)$ by $b(x)$.

Assume $\deg(a(x)) \geq \deg(b(x))$. Set $r_{-1}(x) = a(x)$ and $r_0(x) = b(x)$, and define inductively (while $r_k(x) \neq 0$)

$$\begin{aligned} q_k(x) &= r_{k-1}(x)//r_k(x) \\ r_{k+1}(x) &= r_{k-1}(x)\%r_k(x), \qquad \text{so that} \\ r_{k-1}(x) &= r_k(x)q_k(x) + r_{k+1}(x). \end{aligned}$$

Rearranging these equalites by solving for $r_{k+1}(x)$, we get a sequence

$$\begin{aligned}
r_1(x) &= r_{-1}(x) - r_0(x)q_0(x) = a(x) - b(x)q_0(x) \\
r_2(x) &= r_0(x) - r_1(x)q_1(x) \\
r_3(x) &= r_1(x) - r_2(x)q_2(x) \\
&\vdots \\
r_{k+1}(x) &= r_{k-1}(x) - q_k(x)r_k(x) \\
&\vdots \\
r_{n+1}(x) &= r_{n-1}(x) - q_n(x)r_n(x) = 0.
\end{aligned}$$

We note that $\deg(r_k(x))$ is a strictly decreasing sequence of nonnegative integers. The process must terminate: for some n, $r_{n+1}(x) = 0$. Now we make use of the following lemma, proved using basic divisibility properties.

Lemma 1.3.4. *For polynomials $a(x), b(x), c(x), s(x)$ in $F[x]$ that satisfy $a(x) = b(x)s(x) + c(x)$, we have $\gcd(a(x), b(x)) = \gcd(b(x), c(x))$.*

Let's apply this to the sequence $r_k(x)$, letting n be minimal such that $r_{n+1}(x) = 0$. We have (since $r_{n+1}(x) = 0$)

$$\gcd(a(x), b(x)) = \gcd(b(x), r_1(x)) = \cdots = \gcd(r_n(x), r_{n+1}(x)) = \gcd(r_n(x), 0) = r_n(x)$$

This argument shows that the Euclidean algorithm produces the gcd of $a(x)$ and $b(x)$.

Exercises 1.3.5. Using the Euclidean algorithm

Express the greatest common divisor of these two polynomials as a polynomial combination of them.

(a) $f = x^4 + x^2$ and $g = x^3 + 1$.

(b) $f = x^6 + 1$ and $g = x^4 + x^3 + x^2 + 1$ as elements of $\mathbb{F}_2[x]$.

Primes, Irreducibles and Unique Factorization

Let $r(x) \in F[x]$ have degree at least one. As with integers, we say $r(x)$ is **irreducible** when it can't be factored in a nontrivial way: whenever $r(x) = a(x)b(x)$ either $a(x)$ or $b(x)$ is a constant (that is, in F). We say $r(x)$ is **prime** when $r(x)|a(x)b(x)$ implies $r(x)|a(x)$ or $r(x)|b(x)$.

As with integers, we have the equivalence of primality and irreducibility, which is a key step towards proving unique factorization.

Theorem 1.3.6 (Prime-Irreducible). *Any nonconstant element of $F[x]$ is irreducible if and only if it is prime.*

Proof. Let $r(x) \in F[x]$ be irreducible; let us show it is prime. Suppose that $p(x)|a(x)b(x)$ for some $a(x)$ and $b(x)$ in $F[x]$. We need to show $p(x)|a(x)$ or $p(x)|b(x)$. If $p(x)$ divides $a(x)$ we are done, so suppose it does not divide $a(x)$. Since $p(x)$ is irreducible, the GCD of $a(x)$ and $p(x)$ is 1. By the GCD Theorem, there are polynomials $u(x), v(x)$ such that

$$1 = a(x)u(x) + p(x)v(x)$$

Multiplying by $b(x)$

$$b(x) = a(x)b(x)u(x) + p(x)b(x)v(x)$$

Since $p(x)|a(x)b(x)$ we have that $p(x)$ divides the right hand side, and consequently $p(x)$ divides $b(x)$. Thus we have shown that $p(x)$ is prime.

Suppose now that $p(x)$ is prime; we will show it is irreducible. Let $p(x) = a(x)b(x)$ be a factorization of $p(x)$. We must show one of $a(x)$ or $b(x)$ is a constant. Since $p(x)$ is prime and it divides (in fact equals) the product $a(x)b(x)$ it must divide one of the factors. Without loss of generality, say $p(x)|a(x)$. We then have $\deg(a(x)) \geq \deg(p(x))$. On the other hand, since $p(x) = a(x)b(x)$ we have by additivity of degrees for a product of polynomials,

$$\deg(p(x)) = \deg(a(x)) + \deg(b(x)) \geq \deg(a(x))$$

We conclude that $\deg(p(x)) = \deg(a(x))$ and $\deg(b(x)) = 0$. Thus $b(x)$ is a constant.

□

Finally, we can establish uniqueness of factorization. The proof differs slightly from the proof for integers.

Theorem 1.3.7 (Unique Factorization). *Let $a(x) \in F[x]$ be nonzero. There is a nonnegative integer t, a constant $u \in F$, distinct monic irreducible polynomials $p_1(x), \ldots, p_t(x)$, and positive integers $e_1, \ldots, e_t$ such that*

$$a(x) = u\big(p_1(x)\big)^{e_1} \cdots \big(p_t(x)\big)^{e_t}$$

Each of t, $p_i(x)$, e_i and u is uniquely determined, up to reordering of the $p_i(x)^{e_i}$.

Proof. We prove existence first, then uniqueness, both by induction on the degree of $a(x)$. Clearly, a polynomial of degree one cannot be factored as a product of polynomials of positive degree, since the degree of a product is the sum of the degrees of the factors. Thus a polynomial of degree one is irreducible and has a unique factorization $a_1x + a_0 = a_1(x + a_0/a_1)$. This gives the base step for both existence and uniqueness.

Let us assume that all polynomials of degree less than n can be factored as a product of a constant and monic irreducibles, and prove that any polynomial $a(x)$ of degree n can also be factored in this way. If $a(x)$ is irreducible, we are done (again using $t = 1$, $e_t = 1$ and factoring out the leading coefficient). Otherwise $a(x) = b(x)c(x)$ with both $b(x)$ and $c(x)$ having degree less than n. By the induction hypothesis, each of $b(x)$ and $c(x)$ can be factored as a product of irreducibles, let's say there are t distinct irreducible polynomials involved altogether in factorizations of $b(x)$ and $c(x)$. Allowing some of the exponents to be 0 we have

$$\begin{aligned}
b &= p_1(x)^{d_1} \cdots p_t(x)^{d_t} \qquad \text{and} \\
c &= p_1(x)^{e_1} \cdots p_t(x)^{e_t} \qquad \text{so} \\
a(x) &= b(x)c(x) = \left(p_1(x)^{d_1} \cdots p_t(x)^{d_t}\right)\left(p_1(x)^{e_1} \cdots p_t(x)^{e_t}\right) \\
a(x) &= b(x)c(x) = p_1(x)^{d_1+e_1} \cdots p_t(x)^{d_t+e_t}
\end{aligned}$$

To prove uniqueness, suppose $a(x)$ has two distinct factorizations.

$$a(x) = up_1(x)^{d_1} \cdots p_s(x)^{d_s} = vq_1(x)^{e_1} \cdots q_t(x)^{e_t}$$

with d_i, e_i positive. Then u and v are both the leading term of $a(x)$, so they are equal. Since $p_1(x)$ is irreducible, it is prime, and since it is a factor of $a(x)$ it must divide one of the factors of $a(x)$ in the q factorization; say $p_1(x)$ divides $q_j(x)$. Since $q_j(x)$ is irreducible and monic its only monic factor is itself. Thus $p_1(x) = q_j(x)$. Dividing both sides by $p_1(x)$ we get $a(x)/p_1(x)$ is a polynomial of degree less than n. It therefore has a unique factorization. Consequently, $s = t$ and for all i, $p_i = q_i$ and $d_i = e_i$. □

Polynomial modulus

We now extend the technique of modular arithmetic to the polynomial ring over a field.

Let $m(x) \in F[x]$ have degree d. Polynomials $a(x)$ and $b(x)$ are **congruent modulo** $m(x)$ when $m(x)$ divides $a(x) - b(x)$.

Theorem 1.3.8. *Congruence modulo $m(x)$ is an equivalence relation. The set of polynomials of degree less than $\delta = \deg(m(x))$ is a system of representatives for congruence modulo $m(x)$. That is, each polynomial is congruent modulo $m(x)$ to its remainder when divided by $m(x)$, which has degree less than δ.*

Any constant multiple of $m(x)$ will define the same equivalence relation as $m(x)$ so we usually take $m(x)$ to be monic. We will write $[a(x)]_{m(x)}$ for the congruence class of $a(x)$ modulo $m(x)$ whenever we need to be careful to distinguish between $a(x)$, otherwise we will omit the subscript if the modulus is obvious.

Theorem 1.3.9 (Arithmetic modulo $m(x)$)**.** *Suppose that* $a(x) \equiv b(x) \bmod m(x)$ *and* $r(x) \equiv s(x) \bmod m(x)$. *Then* $a(x) + r(x) \equiv b(x) + s(x) \bmod m(x)$ *and* $a(x)r(x) = b(x)s(x) \bmod m(x)$. *Thus, arithmetic on congruence classes modulo* $m(x)$ *is well-defined.*

- $[a(x)] + [r(x)] = [a(x) + r(x)]$
- $[a(x)] * [r(x)] = [a(x)r(x)]$

Proof. Suppose that $a(x) \equiv b(x) \bmod m(x)$ and $r(x) \equiv s(x) \bmod m(x)$. We have $a(x) = b(x) + u(x)m(x)$ and $r(x) = s(x) + v(x)m(x)$ for some polynomials $u(x)$ and $v(x)$. Then

$$\begin{aligned} &a(x) + r(x) = b(x) + s(x) + \big(u(x) + v(x)\big)m(x), \qquad \text{so} \\ &a(x) + r(x) \equiv b(x) + s(x) \bmod m(x) \end{aligned}$$

We also have

$$a(x)r(x) = b(x)s(x) + \Big(a(x)v(x) + b(x)u(x) + u(x)v(x)m(x)\Big)m(x)$$

so $a(x)r(x) \equiv b(x)s(x) \bmod m(x)$.

This shows that no matter what element of a congruence class is used to represent a class, arithmetic operations modulo $m(x)$ will give the same result. □

We will write $F[x]/m(x)$ for the set of equivalence classes modulo $m(x)$, with the operations $+$ and $*$ as designated above.

Exercises 1.3.10. Invertible elements in $F[x]/mx)$

(a) Let $m(x)$ be an irreducible monic polynomial in $F[x]$. Let $[a(x)] \in F[x]/m(x)$ with $a(x)$ not divisible by $m(x)$. Use the GCD Theorem 1.3.3 to show there is some $r(x) \in F[x]$ such that $[a(x)][r(x)] = [1]$.

(b) Extend this result, partially, to other $F[x]/m(x)$. If $a(x) \in F[x]$ is such that $a(x)$ is coprime to $m(x)$, then there is some $r(x) \in F[x]$ such that $[a(x)][r(x)] = [1]$.

1.4 A Roadmap for this Book

Main themes

We have introduced three *categories* of algebraic objects: groups, rings, and fields. We have a few examples of each type. In the rest of the book we will study some fundamental questions for each of these categories:

- What are some characteristic first examples?
- How can we construct new (potentially more complex) objects from simpler ones?
- What types of functions are there between objects of the category (from groups to groups, or rings to rings) that *respect the architecture* of that category (the operations, identities, and inverses)? These functions are called homomorphisms.
- When are two objects essentially the same? (When there is a bijective homomorphism between them.)
- Can we break an arbitrary object in the category into constituent parts that are easier to understand? (This is the reverse question to the one about constructing new more complicated objects from simpler objects.)

For Groups

- Classify finite, and finitely generated groups.
- Outline the challenges and what is known about classifying all finite groups.
- Expand on the discussion of symmetry with group actions
- Apply group actions to understanding the structure of a finite group using the Sylow theorems.

For Rings

- Explore constructions, in particular polynomial rings and rings of fractions.
- Explore generalizations of the theorems in Sections 1.1 and 1.3 to other rings, particularly polynomial rings over a field in several indeterminates. What can we say about unique factorization? About division? About the GCD theorem (more specifically, the linear combination used to prove the GCD theorem for integers).

For Fields

- Describe the construction and the structure of all finite fields. Expanding on the observations previously that $\mathbb{Z}/p$ is a field for p prime and that $F[x]/m(x)$ is a field when $m(x)$ is irreducible.
- Describe number fields, these are fields that, like $\mathbb{Q}(i)$ are derived from the rationals using irreducible polynomials over $\mathbb{Q}$.
- Show the relationship between group theory and field theory developed by Galois.

Chapter 2

Groups

2.1 Groups and Subgroups

The material in this section is a quick summary of the most fundamental properties of groups. I have omitted some proofs because they are are fairly routine, are good exercises for the reader, and are available from many sources. See in particular the book of Hungerford [Hun12]. It is worthwhile reviewing the proofs as you read!

First let us recall the definition of a group.

Definition 2.1.1. A **group** is a set G with a binary operation $*$ and a unary operation denoted $a \longmapsto a^{-1}$ satisfying the following properties.

(1) Associativity of $*$: For all $a, b, c \in G$, $(a * b) * c = a * (b * c)$.

(2) Identity for $*$: There is an element in G, usually denoted e, such that $e * a = a = a * e$ for all $a \in G$.

(3) Inverses for $*$: For each $a \in G$ the unary operation $a \longmapsto a^{-1}$ gives the multiplicative inverse for a. That is, $a * a^{-1} = e = a^{-1} * a$.

A group which also satisfies $a * b = b * a$ is called **commutative** or **abelian** (after the mathematician Abel).

The operation is usually called a product. So $a * b$ is the product of a and b. One must be careful, because, in a nonabelian group, the products $a * b$ and $b * a$ are not necessarily equal. In abelian groups the operation is often called addition and is written with a $+$ sign, while the identity is written as 0.

The most basic properties are contained in the following proposition. The proofs of all of these are called “card tricks” by a friend of mine. Any algebraist should have these up a sleeve since similar cleverness is used in other contexts.

Proposition 2.1.2 (Basic Properties)**.** *Let $G, *$ be a group. Then*

(1) The identity element is unique.

(2) The inverse of any element is unique.

(3) The cancellation law holds: $a * b = a * c$ implies $b = c$ (and similarly for cancellation on the right).

(4) If $a * g = g$ for some $g \in G$, then $a = e_G$.

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

(6) $(a^{-1})^{-1} = a$.

Let G be a group. When there is risk of confusion, with more than one group under consideration, we will use $*_G$ for the operation on the group G and e_G for the identity element. Otherwise we will not subscript with G. In fact, unless there is some reason to be very clear we rarely write the group operation: $g_1 g_2$ means $g_1 *_G g_2$. For a positive integer n, g^n is shorthand for $\underbrace{gg \cdots g}_{n \text{ factors}}$ and g^{-n} is shorthand for $\underbrace{g^{-1}g^{-1} \cdots g^{-1}}_{n \text{ factors}}$. It is straightforward to check that the usual rules for exponents apply.

Exercises 2.1.3. Basic Properties of Groups

(a) Prove each of the properties in Proposition 2.1.2

(b) Show that it makes sense to define $g^0 = e$.

(c) Prove by induction that for a positive integer n, $(g^{-1})^n = (g^n)^{-1}$.

For a group in which the operation is $+$ and the identity is 0 (in particular, the group must be abelian), the sum $\underbrace{g + g + \cdots + g}_{n \text{ terms}}$ is written ng. Think of this as repeated addition, not as multiplication: the group just has one operation, and n is an integer, not necessarily an element of the group. For a negative integer $-n$ (with $n > 0$), we define $(-n)(g) = \underbrace{(-g) + (-g) + \cdots + (-g)}$ and show this is equal to $-(ng)$ by the same argument as used in the previous exercise.

The first examples come from a familiar place, the integers, as discussed in Section 1.2.

Example 2.1.4. The integers $\mathbb{Z}$ form a group with operation $+$, identity element 0 and inversion operation $a \longmapsto -a$. The elements 1 and -1 generate the group $\mathbb{Z}$ in the sense that by applying inversion and repeated addition we can get all the other elements of $\mathbb{Z}$. This is not true for other elements.

The set of multiples of n in $\mathbb{Z}$ also is a group under $+$ with identity 0. Adding two multiples of n gives another multiple of n, and the additive inverse of a multiple of n is also a multiple of n. We will denote this group by $n\mathbb{Z}$. Later in this section we introduce the abstract definition of a subgroup. Here $n\mathbb{Z}$ is a subgroup of $\mathbb{Z}$. The elements n and $-n$ generate $n\mathbb{Z}$ just as 1 and -1 generate $\mathbb{Z}$.

The integers modulo n, introduced in Section 1.1, also form a group under $+$. We can write the elements as $0, 1, 2, \ldots, n-1$, but these are really shorthand for congruence classes. When consider the integers modulo n as a group (ignoring multiplication) we will write it as $\mathbb{Z}_n$. An interesting question is: what elements generate $\mathbb{Z}_n$?

Definition 2.1.5. A single element g of a group G **generates** G when any element of G is equal to g^n for some $n \in \mathbb{Z}$ (or, if the operation of the group is addition, any element is equal to ng for some n). Such a group is said to be **cyclic**.

Let S be a subset of a group G. We say that G **is generated** by S if any element of G is equal to a product (with an arbitrary number of terms) of elements of S and elements of $\left\{s^{-1} : s \in S\right\}$.

A group is $t-$**generated** if there is a subset S of G with t elements that generates G.

Exercises 2.1.6. A 2-GENERATED GROUP.

The Cartesian product $\mathbb{Z} \times \mathbb{Z}$ is a group under coordinatewise addition with identity element $(0, 0)$ and inverse operator $(a, b) \longmapsto (-a, -b)$.

(a) Show that it is not possible to generate all elements by repeated addition of a single element. [Show that if (a, b) generates the group then both a and b have to be ± 1. Then show that $(1, 1)$ can't generate all elements of $\mathbb{Z} \times \mathbb{Z}$.]

(b) Show that this group is $2-$generated.

The Cartesian product $\mathbb{Z}/m \times \mathbb{Z}/n$ under coordinatewise addition (and using coordinatewise identity elements and inversion) is also a group. In certain cases it is possible to have a single element generate all elements by repeated addition.

Exercises 2.1.7. MORE ON GENERATORS FOR GROUPS.

(a) Consider $\mathbb{Z}_n$ for $n = 2, 3, 4, 5, 6, 7$. Which elements $a \in \mathbb{Z}_n$ generate all of $\mathbb{Z}_n$?

(b) Experiment with some small integers m and n to find cases in which $\mathbb{Z}_m \times \mathbb{Z}_n$ is generated by a single element and other cases in which it is not.

The following proposition introduces another very tangible example of a group, which is derived from the ring $\mathbb{Z}/n$. This is an elaboration of Exercise 1.1.18 concerning multiplication in $\mathbb{Z}/n$. We say that an element $[a]$ in $\mathbb{Z}/n$ is a **unit** if $[a]$ has a multiplicative inverse modulo n, that is, there is some $[u] \in \mathbb{Z}_n$ such that $[a][u] = [1]$.

Proposition 2.1.8. *The congruence class $[a]$ in $\mathbb{Z}/n$ is a unit if and only if a is coprime to n. The units in $\mathbb{Z}/n$ form a group under multiplication. This group is denoted U_n.*

Proof. Let $n > 1$ be an integer and let $a \in \mathbb{Z}$. We first show that a is coprime to n if and only if $[a]$ has a multiplicative inverse modulo n. If a and n are coprime then, by the GCD Theorem 1.1.4, there exist integers u and v such $au + vn = 1$. Reducing modulo n we have

$$[a][u] = [a][u] + [v][0] = [a][u] + [b][n] = [1]$$

Conversely if $[a]$ has an inverse modulo n, say $[a][u] = [1]$ then $au - 1$ must be a multiple of n. But if $au - 1 = vn$ then the smallest positive linear combination of a and n is 1. By the GCD Theorem a and n are coprime.

Let U_n be the set of elements in $\mathbb{Z}/n$ that have multiplicative inverses. The identity element of U_n is $[1]$. Clearly if $[a]$ is in U_n and $[u]$ is such that $[a][u] = [1]$ then $[u]$ is the inverse of $[a]$ and vice-versa. If $[a], [b] \in U_n$ and their inverses are, respectively, u and v, then by associativity and commutativity, $([a][b])([u][v]) = ([a][u])([b][v]) = [1][1] = [1]$ so ab is also invertible. Thus multiplication is an operation on U_n (it maps $U_n \times U_n$ to U_n); there is an identity element, $[1]$, and a unary operation corresponding to inversion. □

The groups U_n are interesting because their structure is not immediately obvious. The additive group $\mathbb{Z}_n, +$ is easy to understand; it is cyclic, generated by $[1]$. Some groups U_n are cyclic, but some are not. We won't fully determine the structure of U_n until we establish the Chinese Remainder Theorem for rings (or specifically $\mathbb{Z}/n$) in Theorem 4.6.15.

Definition 2.1.9. The **cardinality** of a group G, written $|G|$, is just the cardinality of the underlying set. It is also called the **order** of the group.

Many of our examples will be finite groups and we will be studying some of the properties that go into understanding the structure of finite groups.

Example 2.1.10. Recall the definition of the dihedral group D_n, which is the symmetry group of a regular n-gon (1.2.4). We showed that D_n has $2n$ elements and that it is 2-generated—by r, the rotation by $2\pi/n$, and any reflection, t_i. (Well, we did this for D_5, but the same argument holds for D_n).

Section 1.2 also introduced S_n, the group of all possible permutations on $\{1, 2, \ldots, n\}$. This group has cardinality $n!$. Composing them is just composing functions. The inverse permutation involves flipping the two rows of the permutation and then, for convenience, rearranging the columns so that the first row is increasing order.

$$\text{For} \qquad \pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 1 & 5 & 2 \end{pmatrix}, \qquad \text{the inverse is } \pi^{-1} \quad = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 1 & 4 \end{pmatrix}.$$

The Order of an Element

Definition 2.1.11. For $g \in G$ the **order of the element** g is the smallest positive integer n such that $g^n = e$, if such an n exists. If no such n exists then g has infinite order. We use $|g|$ or $\mathrm{ord}(g)$ for the order of g.

The **exponent** of G is the least common multiple of the orders of the elements of G, if such an integer exists, that is $\exp(G) = \mathrm{lcm}\,\{\mathrm{ord}(g) : g \in G\}$. If no such element exists one can say the exponent is infinite.

Exercises 2.1.12. Basic properties of order.

(a) Only the identity element of a group has order 1.

(b) Every nonzero element of $\mathbb{Z}$ has infinite order.

(c) In $\mathbb{Z}_n$ some elements have order n, but other non-identity elements may have a different order.

(d) For any finite group there is a well defined exponent, but an infinite group may or may not have one.

Theorem 2.1.13 (Order Theorem)**.** *Let g be an element of the group G.*

(1) If g has infinite order, then elements g^t for $t \in \mathbb{Z}$ are all distinct.

(2) If g has order n then

(a) $g^t = g^s$ if and only if $t \equiv s \mod n$. In particular, $g^t = e$ if and only if n divides t.

(b) $\mathrm{ord}(g^t) = \frac{n}{\gcd(t,n)}$.

Proof. Suppose $g^t = g^s$ for integers s, t. Then $g^{t-s} = e_G$. If g has infinite order then $s - t = 0$ so $s = t$. This proves item (1).

Suppose g has order n. Then $g^0 = emg^1, g^2, \ldots, g^{n-1}$ must all be distinct, by an argument similar to the previous paragraph. If $g^t = g^s$ for $0 \leq s \leq t < n$ then $g^{t-s} = e$ with $0 \leq t - s < n$. Since the order of g is n, $t - s = 0$, so $t = s$. For an arbitrary integer t use the quotient remainder theorem to write $t = nq + r$. Then $g^t = g^{nq+r} = (g^n)^q g^r = e^q g^r = g^r$. This establishes claim (2a): $g^t = g^s$ if and only if t and s have the same remainder when divided by n.

Now let $d = \gcd(t, n)$ and write $t = da$ and $n = db$. Then a, b have no common factor (otherwise d would not be the gcd) and we observe that $b = n/\gcd(t, n)$.

We now have $(g^t)^b = g^{dab} = (g^n)^a = e$. Furthermore, if $(g^t)^s = e$ then, by (2a), $n = db$ divides $ts = das$. Cancelling d and taking note of b and a being coprime we get b divides s. This establishes (2b). □

Example 2.1.14. The reflections in D_n all have order 2. The rotation by $2\pi/n$ has order n, but some of the other rotations have order less than n. For example in D_6 with r the rotation by $2\pi/6$, r^2 has order 3 and r^3 has order 2.

Exercises 2.1.15. More card tricks.

(a) Suppose that every element of G has order 2. Show that G is abelian.

(b) If G has even order then G has an element of order 2. (Consider the pairing of g with g^{-1}).

Exercises 2.1.16. Order and commutativity.

(a) If $g \in G$ has order m and $h \in H$ has order n, find the order of $(g, h) \in G \times H$.

(b) Suppose that $a, b \in G$ commute (that is $ab = ba$). If $\operatorname{ord}(a)$ and $\operatorname{ord}(b)$ are coprime find the order of ab.

(c) Let A be an abelian group with finite exponent. Show that there is some $a \in A$ such that $\operatorname{ord}(a) = \exp(A)$. [This is a bit more challenging and uses part (b).]

2.2 Subgroups

A key area of investigation in many mathematical subjects is the subsets of a given object that have useful structure. In this section, we treat subsets of a group that are themselves groups.

Definition 2.2.1. A nonempty subset H of a group G is a **subgroup**, when H is a group using the operation $*_G$ on G. We will write $H \leq G$ when H is a subgroup of G (as opposed to $H \subseteq G$ when H is just a subset), and $H < G$ when H is a proper subgroup (that is $H \neq G$).

The following proposition is a sanity check on our definition of subgroup: the identity element and inversion are the same for the subgroup as for the group.

Proposition 2.2.2. *If K is a subgroup of G then $e_K = e_G$ and the inversion operation is the same on K as it is on G.*

Proof. If K is a subgroup of G then it must have an identity element. For any $k \in K$, we have (using $*_K = *_G$) that $e_K *_K k = e_K *_G k = k$. Proposition 2.1.2 (4) shows that it must be the case that $e_K = e_G$. (If something acts like the identity it is the identity!)

Let $k \in K$ and let k^{-1} be its inverse in G. This k^{-1} is also the inverse of k in the subgroup K because $k * k^{-1} = e_G = e_K$. □

(Thank goodness for both of these facts.)

In practice the following proposition is used to check if a subset of a group is a subgroup. We will say that $H \subseteq G$ is **closed under inversion** when for any $h \in H$ the inverse h^{-1} is also in H. We say is H **closed under multiplication** when for any $h, k \in H$, the product $h *_G k$ is also an element of H.

Proposition 2.2.3 (Subgroup Properties)**.** *If H is a nonempty subset of G that is closed under inversion and closed under multiplication in G then H is a subgroup of G (i.e. it also contains e_G).*

If H is a nonempty subset of G such that $h *_G k^{-1} \in H$ for all $h, k \in H$ then H is a subgroup of G.

Proof. Since H is nonempty, it contains some element k. Since H is closed under inversion, $k^{-1} \in H$. Since H is closed under multiplication, $k *_G k^{-1} = e_G \in H$. Thus H satisfies the definition of a group since it has associativity (immediate since $*_H$ is the restriction of $*_G$), an identity element, and inverses (by assumption).

Suppose H is a nonempty subset of G such that $h *_G k^{-1} \in H$ for all $h, k \in H$. For any $k \in H$, setting $h = k$ gives $k *_G k^{-1} = e_G \in H$. Letting $h = e_G$ gives $e_G * k^{-1} = k^{-1} \in H$, so H is closed under inversion. Now for any $h, k \in H$ we know $k^{-1} \in H$, so $h *_G (k^{-1})^{-1} = h *_G k \in H$. This shows H is closed under multiplication. □

When proving that a particular subset of a group is a subgroup one can either show the set is closed under inversion and under multiplication, or use the second property of the theorem. I like the clarity of proving closure under each operation.

Exercises 2.2.4. Subgroup constructions

- **(a)** Let G be a group and let g an arbitrary element of G. Show that $\left\{g^i : i \in \mathbb{Z}\right\}$ is a subgroup of G. This group is called the **cyclic subgroup generated by** g and is written $\langle g \rangle$.
- **(b)** Let G be a group. Show that the set $Z(G) = \{a \in G : ag = ga \text{ for all } g \in G\}$ is an abelian subgroup of G. It is called the **center** of G.
- **(c)** Let H and K be subgroups of G. Show that their intersection is also a subgroup of G.

Note that there is consistency between the order of an element and the order of the subgroup it generates. If $g \in G$ has order n then the set of powers of g is $\left\{g^0 = e_G, g, g^2, \ldots, g^{n-1}\right\}$ (any other power of g is one of these). This set is a subgroup of G of order n. If $g \in G$ has infinite order then it generates a cyclic subgroup that is infinite. (It is not really cycling in the infinite case, but that's the term used!)

Example 2.2.5. As was discussed in Section 1.2, the integers, $\mathbb{Z}$, the rational numbers, $\mathbb{Q}$, the real numbers, $\mathbb{R}$, and the complex numbers, $\mathbb{C}$, are all abelian groups under addition. We sometimes write $\mathbb{Z}, +$ (for example) to emphasize that we are ignoring multiplication, and are just considering the additive properties of $\mathbb{Z}$. Clearly we have a sequence of subgroups.

$$\mathbb{Z} < \mathbb{Q} < \mathbb{R} < \mathbb{C}$$

Exercises 2.2.6. Some subgroups of abelian groups

Let $A, +$ be an abelian group and let m be an integer.

(a) Let $mA = \{ma : a \in A\}$. Show that mA is a subgroup of A.

(b) Let $A[m] = \{a \in A : ma = 0\}$. Show that $A[m]$ is a subgroup of A.

(c) Give an example in which $mA \cap A[m]$ is trivial (just 0) and given an example in which it is not trivial. (Try $\mathbb{Z}_n$ for a few choices of n.)

(d) Compute $m\mathbb{Z} \cap n\mathbb{Z}$ and $\mathbb{Z}[m] \cap \mathbb{Z}[n]$.

Example 2.2.7. We showed in Section 1.2 that D_5 may be identified with a subgroup of S_5. The discussion can be generalized to show D_n identified as a subgroup S_n for any $n \geq 3$.

The permutation group S_4 may also be seen as contained in S_5; it is just the set of all permutations in S_5 that take 5 to 5 (we say these "fix" 5). There are other subgroups that have the exact same structure as S_4. For example the set of all permutations in S_5 that fix 3. These subgroups are all isomorphic (see Section 2.5).

Proposition 2.2.8. *Let $\mathcal{H}$ be a set of subgroups of a group G. Let H be the intersection of all the elements of $\mathcal{H}$,*

$$K = \bigcap_{H \in \mathcal{H}} H$$

Then K is itself a subgroup of G.

Proof. We have to show that K is nonempty and is closed under multiplication and inversion. Since each $H \in \mathcal{H}$ is a subgroup of G, each contains the identity, so $e_G \in K$. If $k \in K$ then k is an element of each H in $\mathcal{H}$. Since each $H \in \mathcal{H}$ is a group, $k^{-1} \in H$ for each $H \in \mathcal{H}$. Thus k^{-1} is in the intersection $K = \bigcap_{H \in \mathcal{H}} H$. Similarly if k and k' are elements of K then k and k' are in each $H \in \mathcal{H}$ and the product kk' is in each $H \in \mathcal{H}$. Thus $kk' \in K$. □

Definition 2.2.9. Let B be a subset of a group G we define the **group generated by** B to be the intersection of all subgroups of G containing B. We denote it $\langle B\rangle$.

If H and K are subgroups of G their **join**, written $H \vee K$, is $\langle H \cup K\rangle$.

By the proposition, $\langle B\rangle$ is a subgroup of G. Furthermore, since by definition it is the intersection of all subgroups of G containing B, it is contained in every subgroup of G containing B. Thus it makes sense to call it the **smallest subgroup of** G **containing** B. The join of two subgroups plays a similar role to the union of sets. The subgroup $H \vee K$ is the smallest subgroup of G that contains both H and K.

The definition is not very useful for computing $\langle B\rangle$. How would one find all subgroups of G containing B and then find the intersection of them?! It is more practical to adopt a constructive (or "bottom up") approach to find all the elements of $\langle B\rangle$. That is, if $b_1, b_2 \in B$ then $b_1^k b_2^m \in \langle B\rangle$ for any integers k, m. This is not simple either(!), but for small examples it can be useful to compute the subgroup $\langle B\rangle$, as we will see below.

The Lattice of Subgroups

Consider a fixed group G and let S be the set of all subgroups of G. For small groups, it is often illuminating to draw a diagram showing all these subgroups and the containment relationships among them. There is no simple and efficient process for this in general, but there are some useful heuristics for small examples. One is to find the extremal cases of subgroups.

Definition 2.2.10. Let G be a group. A proper subgroup M of G is **maximal** when there is no subgroup H of G satisfying $M < H < G$. Similarly, a nontrivial subgroup M of G is **minimal** when there is no subgroup H of G satisfying $\{e_G\} < H < M$.

For any group G a single element $a \in G$ generates a subgroup $\langle a\rangle$, which we call a cyclic subgroup. Any minimal subgroup must be cyclic (but the converse is not true!). This suggests a general approach: work up from the trivial subgroup $\{e_G\}$ to construct the lattice of subgroups.

(1) Find all subgroups generated by 1 element. Take account of containment relationships among them.

(2) Find all subgroups generated by 2 elements by adding a new element to the 1-generated subgroups.

(3) See if there are subgroups that require 3 generators by adding an element to the 2-element subgroups.

Note that there are usually several ways to generate a particular subgroup.

Here are some examples.

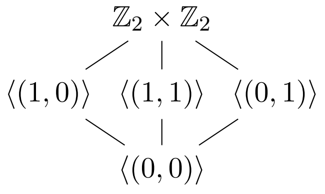


Figure 2.1: The lattice diagram for $\mathbb{Z}_2 \times \mathbb{Z}_2$.

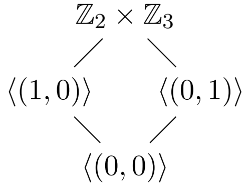


Figure 2.2: The lattice diagram for $\mathbb{Z}_2 \times \mathbb{Z}_3$.

Example 2.2.11. The lattice of subgroups of $\mathbb{Z}$ is infinite, but we can get some sense for its structure. We have shown that for an integer n, the multiples of n, $n\mathbb{Z}$, form a subgroup of $\mathbb{Z}$. The integers n and $-n$ generate the same group, so we may restrict our attention to nonnegative integers. I claim these are the only subgroups of $\mathbb{Z}$ and that the subgroups for distinct $n \in \mathbb{N}$ are unequal. Suppose A is a subgroup of $\mathbb{Z}$. If A has no positive elements then, since it is closed under (additive) inverses it must also have no negative elements, and $A = \{0\} = 0\mathbb{Z}$. Suppose A does have positive elements, and let n be the smallest positive element

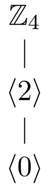


Figure 2.3: The lattice diagram for $\mathbb{Z}_4$.

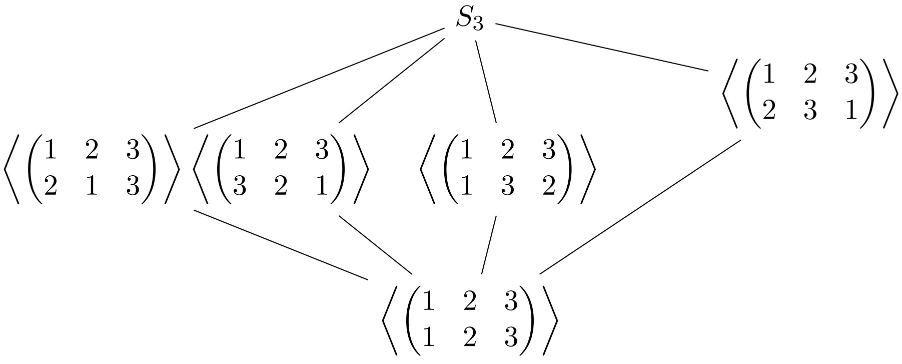

S_3
$\left\langle\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}\right\rangle$
$\left\langle\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}\right\rangle$
$\left\langle\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}\right\rangle$
$\left\langle\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}\right\rangle$
$\left\langle\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix}\right\rangle$

Figure 2.4: The lattice diagram for S_3.

of A (we are using the Well-Ordering Principle). Let a be another element of A. By the Quotient-Remainder Theorem $a = qn + r$ for some $0 \leq r < n$. Since a and qn are both in A, $a - qn = r \in A$. By assumption, n is the smallest positive element of A, so r must be 0. Thus every element of A is a multiple of n, and we have shown $A = n\mathbb{Z}$.

The lattice of subgroups of $\mathbb{Z}$ simply reflects divisibility properties: $n\mathbb{Z} \leq d\mathbb{Z}$ if and only if $d \mid n$. The maximal subgroups of $\mathbb{Z}$ are generated by prime numbers. A prime p has no divisors except $\pm p$ and ± 1 so the only subgroup of $\mathbb{Z}$ properly containing $p\mathbb{Z}$ is $\mathbb{Z}$ itself. There are no minimal subgroups of $\mathbb{Z}$ because for any $n > 0$ there are (infinitely many) subgroups such as $2n\mathbb{Z}$, $3n\mathbb{Z}$, $4n\mathbb{Z}$, that are proper subgroups of $n\mathbb{Z}$ and are not equal to $\{0\}$.

Exercises 2.2.12. Lattice Diagrams for Groups

(a) Draw the subgroup lattice diagram for $\mathbb{Z}_{45}$.

(b) Draw the subgroup lattice diagram for $\mathbb{Z}_{60}$.

(c) Draw the subgroup lattice diagram for $\mathbb{Z}_2 \times \mathbb{Z}_4$.

(d) Draw the subgroup lattice diagram for $\mathbb{Z}_3 \times \mathbb{Z}_4$.

(e) Find all subgroups of $\mathbb{Z}_4 \times \mathbb{Z}_4$. Describe the logic of your process for finding them. Present them in an organized fashion. Draw the lattice if you can.

2.3 Group Homomorphisms

In any algebraic subject a key starting point is to identify functions that are appropriate to study. These functions are typically called homomorphisms (from classical Greek: hom meaning “same” and morph “shape”) because they preserve the algebraic structures on which they act.

Definition 2.3.1. Let G and H be groups. A function $\varphi : G \longrightarrow H$ is a homomorphism when

(1) $\varphi(g_1 *_G g_2) = \varphi(g_1) *_H \varphi(g_2)$ for all $g_1, g_2 \in G$, and

(2) $\varphi(e_G) = e_H$, and

(3) $\varphi(g^{-1}) = (\varphi(g))^{-1}$ for all $g \in G$.

I like to speak informally about a homomorphism as a function that **respects structure**: A homomorphism of groups "respects" the property of the identity element, multiplication, and inversion.

It is fairly easy to show that the first item in the definition of homomorphism implies the other two. This result and another important result are contained in the following proposition.

Proposition 2.3.2 (Homomorphisms)**.** *Let G, H, K be groups.*

If $\varphi : G \longrightarrow H$ is a function such that $\varphi(g_1 *_G g_2) = \varphi(g_1) *_H \varphi(g_2)$ then φ is a group homomorphism.

If $\varphi : G \longrightarrow H$ and $\theta : H \longrightarrow K$ are group homomorphisms then the composition $\theta \circ \varphi$ is also a group homomorphism.

Exercises 2.3.3. PROVE THE PROPOSITION.

(a) Assuming that φ respects multiplication, show that it also takes the identity of G to the identity of H (use $e_G * e_G = e_G$) and that it respects inversion (use $gg^{-1} = e_G$).

(b) Prove that the composition of homomorphisms is a homomorphism.

The simplest type of a homomorphism is the **inclusion** of a subgroup H of G into G. That is, when $H < G$, then there is a function $H \longrightarrow G$ that takes elements of H to themselves (now thought of as elements of G). A homomorphism of groups may be surjective, injective or neither. A injective homomorphism $\varphi : G \longrightarrow H$ is often called an **embedding**.

Proposition 2.3.4. *Let g be an element of a group G. There is homomorphism*

$$\begin{aligned} \varphi : \mathbb{Z} &\longrightarrow G \\ t &\longmapsto g^t \end{aligned}$$

When $g \in G$ has infinite order, this homomorphism is injective.

When $g \in G$ has order n, the function below is an injective homomorphism.

$$\begin{aligned} \varphi : \mathbb{Z}_n &\longrightarrow G \\ t &\longmapsto g^t \end{aligned}$$

Proof. Let g be an element of a group G. Consider the function $\mathbb{Z} \longrightarrow G$ taking t to g^t. The function is a homomorphism because for $s, t \in \mathbb{Z}$, $s + t$ maps to $g^{s+t} = g^s * g^t$ and this is the product of the images of s and t.

Now assume that g has infinite order. Suppose that $g^s = g^t$. Then $g^{s-t} = e_G$. Since g has infinite order $s - t = 0$, so $s = t$. This proves injectivity.

The proof for g of finite order is similar. □

Notice that the homomorphisms of the proposition are completely defined by the requirement that 1 maps to g. For if 1 maps to g then $1 + 1$ maps to $g * g$, and inductively, a positive number t must map to g^t. Since homomorphisms respect inversion, $-t$ maps to g^{-t}.

Proposition 2.3.5. *Let $\varphi : G \longrightarrow H$ be a homomorphism. For any subgroup G' of G, the image of G', which we write $\varphi(G')$, is a subgroup of H.*

Proof. We need only show that the image is closed under inversion and multiplication. Consider an element of $\varphi(G')$. We may write it as $\varphi(g)$ for some $g \in G'$. By the properties of a homomorphism

$$\varphi(g^{-1})\varphi(g) = \varphi(g^{-1}g) = \varphi(e_G) = e_H$$

This shows that the inverse of $\varphi(g)$ is $\varphi(g^{-1})$. Since G' is a subgroup of G, $g^{-1} \in G'$, and we can conclude that $\big(\varphi(g)\big)^{-1} \in \varphi(G')$.

Consider two elements of $\varphi(G')$, which we may write as $\varphi(g_1)$ and $\varphi(g_2)$ for $g_1, g_2 \in G'$. Their product is $\varphi(g_1)\varphi(g_2) = \varphi(g_1 g_2)$, and this is in $\varphi(G')$ since $g_1 g_2 \in G'$. We have shown that $\varphi(G)$ is closed under inversion and under multiplication so it is a subgroup of G. □

The Kernel of a Homomorphism

Definition 2.3.6. Let $\varphi : G \longrightarrow H$ be a homomorphism. The **kernel** of φ is the set of elements that map to the identity in H, that is

$$\ker(\varphi) = \varphi^{-1}(e_H) = \{g \in G : \varphi(g) = e_H\}.$$

Proposition 2.3.7. *Let $\varphi : G \longrightarrow H$ be a homomorphism. The kernel of φ is a subgroup of G. Furthermore, for any $a \in \ker(\varphi)$ and any $g \in G$, $gag^{-1} \in \ker(\varphi)$.*

Proof. By the definition of homomorphism $e_G \in \ker(\varphi)$. Suppose $a \in \ker(\varphi)$. We have

$$e_H = \varphi(e_G) = \varphi(aa^{-1}) = \varphi(a)\varphi(a^{-1}) = e_H\varphi(a^{-1}) = \varphi(a^{-1})$$

So a^{-1} is also in $\ker(\varphi)$. If b is another element in $\ker(\varphi)$ then $\varphi(ab) = \varphi(a)\varphi(b) = e_H e_H = e_H$. Thus $\ker(\varphi)$ is closed under inversion and multiplication, so it is a subgroup of G. The final claim is a similar computation.

$$\begin{aligned}\varphi(gag^{-1}) &= \varphi(g)\varphi(a)\varphi(g^{-1}) = \varphi(g)e_H\varphi(g^{-1})\\ &= \varphi(g)\varphi(g^{-1}) = \varphi(gg^{-1}) = \varphi(e_G) = e_H\end{aligned}$$

□

The kernel of a homomorphism and the property described in the last sentence of the proposition are important concepts, as we will see in Section 2.8 Here is another important aspect of the kernel, it gives a simple test for injectivity. Recall that to test if a function $f : A \longrightarrow S$ is injective we show that for $a, b \in A$, $f(a) = f(b)$ implies $a = b$.

Proposition 2.3.8. *A homomorphism $\varphi : G \longrightarrow H$ is injective if and only if the kernel is trivial,* $\ker(\varphi) = \{e_G\}$*.*

Proof. For any homomorphism $\varphi(e_G) = e_H$, so injectivity forces $\ker(\varphi) = \{e_G\}$.

Suppose that the only element of $\ker(\varphi)$ is e_G, let us show that φ is injective. Let $a, b \in G$ be such that $\varphi(a) = \varphi(b)$. Using the properties of a homomorphism and $\varphi(a) = \varphi(b)$ we have

$$phi(ab^{-1}) = \varphi(a)\varphi(b^{-1}) = \varphi(a)(\varphi(b))^{-1} = \varphi(a)(\varphi(a))^{-1} = e_H$$

This forces $ab^{-1} = e_G$, and multiplying on the right by b we get $a = b$. This shows that φ is injective. □

Exercises 2.3.9. The preimage of a subgroup

Prove this extension of Proposition 2.3.7 and compare with Proposition 2.3.5.

(a) Let $\varphi : G \longrightarrow H$ be a homomorphism. For any subgroup H' of H, the preimage of H', which we write $\varphi^{-1}(H')$, is a subgroup of G.

Isomorphisms

Definition 2.3.10. A homomorphism φ that is also a bijection (one-to-one and onto) is called an **isomorphism**. When there exists an isomorphism $\varphi : G \longrightarrow H$ we say G and H are **isomorphic** and write $G \cong H$.

The following proposition shows that the relation of being isomorphic satisfies symmetry and transitivity so it determines an equivalence relation on any set of groups.

Proposition 2.3.11 (Isomorphisms)**.** *If φ is an isomorphism of groups, then the inverse function φ^{-1} is also an isomorphism of groups.*

If $\varphi : G \longrightarrow H$ and $\theta : H \longrightarrow K$ are group isomomorphisms then the composition $\theta \circ \varphi$ is also a group isomorphism.

On any set of groups $\mathcal{G}$, the relation of being isomorphic is an equivalence relation.

Proof. Let $\varphi : G \longrightarrow H$ be an isomorphism of groups. By definition, φ is both injective and surjective, so there is a well defined inverse function, φ^{-1}. We must show that φ^{-1} is a homomorphism.

Let h_1 and h_2 be two elements of h. Since φ is surjective there are two elements g_1 and g_2 such that $\varphi(g_1) = h_1$ and $\varphi(g_2) = h_2$. Since φ is injective these two elements are uniquely defined. We now show that φ^{-1} respects products, which is sufficient to show it is a homomorphism.

$$\begin{aligned}\varphi^{-1}(h_1h_2) &= \varphi^{-1}\left(\varphi(g_1)\varphi(g_2)\right)\\ &= \varphi^{-1}\left(\varphi(g_1g_2)\right) \qquad \text{since } \varphi \text{ is a homomorphism}\\ &= g_1g_2 \qquad \text{since } \varphi \text{ and } \varphi^{-1} \text{ are inverse functions}\\ &= \varphi^{-1}(h_1)\varphi^{-1}(h_2)\end{aligned}$$

The composition of two bijections is a bijection, and by Proposition 2.3.5 we know that the composition of homomorphisms is a homomorphism. Thus the composition of two isomorphisms is an isomorphism.

On any set of groups we can define a relation as follows: G is related to G if there is an isomorphism from G to H. The relation is clearly reflexive since any group is isomorphic to itself under the identity map. The first part of this proposition shows the relation is symmetric: if G is isomorphic to H then H is also isomorphic to G. The second part establishes transitivity. □

If there is an isomorphism φ from group G to H then G and H have the same algebraic structure. Since φ is a bijection it gives a pairing of elements of G with elements of H. The image of $g \in G$ will have the same order as g does. The image of a subgroup G' of G is a subgroup of H (by Proposition 2.3.5) that is isomorphic to G'. Distinct subgroups of G will be mapped to distinct subgroups of H and the lattice of subgroups of G will have the same structure as the lattice of subgroups of H.

Exercises 2.3.12. Homomorphisms on $\mathbb{Z}_n$

(a) Show that for each $a \in \mathbb{Z}_n$ there is a unique homomorphism $\varphi_a : \mathbb{Z}_n \longrightarrow \mathbb{Z}_n$ such that $\varphi_a(1) = a$.

(b) Under what conditions on a is φ_a an isomorphism? [Try some examples with $n < 10$.]

(c) Identify all subgroups of $\mathbb{Z}_n$.

Exercises 2.3.13. Homomorphisms on cyclic groups

(a) Show that if $d \mid n$ then there is a unique homomorphism φ from $\mathbb{Z}_n$ to $\mathbb{Z}_d$ that takes 1 (in $\mathbb{Z}_n$) to 1 (in $\mathbb{Z}_d$). [Show that respecting the operation of addition forces a unique choice for $\varphi(a)$. Show also that this does give a homomorphism.]

(b) Show that φ is also surjective.

(c) Show that if d does not divide n then there is no homomorphism $\mathbb{Z}_n$ to $\mathbb{Z}_d$ that takes 1 (in $\mathbb{Z}_n$) to 1 (in $\mathbb{Z}_d$). [Try to define it and run into a roadblock.]

Exercises 2.3.14. Homomorphisms and dihedral groups

(a) Show that there is an injective homomorphism from $\mathbb{Z}_n$ into D_n taking 1 to rotation by $2\pi/n$.

(b) How many injective homomorphisms are there from $\mathbb{Z}_n$ into D_n?

(c) Identify all subgroups of D_n for $n = 3, 4, 5, 6$. Draw a lattice diagram showing containment of subgroups.

Exercises 2.3.15. A Perverse Group

(a) Show that $\mathbb{Z}$ is a group under the operation $\square$ defined by $a \square b = a + b - 2$. (What is the identity element? What is the inverse of an element a?)

(b) Find an isomorphism from $\mathbb{Z}, +$ to $\mathbb{Z}, \square$.

Exercises 2.3.16. Order and homomorphisms

Prove the following results about the relationship between the order of an element and the order of its image under a homomorphism.

(a) If $\varphi : G \longrightarrow H$ is a homomorphism, then $\mathrm{ord}(\varphi(g))$ divides $\mathrm{ord}(g)$.

(b) If $\varphi : G \longrightarrow H$ is an isomorphism, then $\mathrm{ord}(\varphi(g)) = \mathrm{ord}(g)$.

The previous exercises give important restrictions on homomorphisms. If you want to create a homomorphism from G to H, each element g in G must go to an element of H that has order dividing $\mathrm{ord}(g)$.

Exercises 2.3.17. Other homomorphisms

(a) Show that there is a nontrivial homomorphism from D_3 to $\mathbb{Z}_2$ but that any homomorphism from D_3 to $\mathbb{Z}_3$ is trivial.

2.4 Some Constructions of Groups

In this section we show two ways to construct new groups from ones that we already have. Both have been touched on briefly; we give more detail here. The first construction is the direct product of groups. The idea is simple (and was illustrated in the first examples of groups that we gave). Given two groups, form the Cartesian product as sets, and apply componentwise operations to get a new group. The construction yields, in a natural way, two types of homomorphisms that are important despite their simplicity. The direct product can also be applied to several groups, not just two. The second construction is quite a bit more mysterious, for a given group G the isomorphisms of G to itself have a group structure that is useful in understanding the properties of G.

The Direct Product

Definition 2.4.1. Let G and H be groups. The Cartesian product $G \times H$, along with the unary operation (of inversion) and the binary operation (of multiplication) below form the **direct product** of G and H.

$$(g,h)^{-1} = (g^{-1}, h^{-1})$$
$$(g_1,h_1) *_{G\times H} (g_2,h_2) = (g_1 *_G g_2, h_1 *_H h_2)$$

The identity element is of course (e_G, e_H).

The following proposition shows that the direct product is in fact a group and gives other important properties. None of the following are surprising and they are routine to prove.

Theorem 2.4.2 (Direct Product)**.** *Let G and H be groups.*

(1) The above definition does, indeed, make $G \times H$ a group.

(2) The associative law for the product of several groups holds: $G_1 \times (G_2 \times G_3) \cong (G_1 \times G_2) \times G_3$.

(3) $G \times H$ is abelian if and only if G and H are abelian.

(4) If G' is a subgroup of G and H' is a subgroup of H then $G' \times H'$ is a subgroup of $G \times H$. In particular $G \times \{e_H\}$ and $\{e_G\} \times H$ are subgroups of $G \times H$.

(5) There is an injective homomorphism $i_G : G \longrightarrow G \times H$ taking g to (g, e_H) (and similarly $i_H : H \longrightarrow G \times H$).

(6) The projection maps $p_G : G \times H \longrightarrow G$ and $p_H : G \times H \longrightarrow H$ are surjective homomorphisms.

(7) The construction and the observations above can be generalized to the direct product of any set of groups $\{G_i : i \in I\}$ indexed by a finite set I. (It extends to infinite index sets I with some modification due to subtle issues.)

Notice that the kernel of the homomorphism p_G is $\{(e_G, h) : h \in H\}$ and this is exactly the image of the homomorphism i_H. Similarly $\ker(p_H) = \operatorname{im}(i_G)$.

Exercises 2.4.3. Subgroups of a direct product

(a) Not all subgroups of $G \times H$ are direct products of subgroups of G and H. Illustrate with some examples: $\mathbb{Z}_2 \times \mathbb{Z}_2$, $\mathbb{Z}_4 \times \mathbb{Z}_4$.

The following result is more subtle and it turns out to be a powerful idea.

Proposition 2.4.4 (Universal Property of the Product)**.** *Let G, H, and T be groups, and let $\varphi : T \longrightarrow G$ and $\psi : T \longrightarrow H$ be homomorphisms. The function $\alpha : T \longrightarrow G \times H$ defined by $t \longmapsto (\varphi(t), \psi(t))$ is a homomorphism. It is the unique homomorphism such that $p_G \circ \alpha = \varphi$ and $p_H \circ \alpha = \psi$.*

Proof. Note first that, by construction, $p_G \circ \alpha = \varphi$ and $p_H \circ \alpha = \psi$. Furthermore, there is no other choice for the definition of α that satisfies these two requirements.

We have to show that α respects inversion and multiplication. Let $t \in T$. We have to show that $\alpha(t^{-1})$ is the inverse of $\alpha(t)$. The subscript on $*$ that we sometimes use to show the group being used is omitted in the following derivation, but it is worthwhile to identify it while reading.

$$\begin{aligned}
\alpha(t) * \alpha(t^{-1}) &= \Big(\varphi(t), \psi(t)\Big) * \Big(\varphi(t^{-1}), \psi(t^{-1})\Big) \\
&= \Big(\varphi(t) * \varphi(t^{-1}), \psi(t) * \psi(t^{-1})\Big) \\
&= \Big(\varphi(t * t^{-1}), \psi(t * t^{-1})\Big) \\
&= (e_H, e_K)
\end{aligned}$$

This proves that $\alpha(t^{-1})$ is the inverse of $\alpha(t)$. Similarly for $t_1, t_2 \in T$,

$$\begin{aligned}
\alpha(t_1) * \alpha(t_2) &= \Big(\varphi(t_1), \psi(t_1)\Big) * \Big(\varphi(t_2), \psi(t_2)\Big) \\
&= \Big(\varphi(t_1) * \varphi(t_2), \psi(t_1) * \psi(t_2)\Big) \\
&= \Big(\varphi(t_1 * t_2), \psi(t_1 * t_2)\Big) \\
&= \alpha(t_1 * t_2)
\end{aligned}$$

This shows α respects products. □

As a corollary of Proposition 2.4.4 we get

Corollary 2.4.5. *Let m and n be positive integers. There is a unique homomorphism $\mathbb{Z}_{mn} \longrightarrow \mathbb{Z}_m \times \mathbb{Z}_n$ that takes $[1]_{mn}$ to $\big([1]_m, [1]_n\big)$. When m and n are coprime this homomorphism is an isomorphism.*

Proof. Exercise 2.3.13 shows that there is a unique homomorphism taking $[1]_{mn}$ to $[1]_m$ and a unique homomorphism taking $[1]_{mn}$ to $[1]_n$. The previous proposition says this pair of homomorphisms extends in a unique way to a homomorphism $\alpha : \mathbb{Z}_{mn} \longrightarrow \mathbb{Z}_m \times \mathbb{Z}_n$ that takes $[1]_{mn}$ to $\big([1]_m, [1]_n\big)$.

When m and n are coprime the kernel of α is trivial because if $\alpha([b]_{mn})$ is $\big([0]_m, [0]_n\big)$ then b must be divisible by both m and n. Since m and n are coprime b is divisible by mn, so $[b]_{mn} = [0]_{mn}$. □

The previous corollary is closely related to the Chinese Remainder Theorem 4.6.15 (which says a bit more than our corollary does). The following exercise broadens the perspective

Exercises 2.4.6. HOMOMORPHISMS AND THE DIRECT PRODUCT OF CYCLIC GROUPS We have shown that there is a homomorphism $\mathbb{Z}_n \longrightarrow \mathbb{Z}_d$ taking 1 in $\mathbb{Z}_n$ to 1 in $\mathbb{Z}_d$ if and only if $d \mid n$. Suppose c and d both divide n. Proposition 2.4.4 says that the two homomorphisms $\mathbb{Z}_n \longrightarrow \mathbb{Z}_d$ (taking $[1]_n$ to $[1]_d]$) and $\mathbb{Z}_n \longrightarrow \mathbb{Z}_c$ (taking $[1]_n$ to $[1]_c]$) give rise to a homomorphism $\mathbb{Z}_n \longrightarrow \mathbb{Z}_c \times \mathbb{Z}_d$.

(a) What is the kernel of the homomorphism $\mathbb{Z}_n \longrightarrow \mathbb{Z}_c \times \mathbb{Z}_d$?

(b) Under what conditions is it injective?

(c) Under what conditions is it surjective?

(d) Illustrate with $n = 8$ and $c = d = 4$. What is the image?

(e) Illustrate with $n = 18$ and $c = 6$ and $d = 9$. What is the image?

Automorphism Groups

We have already noted, in Section 1.2, that the bijections of a set form a group using composition of functions as the group operation. The identity element is the function that takes each element to itself, each bijection has an inverse that is a bijection, and the composition of two bijections is a bijection. We will study bijections of a finite set in the next section. For now, we are interested in bijections of a group to itself that also happen to be homomorphisms.

Proposition 2.4.7 (Automorphisms)**.** *Let G be a group. The set of all isomorphisms from G to itself is a group under composition. This new group is called* $\operatorname{Aut}(G)$*, the group of* **automorphisms** *of G.*

Proof. The identity map id_G is clearly an automorphism of G, so there is at least one automorphism of G. The composition of id_G with any automorphism $\varphi : G \longrightarrow G$ is φ, since the identity map takes each element to itself. Proposition 2.3.11 shows that the inverse of an isomorphism is an isomorphism and the composition of two isomorphisms is an isomorphism. Thus $\operatorname{Aut}(G)$ is a subgroup of the group of bijections of G. □

Exercises 2.4.8. Automorphism groups of some cyclic groups

(a) Show that $\mathrm{Aut}(\mathbb{Z})$ has two elements and $\mathrm{Aut}(\mathbb{Z}) \cong \mathbb{Z}_2$.

(b) Compute $\mathrm{Aut}(\mathbb{Z}_n)$ for $n = 2, 3, 4, 5, 6, 7$. [In each case the answer is a cyclic group.]

(c) Show that $\mathrm{Aut}(\mathbb{Z}_8)$ is not cyclic.

Exercises 2.4.9. Automorphism group of the simplest non-cyclic groups

(a) Show that there is a unique homomorphism of $\mathbb{Z}_2 \times \mathbb{Z}_2$ that swaps $(1, 0)$ with $(0, 1)$. What does it do to the other elements of $\mathbb{Z}_2 \times \mathbb{Z}_2$?

(b) Show that there are two different homomorphisms of $\mathbb{Z}_2 \times \mathbb{Z}_2$ that satisfy the following property: the only element that gets mapped to itself is $(0, 0)$.

(c) Can you extend to $\mathbb{Z}_3 \times \mathbb{Z}_3$? [Think of matrices.]

(d) Can you extend to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$? [Think of matrices.]

2.5 Permutation Groups

In this section we delve more deeply into the structure of the **symmetric group** S_n, the group of permutations of $\{1, \ldots, n\}$. The number of elements in S_n is $n!$. Informally, we may justify this claim by noting that there are n possible images for the number 1. Once the image for 1 is chosen, there are $n - 1$ choices for the number 2. Continuing in this manner we count $n!$ bijections from $\{1, \ldots, n\}$ to itself. One can give a more formal inductive proof.

We will sometimes write an element π of S_n in **tabular form** with i in the top row and $\pi(i)$ in the bottom row.

Exercises 2.5.1. Some computations in S_n

(a) Here are two elements of S_5:

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 1 & 2 & 4 \end{pmatrix} \quad \text{and} \quad \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 2 & 5 \end{pmatrix}.$$

(b) Compute the inverse of each.

(c) Compute the products $\pi\sigma$ and $\sigma\pi$, using the usual convention for compositions: $(\pi\sigma)(i) = \pi(\sigma(i))$. You should see that the results are not equal.

Example 2.5.2. Let $n = 3$, and enumerate the vertices of a triangle clockwise as 1, 2, 3. Each element of D_3 gives rise to a permutation of $\{1, 2, 3\}$.

Let r be rotation clockwise by $2\pi/3$. Then

$$r = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \text{and} \quad r^2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

There are three reflections, each fixes one element of $\{1,2,3\}$ and transposes the other two

$$u_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \qquad u_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \qquad u_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}.$$

This exhausts all permutations of $\{1,2,3\}$ so by enumerating the vertices of the triangle we have established a bijection between D_3 and S_3. This is actually an isomorphism since the operation for D_3 is composition, as it is for S_n.

Exercises 2.5.3. Embeddings into S_n

(a) How many ways are there to embed $\mathbb{Z}_4$ in S_4?

(b) How many ways are there to embed D_4 in S_4?

Cycle Decomposition

Definition 2.5.4. Let $a_1, a_2, \ldots, a_t$ be distinct elements of $\{1, \ldots, n\}$. We use the notation $(a_1, a_2, \ldots, a_t)$ to define an element of S_n called a t-**cycle**. This permutation takes a_i to a_{i+1}, for $i = 1, 2, 3 \ldots, t-1$ and it takes a_t to a_1. Every element of $\{1, \ldots, n\} \setminus \{a_1, \ldots, a_t\}$ is fixed (i.e. taken to itself) by the cycle $(a_1, a_2, \ldots, a_t)$. We will call the set $\{a_1, \ldots, a_t\}$ the **support** of the cycle $(a_1, a_2, \ldots, a_t)$.

A two-cycle is often called a **transposition**.

Two cycles are called **disjoint** when their supports are disjoint sets.

When we use cycle notation we will use id for the identity permutation.

Exercises 2.5.5. Properties of cycles

(a) Show that disjoint cycles commute.

(b) Suppose σ is a t-cycle. For which r is σ^r a t-cycle? What can happen for other r?

(c) Show that for a t-cycle σ, there is an injective homomorphism $\mathbb{Z}_r \longrightarrow S_n$ taking 1 to σ.

Definition 2.5.6. Let $\pi \in S_n$. The **orbit** of $a \in \{1, \ldots, n\}$ under π is the set $\left\{\pi^i(a) : i \in \mathbb{Z}\right\}$.

Let $\pi \in S_n$. A **cycle decomposition** for π is a product of disjoint cycles that is equal to π.

We want to show every permutation has a unique cycle decomposition. The first step is this lemma.

Lemma 2.5.7. *Let $\pi \in S_n$. Any two orbits of π are either equal or disjoint.*

Proof. Suppose two orbits of $\pi \in S_n$ are not disjoint. We will show they are equal. Let a, b, c be distinct elements of $\{1, 2, \ldots, n\}$. Suppose that b is in the orbit of a and also in the orbit of c. We will show that $\operatorname{orb}(a) = \operatorname{orb}(c)$. We have assumed $\pi^i(a) = b$ and $\pi^j(c) = b$ for some $i, j \in \mathbb{Z}$. Then $\pi^i(a) = \pi^j(c)$ so $\pi^{i-j}(a) = c$ so c is in $\operatorname{orb}(a)$. Moreover, anything in the orbit of c must be in the orbit of a since $\pi^k(c) = \pi^{k+i-j}(a)$. The reverse is also true by the same reasoning, so $\operatorname{orb}(a) = \operatorname{orb}(c)$. □

Proposition 2.5.8. *Every permutation in S_n has a cycle decomposition, and it is unique up to reordering the factors.*

Proof. This is just a sketch that should make sense, and one could formalize it using induction. Take an element $a \in \{1, \ldots, n\}$. Since $\{1, \ldots, n\}$ is finite, there is some pair of distinct positive integers such that $\pi^i(a) = \pi^j(a)$. Notice that $\pi^i(a) = \pi^j(a)$ implies $\pi^{i-1}(a) = \pi^{j-1}(a)$ and so forth until $\pi^{i-j}(a) = a$. Thus, there is some minimal positive integer, call it d, such that $\pi^d(a) = a$. It should be clear that, for $m \in \mathbb{Z}$, $\pi^m(a) = \pi^r(a)$ for r the remainder when m is divided by d. Now consider the cycle $(a, \pi(a), \ldots, \pi^{r-1}(a))$. The orbit of a is this set of elements, $\operatorname{orb}(a) = \big\{a, \pi(a), \ldots, \pi^{r-1}(a)\big\}$. Consequently, π can be written as the product of $(a, \pi(a), \ldots, \pi^{r-1}(a))$ and some other permutation that fixes each element in $\operatorname{orb}(a)$. Now choose an element of $\{1, \ldots, n\} \setminus \operatorname{orb}(a)$ and look at its orbit; continue. □

A permutation $\pi \in S_n$ may take an element to itself. We say π **fixes** $a \in \{1, \ldots, n\}$ when $\pi(a) = a$. We can denote this one-cycle by (a). Usually we don't write the one cycles in the cycle decomposition if the context is clear. For example, $(1, 3, 5)(2, 7)$ as an element of S_7 is really $(1, 3, 5)(2, 7)(4)(6)$.

Definition 2.5.9. We will call the list of cycle lengths, in decreasing order, the **signature** of the permutation.

For example, the permutation π in S_5 from Exercise 2.5.1 has cycle decomposition $\pi = (1, 3)(2, 5, 4)$ and signature $3, 2$. If we consider π as an element of S_6, we have $\pi = (1, 3)(2, 5, 4)(6)$ and the signature is $3, 2, 1$. The permutation $(1, 3, 5)(2, 7)$ as an element of S_7 has signature $3, 2, 1, 1$.

Exercises 2.5.10. The signature of a permutation

Prove the following results about the signature of a permutation.

(a) For $\pi \in S_n$, the sum of the signature list is n.

(b) If $\pi = \sigma_1 \sigma_2 \cdots \sigma_r$ is a cycle decomposition, then $\pi^k = \sigma_1^k \sigma_2^k \cdots \sigma_r^k$. Under what conditions is this also a cycle decomposition in the sense that each σ_i^k is a cycle?

(c) The order of $\pi \in S_n$ is the lcm of the signature list.

Exercises 2.5.11. A sanity check on symmetric groups

For $n = 4$ and $n = 5$ do the following to check that all elements of S_n are accounted for.

(a) Identify all possible signatures for elements of S_n and find the order of an element with the given signature.

(b) For each possible signature in S_n, count how many elements have that signature. Then check that you get the correct total number of elements in S_n.

(c) What is the exponent of S_n for $n = 4, 5, 6, 7$?

Transpositions and the Alternating Group

There is another factorization that is important.

Proposition 2.5.12. *Every permutation can be written as a product of transpositions.*

Proof. Since every permutation is a product of cycles, it is enough to show that every cycle is a product of transpositions. This is shown by verifying that

$$(a_1, a_2, \dots, a_t) = (a_1, a_2) * (a_2, a_3) * \cdots * (a_{t-2}, a_{t-1}) * (a_{t-1}, a_t)$$

Recall that we treat permutations are functions and we apply the rightmost permutation first. One can see that a_t gets mapped to a_{t-1} then a_{t-2} and so forth, until the final transposition is applied and takes a_2 (the image of a_t at this point) to a_1. Similar arguments apply to the other a_i. □

We may interpret the previous result as saying that S_n is generated by transpositions. That is somewhat good news: there are $n!$ elements of S_n but only $\binom{n}{2}$ transpositions. Thus $n(n-1)/2$ elements of S_n are enough to generate S_n. In fact we can do much better!

Exercises 2.5.13. Generators for S_n

(a) Show that S_n is generated by the $n - 1$ elements $(1, k)$ for $k = 2, \dots, n$. [Show that you can get an arbitrary transposition by conjugating $(1, k)$ by some $(1, j)$, see Definition 2.7.10.]

(b) Show that S_n is generated by 2 elements: $(1, 2)$ and $(1, 2, 3, \dots, n - 1, n)$. [Show that you can get all $(1, k)$ from these two using conjugation and then apply the previous exercise.]

Exercises 2.5.14. Preparing for the next proposition

(a) Let a, b, c be distinct elements of $\{1, \ldots, n\}$. Write down all possible factorizations of the 3-cycle (a, b, c) as a product of 2 transpositions.

(b) Let a be an element of $\{1, \ldots, n\}$. Let τ_1, τ_2 be transpositions in S_n with $\tau_1 \neq \tau_2$. Show that there exist transpositions $\sigma_1, \sigma_2 \in S_n$ such that $\sigma_1\sigma_2 = \tau_1\tau_2$ and a is not in the support of σ_2. [You will need to consider a few different cases depending on whether a is in the support of τ_1 or τ_2.]

We know from the previous proposition that a permutation can be written as a product of transpositions. This "factorization" is not unique, for example $\mathrm{id} = (1, 2)(1, 2) = (1, 3)(1, 3)$, but the next proposition shows that the parity of the factorization is.

Proposition 2.5.15. *The identity element of S_n cannot be written as the product of an odd number of transpositions.*

Consequently, no permutation can be written as a product of an even number of transpositions and also as a product of an odd number of transpositions.

Proof. We will show that if id is the product of n transpositions then it is the product of $n-2$ transpositions. Consequently, if it is the product of an odd number of transpositions, inductively we could show that id is a single transposition. This is clearly false.

Suppose that $\mathrm{id} = \tau_1 \cdots \tau_m$ with $\tau_i = (b_{2i-1}, b_{2i})$. The b_i are not necessarily distinct, except $b_{2i} \neq b_{2i-1}$ so that τ_i is indeed a transposition. Let $a = b_1$. Let k be the largest integer such that a is in the support of τ_k (so either b_{2k-1} or b_{2k} is equal to a). Note that $k \neq 1$ because if a was only in the support of τ_1 then $\tau_1 \cdots \tau_m(a) = \tau_1(b_1) = b_2 \neq a$ and the factorization would not be the identity.

Using the previous exercise we can rewrite the factorization of the identity replacing $\tau_{k-1}\tau_k$ with $\sigma_{k-1}\sigma_k$ in which a is not in the support of σ_k (the indexing of $k-1$ and k on σ_{k-1} and σ_k is just for notational convenience). We have a new factorization of id with m terms, but now, only the transpositions $\tau_1, \ldots, \tau_{k-2}$ and σ_{k-1} can have a in the support. If $\tau_{k-2} = \sigma_{k-1}$ we can cancel and get a shorter factorization of the identity using $n-2$ transpositions, as claimed. Otherwise we repeat the process: find the largest index such that the transposition with that index has a in the support; use the exercise to move a into a lower index term; cancel if possible; if not repeat. Eventually we either get a cancellation, or we arrive at a factorization $\tau_1\sigma_2 \cdots \tau_{k+1}\tau_{k+2} \cdots \tau_m$ in which only the first two transpositions τ_1 and σ_2 have a in their support. Then $\tau_1\sigma_2(a) = \mathrm{id}(a) = a$. This is possible only if $\tau_1 = \sigma_2$. Thus we may cancel and get id equal to the product of $m-2$ transpositions as claimed.

For the second part, suppose that π is the product of transpositions in two ways: $\pi = \sigma_1\sigma_2 \ldots \sigma_m = \theta_1\theta_2 \ldots \theta_k$. Then $\mathrm{id} = \sigma_1\sigma_2 \ldots \sigma_m\theta_1^{-1}\theta_2^{-1} \ldots \theta_k^{-1}$. Thus $m+k$ must be even, and this implies that m and k must have the same parity. □

We now have an important and easy consequence.

Proposition 2.5.16. *The set of even parity permutations forms a subgroup of S_n. This is called the* **alternating group** *and is denoted A_n.*

Furthermore, there is a homomorphism from S_n to $\mathbb{Z}_2$ whose kernel is A_n.

Proof. Define a function from f from S_n to $\mathbb{Z}_2$ that takes even permutations to 0 and odd permutations to 1. This is well defined by the proposition. Let π and σ have the following decompositions as products of transpositions $\pi = \tau_1 \cdots, \tau_k$, $\sigma = \tau_1' \cdots, \tau_m'$. Then $\pi\sigma$ has the following decomposition as a product of transpositions:

$$\pi\sigma = \tau_1 \cdots, \tau_k \tau_1' \cdots, \tau_m'$$

Consequently,

$$\begin{aligned} f(\pi\sigma) &= m + k \mod 2 \\ &= f(\pi) + f(\sigma) \end{aligned}$$

This shows f is a homomorphism.

By construction, the kernel of f is A_n, the set of even permutations. Since the kernel of a homomorphism is a subgroup of the domain, A_n is a subgroup of S_n.

□

Exercises 2.5.17. Parity of a t-cycle

(a) Show that the parity of a t-cycle is the *not* $t \mod 1$.

(b) Find a formula for the parity of $\sigma \in S_n$ that uses the signature of σ.

Exercises 2.5.18. Details for proving the simplicity of A_n

(a) Suppose that σ is a k-cycle and τ is an m-cycle and there is exactly one element of $\{1, \ldots, n\}$ that is in the support of both σ and τ. Show that $\sigma\tau$ is a $(k + m - 1)$-cycle.

(b) Show that the product of two disjoint transpositions can also be written as the product of two 3-cycles.

(c) Use part (a) (with $k = m = 2$) and part (b) to prove that A_n is generated by 3 cycles.

(d) Compute $(1, 2, a)(1, b, 2)$ for a, b distinct and not equal to 1 or 2. Use the result as motivation to show that the 3-cycles of the form $(1, 2, a)$ generate A_n for $n \geq 4$.

Exercises 2.5.19. More on A_n

(a) Find all subgroups of A_4. Draw a diagram of the subgroup lattice.

(b) What is the intersection of A_4 and D_4 (generated by $(1, 2, 3, 4)$ and $(1, 3)$?

(c) There are other subgroups of S_4 isomorphic to D_4, for example the group generated by $(1,3,4,2)$ and $(1,3)$. What is the intersection of these versions of D_4 with A_4?

Cayley's Theorem

For any set T, the set of bijections from T to itself forms a group under composition with the indentity map, $\mathrm{id}(t) = t$, acting as identity element and f mapping to the inverse of a function f as the operation of inversion. In a manner similar to our notation for permutations of $\{1, \ldots, n\}$, we will write the group of bijections of an arbitrary set T as S_T. The elements of S_T will also be called permutations of T. It should be clear that if $|T| = n$ then $S_T \cong S_n$.

For a group G, we can forget that G is a group and just look at arbitrary bijections (set maps) from G to itself, that is S_G. The next theorem shows there is an injective homomorphism from G to S_G.

Theorem 2.5.20. *Any group G is isomorphic to a subgroup of S_G, the group of (set) bijections of G to itself. If $|G| = n$ there is an embedding of G in S_n.*

Proof. For each $a \in G$, left multiplication by a maps elements of G to elements of G. Let us call this map $\lambda_a : g \longrightarrow ag$. We can see that λ_a is a permutation of G as follows. For any $g \in G$, $\lambda_a(a^{-1}g) = a(a^{-1}g) = g$, so $a^{-1}g$ is a preimage for g. Since g was arbitrary, λ_a is surjective. We also have λ_a is injective because $\lambda_a(g) = \lambda_a(g')$ implies $ag = ag'$, which by cancellation in G gives $g = g'$.

(We could also prove that λ_a is a permutation of G by showing that $\lambda_{a^{-1}}$ and λ_a are inverse functions of each other: both compositions, $\lambda_a \circ \lambda_{a^{-1}}$ and $\lambda_{a^{-1}} \circ \lambda_a$ give the identity map.)

Define $\lambda : G \longrightarrow S_G$ by $\lambda : a \longrightarrow \lambda_a$. Since $\lambda_a(e) = a$, we have $\lambda_a = \lambda_b$ can only be true if $a = b$. Thus λ is injective. To show it is a homomorphism we have to show that $\lambda_{ab} = \lambda_a \circ \lambda_b$. The following computation does that. We have for all $g \in G$,

$$\lambda_{ab}(g) = (ab)g = a(bg) = \lambda_a(bg) = \lambda_a(\lambda_b(g)) = (\lambda_a \circ \lambda_b)(g)$$

For a finite set T of cardinality n, we noted earlier that S_T is isomorphic to S_n. It is worth explicitly giving a construction of an embedding of G into S_n (for $|G| = n$) by enumerating the elements of G so $G = \{g_1, \ldots, g_n\}$. For any $a \in G$, we have shown that λ_a, left multipication by a, permutes the elements of G. Define $\varphi_a \in S_n$ by $\varphi_a(i)$ is the unique j such that $ag_i = g_j$. We may then write $ag_i = g_{\varphi_a(i)}$. Observe that

$$g_{\varphi_{ab}(i)} = abg_i = a(bg_i) = a(g_{\varphi_b(i)}) = g_{\varphi_a(\varphi_b(i))} = g_{\varphi_a \circ \varphi_b(i)}$$

This shows that $\varphi_{ab} = \varphi_a \circ \varphi_b$ so the function φ is a homomorphism.

□

Exercises 2.5.21. Cayley's Theorem: three examples

(a) Let $n = 5$ and think of $\mathbb{Z}_n$ in the usual way as $\{0, 1, 2, 3, 4\}$ with addition modulo n. For each $a \in \mathbb{Z}_n$ write down in tabular form the function on $\mathbb{Z}_n$ defined by addition of a.

(b) Show that part (a) defines a function from $\mathbb{Z}_5$ to S_5, provided you think of S_5 as the group of permutations of $\{0, 1, 2, 3, 4\}$. Show that this function is a homomorphism.

(c) Now consider $\mathbb{Z}_2 \times \mathbb{Z}_2$. Enumerate the 4 elements in any way you choose as a_1, a_2, a_3, a_4. For each a_i define a permutation σ_i by $a_i a_1 = a_{\sigma_i(1)}$, $a_i a_2 = a_{\sigma_i(2)}$, $a_i a_3 = a_{\sigma_i(3)}$, $a_i a_4 = a_{\sigma_i(4)}$.

(d) Show in part (c) that this gives a homomorphism from $\mathbb{Z}_2 \times \mathbb{Z}_2$ to S_4.

(e) Similarly, the next steps define a homomorphism from D_3 to S_6. Enumerate the elements of as follows

$$D_3 = \left\{a_1 = r^0, a_2, = r, a_3 = r^2, a_4 = t, a_5 = rt, a_6 = r^2 t\right\}$$

For each a_i define a permutation σ_i in S_6. Since a_1 is the identity in D_3, σ_1 is the identity permutation in S_3. One can see that σ_2 is given by $\sigma_2(i) = k$ whenever $ra_i = a_k$. Verify that each σ_i is indeed a permutation by writing it in permutation notation.

(f) Verify in three examples that for any $a, b \in D_3$, the permutation corresponding to ab equals the product of the permutations corresponding to a and b.

(g) Which elements of D_3 correspond to odd permutations in S_6?

Exercises 2.5.22. The product of two symmetric groups

(a) Let A and B be disjoint subsets of $\{1, \ldots, n\}$. Explain how to think of $S_A \times S_B$ as a subgroup of S_n.

(b) Generalize to any partition of $\{1, \ldots, n\}$.

Exercises 2.5.23. Counting elements of S_n

(a) Let n be a positive integer and $k > n/2$. Find a formula for the number of elements of S_n that include a k-cycle.

(b) Use Stirling's formula to approximate the formula you just computed.

(c) Estimate the probability that a random element of S_n has a cycle of length larger than $n/2$.

2.6 Generators and Relations

There is another way to describe groups: using "generators" and "relations." A set of generators for the group is given, and then certain properties that must be satisfied by those generators are listed. The latter are called the relations. Identifying the generators and relations for a group is called **giving a presentation** of the group. The group is then the set of all possible products of the generators and their inverses (these are informally called "words"). The relations say that some words are equal to the identity, so they give a way to simplify words. The description via generators and relations can be straightforward in the simplest instances, but it is quite subtle in general. We give a few examples here and defer a more thorough treatment to Section 3.4.

We have noted that $\mathbb{Z}$ requires only one generator, as do the groups $\mathbb{Z}_n$. We may describe $\mathbb{Z}_n$ (actually a group isomorphic to it) using the generator a, and the relation $a^n = 1$, in which I use 1 for the identity element. This would be written

$$\langle a | a^n = 1 \rangle$$

The homomorphism $\mathbb{Z}_n \longrightarrow \langle a | a^n = 1 \rangle$ taking 1 to a is clearly an isomorphism.

We could also describe $\mathbb{Z}_n$ in other ways. For example when p and q are distinct primes (or even just coprime to each other), we could use two generators.

$$\mathbb{Z}_{pq} \cong \langle a, b | a^p = 1, b^q = 1, ab = ba \rangle$$

This tells us that a is an element of order p, b is an element of order q and that a and b commute. The latter relation could also be written $aba^{-1}b^{-1} = 1$.

Exercises 2.6.1. A PRESENTATION OF $\mathbb{Z}_{pq}$

(a) Show that the presentation for $\mathbb{Z}_{pq}$ in the previous paragraph is isomorphic to the group with presentation $\langle c | c^{pq} = 1 \rangle$ via the function $c \to ab$.

(b) Show that the same presentation works provided only that p and q are coprime.

(c) Find a presentation for $\mathbb{Z}_{pqr}$ with p, q, r pairwise coprime.

The dihedral group D_n has a presentation as follows

$$\langle a, b | a^n = 1, b^2 = 1, ba = a^{n-1}b \rangle$$

The generator a is clearly playing the role of rotation by $2\pi/n$ and b the role of a reflection. The final relation tells us that in any product using a and b we can switch any occurrence of ba to be $a^{n-1}b$ and thereby rearrange so that all the as are on the left and all the bs on the right. So, just using these relations we know that any element of this group can be uniquely written $a^i b^k$ for $i \in \{0, \ldots, n-1\}$ and $k \in \{0,1\}$. The group product is easily summarized by the following

$$\begin{aligned} (a^i)(a^j b^k) &= a^{i+j} b^k \\ (a^i b)(a^j b^k) &= a^{i-j} b^{k+1} \end{aligned}$$

Exponents on a are computed modulo n and on b modulo 2. It is easy to verify that our earlier discussion of D_n (Section 1.2) is consistent with the description here: for any reflection t_i, $t_i r = r^{-1} t_i$.

Exercises 2.6.2. THE **Quaternion Group**

Consider the presentation

$$Q = \langle a, b \mid a^4 = 1, b^2 = a^2, ba = a^{-1} b \rangle$$

(a) Show that Q has 8 elements. List them in a useful fashion and show how to multiply them as we did for the dihedral group.

(b) Find the order of each element of Q.

(c) Draw the lattice diagram for this group.

(d) Show that no two of the groups $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, $\mathbb{Z}_8$, D_4, and Q are isomorphic. [Investigate the number of elements of order 4. Or, Compare lattice diagrams.]

Exercises 2.6.3. THE INFINITE DIHEDRAL GROUP

Let

$$D_\infty = \langle a, b \mid b^2 = 1 \rangle$$

(a) Show that D_∞ is a symmetry group of the following diagrams.

(b) What other symmetries do the diagrams have that are not captured by D_∞?

2.7 Cosets and Conjugates

Let H be a subset of G. In this section we define cosets of H and show an extremely important result (Lagrange's Theorem) that the cosets form a partition of G. We also show there is another interesting partition of G that is determined by a relation called conjugacy.

The following bit of notation is useful.

Notation 2.7.1. Let S and T be subsets of a group G.

$$ST = \{st : s \in S, t \in T\}$$

So, ST is the set of all products of an element in S (on the left) and an element of T (on the right). Similarly, for $g \in G$, $gS = \{gs : s \in S\}$. We may use analogous notation for the set of all products from 3 or more sets.

If a group is abelian and the operation is $+$ we write $S + T$ instead of ST.

Proposition 2.7.2. *Associativity holds for the notation in 2.7.1. If S, T, U are subsets of G then $(ST)U = S(TU)$.*

Proof.

$$\begin{aligned}(ST)U &= \{st : s \in S, t \in T\}\, U \\ &= \{(st)u : s \in S, t \in T, u \in U\} \\ &= \{s(tu) : s \in S, t \in T, u \in U\} \\ &= S\,\{tu : t \in T, u \in U\} \\ &= S(TU)\end{aligned}$$

□

The proposition shows that we may write unambigously STU for the product of three sets (taken in the order given). Notice that ST and TS are not necessarily equal when the group G is not abelian.

Exercises 2.7.3. Revisiting the properties of a subgroup.

(a) Let T be a nonempty subset of a group G. Prove that T is a subgroup of G if and only if $TT = T$ and $T^{-1}T = T$.

(b) Let T be a nonempty subset of the finite group G. Prove that $TT = T$ if and only if T is a subgroup of G.

(c) Give an example to show that for an arbitrary group G and nonempty subset T, $TT = T$ is *not sufficient* to ensure T is a subgroup of G.

Definition 2.7.4. Let $H \leq G$ and let $g \in G$. The set gH is a **left coset** of H in G. Similarly, Hg is a **right coset** of H in G.

We will prove several results for left cosets. There are analogous results for right cosets.

Lemma 2.7.5. *Let G be a group and H a subgroup of G. The function*

$$\begin{aligned} \lambda_g : H &\longrightarrow gH \\ h &\longmapsto gh \end{aligned}$$

is a bijection.

Proof. The function λ_g is a surjection by the definition of gH. Suppose $gh = gh'$. Multiplying on the left by g^{-1} gives $h = h'$. This shows λ_g is injective. □

Lemma 2.7.6. *Let G be a group and H a subgroup of G. For $a, g \in G$, if $gH \cap aH \neq \emptyset$ then $gH = aH$.*

Proof. First, we show that if $g \in aH$ then $gH \subseteq aH$. Let $g \in aH$, so there is some $k \in H$ such that $g = ak$. For any $h \in H$, we have $gh = akh$. This is an element of aH because $kh \in H$ since H is a subgroup of G. This shows $gH \subseteq aH$.

Suppose $gH \cap aH$ is nonempty, containing some element x. Then there are $h, k \in H$ such that $x = gh = ak$. Then $g = akh^{-1} \in aH$ and similarly $a = ghk^{-1} \in gH$. From the previous paragraph, we have $aH \subseteq gH$ and $gH \subseteq aH$, so $aH = gH$. □

Theorem 2.7.7 (Lagrange)**.** *Let G be a group with subgroup H. The set of cosets of H form a partition of G.*

Consequently, if G is a finite group with subgroup H then the order of H divides the order of G. In particular, the order of any element of G divides $|G|$.

Proof. Any $g \in G$ is in some coset, namely gH, so the cosets cover G. The previous lemma shows that any two unequal cosets are disjoint. Thus the cosets partition G.

Suppose G is finite. Since the cosets of H partition G, there are elements $a_1, \dots, a_t$ such that G is the disjoint union of $a_1H, a_2H, \dots, a_tH$. The cosets of H all have the same number of elements by Lemma 2.7.5. Thus $|G| = \sum_{i=1}^{t} |a_iH| = t|H|$, and the number of elements of G is a multiple of $|H|$.

For any $a \in G$ the number of elements in the subgroup $\langle a \rangle$ is $\mathrm{ord}(a)$. So $\mathrm{ord}(a)$ divides $|G|$. □

Definition 2.7.8. Let $H \leq G$. The **index** of H in G, written $[G : H]$, is the number of cosets of H in G, which may be infinite.

If G is finite and $H \leq G$ then $[G : H] = |G|/|H|$, since all cosets have $|H|$ elements.

Exercises 2.7.9. The exponent of a group

Recall that the exponent of a group G is the lcm of the orders of the elements (if this is finite).

(a) Give an example to show that there may not be an element in G whose order is the exponent of G, even if G is finite.

(b) Let G be a finite group with finite exponent t. Show that t divides the order of G.

Conjugation

Now we consider conjugation.

Definition 2.7.10. Let $a \in G$ and $g \in G$. The element aga^{-1} is called the **conjugation of** g **by** a. If S is a subset of G, we define aSa^{-1} to be $\{asa^{-1} : s \in S\}$. It is the **conjugation of** S **by** a.

Proposition 2.7.11. *Conjugacy on a group G determines an equivalence relation.*

Proof. Let G be a group and define a relation on G by a is related to b if there is some g such that $b = gag^{-1}$. The relation is reflexive, because for any $a \in G$, $eae^{-1} = a$. The relation is symmetic, because if a is related to b (say $b = gag^{-1}$) then we also have $a = g^{-1}b(g^{-1})^{-1}$ so b is related to a. Finally suppose a is related to b (again $b = gag^{-1}$) and b is related to c (so there is some $h \in G$ with $c = hbh^{-1}$). Then

$$c = hbh^{-1} = h(gag^{-1})h^{-1} = (hg)a(hg)^{-1}$$

This shows that a is related to c. We have shown that conjugacy determines an equivalence relation on G. □

Exercises 2.7.12. The **Centralizer** of an element

For $a \in G$ we define the centralizer of a to be $C(a) = \{g \in G : ga = ag\}$.

(a) Show that $C(a)$ is a subgroup of G.

(b) Let G be a finite group. Show that the number of elements of G conjugate to a is $|G|/|C(a)|$. [Consider the cosets of $C(a)$.]

Exercises 2.7.13. Conjugation in the symmetric and dihedral groups

(a) Show that A_n is invariant under conjugation: for any $\pi \in S_n$, $\pi A_n \pi^{-1} = A_n$.

(b) Let C_n be the rotation subgroup of D_n. Find two elements of C_4 that are conjugate as elements of D_4 but are not conjugate as elements of C_4.

(c) Find two elements of D_4 that are conjugate as elements of S_4 but are not conjugate as elements of D_4.

Exercises 2.7.14. Conjugates and subgroups.

Let H be a subgroup of a group G.

(a) Let $a \in G$. Show that aHa^{-1} is a subgroup of G.

(b) Show that there is an isomorphism between H and aHa^{-1}.

Proposition 2.7.15. *Let $\pi \in S_n$. For any $\sigma \in S_n$, the signature of σ and the signature of $\pi\sigma\pi^{-1}$ are the same.*

One proof is contained in the following suite of exercises.

Exercises 2.7.16. Conjugation and cycle decomposition.

Consider conjugation by $\pi \in S_n$.

(a) Let $(a_1, a_2, \dots, a_k) \in S_n$ be a k-cycle, so the a_i are distinct. Show that

$$\pi * (a_1, a_2, \dots, a_k) * \pi^{-1} = \Big(\pi(a_1), \pi(a_2), \dots, \pi(a_k)\Big)$$

[Consider two cases, $b = \pi(a_i)$ for some i, and $b \notin \{\pi(a_1), \pi(a_2), \dots \pi(a_k)\}$. Explain why this breakdown into two cases makes sense.]

(b) If A and B are disjoint subsets of $\{1, \dots, n\}$ show that $\pi(A)$ and $\pi(B)$ are also disjoint.

(c) If $\sigma = \sigma_1\sigma_2\cdots\sigma_k$ is the cycle decomposition of σ, find the cycle decomposition of $\pi\sigma\pi^{-1}$ and justify your answer.

(d) Conclude that the conjugation of any $\sigma \in S_n$ by π has the same signature as σ.

2.8 Normality and the First Isomorphism Theorem

Let's pause for a moment to think about homomorphisms, particularly the special cases: injective homomorphisms (one-to-one) and surjective homomorphisms (onto).

Suppose H is a subgroup of G. There is a injective function from H to G, which is called the **inclusion map**, that simply takes $h \in H$ to itself, as an element of G. Since H is a subgroup of G (it's multiplication is the same as the one on G), the inclusion map is an injective homomorphism from H to G.

On the other hand, suppose that H and G are arbitrary groups and that $\varphi : H \longrightarrow G$ is an injective homomorpism. Proposition 2.3.5 shows that $\varphi(H)$ is a subgroup of G. Thus the bijection $\varphi : H \longrightarrow \varphi(H)$ is actually a homomorphism of groups. This shows that the image of an injective homomorphism $\varphi : H \longrightarrow G$ is a subgroup of G that is isomorphic to H. Thus, the study of injective homomorphisms is essentially the study of subgroups and their automorphisms.

This section and Section 2.10 are focused on surjective homomorphisms, which are intimately related to subgroups that have a special property, treated in the next proposition.

Theorem 2.8.1 (Normal Subgroups)**.** *Let N be a subgroup of G. The following are equivalent.*

(1) $Na = aN$ for all $a \in G$.

(2) $aNbN = abN$ for all $a, b \in G$.

(3) $aNa^{-1} \subseteq N$ for all $a \in G$.

(4) $aNa^{-1} = N$ for all $a \in G$.

Proof. We prove a series of implications that shows the conditions are equivalent.
(1) $\implies$ (2): Assume $Na = aN$ for all $a \in G$. Then

$$(aN)(bN) = a(Nb)N = a(bN)N = (ab)N$$

Here we have used asociativity, then the assumption in (1), and finally, $NN = N$ since N is a subgroup of G.
(2) $\implies$ (3): Assume $aNbN = abN$ for all $a, b \in G$. Set $b = a^{-1}$. Then, using (2), $aNa^{-1}N = aa^{-1}N = eN$. In particular, this shows that $aNa^{-1} \subseteq N$.
(3) $\implies$ (4): Assume $aNa^{-1} \subseteq N$ for all $a \in G$. For any $a \in G$, applying (3) to a^{-1}, we have that $a^{-1}Na \subseteq N$. Conjugating by a, we get

$$a(a^{-1}Na)a^{-1} \subseteq aNa^{-1}$$

The left hand side is N. Thus, assuming (3) we have both $aNa^{-1} \subseteq N$ and $N \subseteq aNa^{-1}$, which proves (4).
(4) $\Longrightarrow$ (1): Assume $aNa^{-1} = N$ for all $a \in G$. Multiplying $aNa^{-1} = N$ on the right by a gives (1).

These arguments may seem slippery since they involve computations with sets. The proofs can also be done elementwise. Consider (4) $\Longrightarrow$ (1). Let $a \in G$. Given any $n \in N$, we know $ana^{-1} \in N$, say $ana^{-1} = n'$. Then $an = n'a \in Na$. Since n was arbitrary, $aN \subseteq Na$. The reverse containment is proven analogously, using $a^{-1}na \in N$. □

Definition 2.8.2. A group satisfying the conditions of the theorem is called **normal**. We write $N \trianglelefteq G$ for N a normal subgroup of G.

Exercises 2.8.3. A WEAKER REQUIREMENT FOR NORMALITY

(a) Let H be a subgroup of a group G such that for any $a \in G$ there is a b in G such that $aH = Hb$. (Every left coset is also a right coset, but not necessarily defined by the same element of G.) Prove that H is normal in G.

Let N be normal in G. Suppose $aN = bN$ and $rN = sN$. Then $a \in bN$ and $r \in sN$, so $ar \in bNsN = bsN$. By Lemma 2.7.6, $arN = bsN$. Consequently, there is a well-defined operation on cosets of N in G that takes the pair (aN, bN) to abN (it doesn't matter which element we choose to represent each coset, their product always defines the same coset). The next theorem shows that this gives a group structure on the cosets of N in G.

Theorem 2.8.4. *Let N be a normal subgroup of G. Let G/N be the set of cosets of N in G with the binary operation by $aN * bN = abN$. Then G/N is a group.*

Proof. We have proven above that the product $aNbN$ is well defined and equal to abN. Associativity is inherited from associativity of $*_G$ (check!). The identity is eN. The inverse of aN is $a^{-1}N$. □

We call G/N the **quotient** of G by N and the homomorphism $G \longrightarrow G/N$ is called the **quotient map**. Some sources call G/N a **factor** group.

Every subgroup of an abelian group A is normal in A, so for any subgroup B of A there is quotient group A/B.

Example 2.8.5. In $\mathbb{Z}$ the only subgroups are $n\mathbb{Z}$. The quotient group $\mathbb{Z}/n\mathbb{Z}$ has the distinct elements $a + n\mathbb{Z}$ for $a \in \{0, \ldots, n-1\}$. Clearly this is just another way to think about the additive group of integers modulo n. It is isomorphic to $\mathbb{Z}_n$.

Exercises 2.8.6. ADDITIONAL PROPERTIES OF NORMAL SUBGROUPS.

(a) Let N be a normal subgroup of G. For any subgroup H of G, $H \cap N$ is a normal subgroup of H.

(b) If $\varphi : G \longrightarrow H$ is a homomorphism and N is normal in H, then $\varphi^{-1}(N)$ is normal in G.

(c) Show that any subgroup of index 2 is normal.

Exercises 2.8.7. The Center of of a group

We defined $Z(G)$, the center of G to be the set of elements in G that commute with all elements of G, $Z(G) = \{a \in G : ag = ga \text{ for all } g \in G\}$.

(a) Prove that any subgroup of the center of G, including $Z(G)$ itself, is normal in G.

(b) Find the center of D_4; it is not trivial.

(c) Show that the centers of D_3 and D_5 are trivial.

Exercises 2.8.8. Example of normal subgroups.

(a) Find all normal subgroups of D_4, D_5, and D_6.

(b) Find all normal subgroups of A_4.

(c) Find all normal subgroups of the quaternions, Q.

Isomorphism and Factor Theorems

We are now in the position to say more about the relationship between homomorphisms and normal subgroups.

Theorem 2.8.9 (First Isomorphism)**.** *Let $\varphi : G \longrightarrow H$ be a surjective homomorphism with kernel K. Then K is a normal subgroup of G and G/K is isomorphic to H.*

Proof. We showed in Proposition 2.3.7 that the kernel of any homomorphism of groups is a normal subgroup of the domain.

Let $g \in aK$, so $g = ak$ for some $k \in K$. Then $\varphi(g) = \varphi(a)\varphi(k) = \varphi(a)$. Consequently, all elements of a fixed coset of K ahve the same image under φ, so there is a well defined map $\tilde{\varphi} : G/K \to H$ taking aK to $\varphi(a)$.

To show $\tilde{\varphi}$ is a homomorphism, let aK and bK be elements of G/K. Since K is normal, $\tilde{\varphi}(aKbK) = \tilde{\varphi}(abK) = \varphi(ab)$ by the definition of multiplication in G/K and the definition of $\tilde{\varphi}$. Since φ is a homomorphism, $\varphi(ab) = \varphi(a)\varphi(b) = \tilde{\varphi}(aK)\tilde{\varphi}(bK)$. thus $\tilde{\varphi}(aKbK) = \tilde{\varphi}(aK)\tilde{\varphi}(bK)$, which shows $\tilde{\varphi}$ is a homomorphism.

Since φ is surjective, for any $h \in H$ there is some $a \in G$ such that $\varphi(a) = h$. Then $\tilde{\varphi}(aK) = h$, so $\tilde{\varphi}$ is surjective.

To show that $\tilde{\varphi}$ is injective, suppose $\tilde{\varphi}(aK) = e_H$. Then $\varphi(a) = e_H$ so $a \in K$ and $aK = e_G K$. Thus the kernel of $\tilde{\varphi}$ just contains just the identity element of G/K. □

Here is an important use of the First Isomorphism theorem.

Theorem 2.8.10. *Let $G_1, G_2, \ldots G_r$ be groups and let $N_1, N_2, \ldots, N_r$ be normal subgroups, $N_i \trianglelefteq G_i$. There is a well defined map*

$$\begin{aligned}(G_1 \times G_2 \times \cdots \times G_r)/(N_1 \times N_2 \times \cdots \times N_r) &\longrightarrow (G_1/N_1) \times (G_2/N_2) \times \cdots \times (G_r/N_r)\\ (g_1, g_2, \ldots, g_r)\,(N_1 \times N_2 \times \cdots \times N_r) &\longmapsto (g_1N_1, g_2N_2, \ldots g_rN_r)\end{aligned}$$

and it is an isomorphism.

Proof. By the Direct Product Theorem 2.4.2, the projection of $G_1 \times G_2 \times \cdots \times G_r$ onto G_i is a homomorphism. Composing this with the quotient map $G_i \longrightarrow G_i/N_i$, we get maps $G_1 \times G_2 \times \cdots \times G_r \longrightarrow G_i/N$. Proposition 2.4.4 then gives a homomorphism

$$G_1 \times G_2 \times \cdots \times G_r \xrightarrow{\varphi} (G_1/N_1) \times (G_2/N_2) \times \cdots \times (G_r/N_r),$$

To be specific, let us show that φ respects products. Let $(g_1, g_2, \ldots, g_r)$ and $(g'_1, g'_2, \ldots, g'_r)$ be elements of $G_1 \times G_2 \times \cdots \times G_r$. Then

$$\begin{aligned}\varphi\Big((g_1, g_2, \ldots, g_r) * (g'_1, g'_2, \ldots, g'_r)\Big) &= \varphi\Big(\,(g_1g'_1, g_2g'_2, \ldots, g_rg'_r)\,\Big)\\ &= (g_1g'_1N_1, g_2g'_2N, \ldots, g_rg'_rN_r)\\ &= (g_1N_1, g_2N_2, \ldots, g_rN_r) * (g'_1N_1, g'_2N_2, \ldots, g'_rN_r)\\ &= \varphi(g_1, g_2, \ldots, g_r) * \varphi(g'_1, g'_2, \ldots, g'_r)\end{aligned}$$

We used, in order, the definition of multiplication in $G_1 \times G_2 \times \cdots \times G_r$, the definition of φ, the definition of multiplication in $G_1/N_1 \times G_2/N/2 \times \cdots \times G_r/N_r$ (and the N_i being normal), and finally, the definition of φ.

The kernel of φ is the set of $(g_1, \ldots, g_r)$ such that $g_1N_1, g_2N_2, \ldots, g_rN_r = N_1 \times N_2 \times \cdots \times N_r$. Each g_i must be in N_i. So, the kernel is $e_1N_1 \times e_2N_2 \times \cdots \times e_rN_r$. Surjectivity is easy to check, so the first isomorphism theorem now gives the result. □

A generalization of the first isomorphism theorem that we will often use treats the case when $\varphi : G \longrightarrow H$ is not necessarily surjective.

Theorem 2.8.11 (Factor)**.** *Let $\varphi : G \longrightarrow H$ be a homomorphism of groups with kernel K. Let N be a normal subgroup of G that is contained in K. Then φ can be factored into the canonical surjective homomorphism $\pi : G \longrightarrow G/N$ followed by a homomorphism $\bar{\varphi} : G/N \longrightarrow H$.*

By letting $N = K$ we conclude that any homomorphism can be factored into a surjective homomorphism followed by an injective homomorphism.

Proof. Define $\tilde{\varphi}$ by $gN \longmapsto \varphi(g)$. This is well defined because N is contained in the kernel of φ so for any $n \in N$, $\varphi(gn) = \varphi(g)\varphi(n) = \varphi(g)e_H = \varphi(g)$. From this definition it is immediate that $\tilde{\varphi} \circ \pi = \varphi$.

The proof that $\tilde{\varphi}$ is a homomorphism is similar to the proof of the First Isomorphism Theorem.

When $N = K$, we want to show that $\tilde{\varphi}$ is injective. Suppose $\tilde{\varphi}(gN) = e_H$. By the definition of $\tilde{\varphi}$, we have $\varphi(g) = e_H$. Thus $g \in K$, and therefore $gK = eK$, the identity element of G/K. □

The Factor Theorem and First Isomorphism Theorem give us a framework for understanding the material that we have seen earlier. As we said above, the group $\mathbb{Z}_n$ is just the quotient of $\mathbb{Z}$ by its normal subgroup $n\mathbb{Z}$.In the context of groups we have given a shorthand notation to $\mathbb{Z}/n\mathbb{Z}$, calling it $\mathbb{Z}_n$.

Recall that Proposition 2.3.4 says that given any group G and $g \in G$ there is a homomorphism $\mathbb{Z} \longrightarrow G$ taking 1 to g. If g has infinite order then the cyclic group, $\langle g \rangle$, is isomorphic to $\mathbb{Z}$. If g has finite order n then the factor theorem says that $\mathbb{Z}_n$ is isomorphic to $\langle g \rangle$ via the homomorphism taking 1 (in $\mathbb{Z}_n$) to g.

2.9 More Examples of Groups

Before continuing the theoretical development we introduce a few more interesting families of groups. The first set are abelian, derived from the number systems discussed in Section 1.2. The second family is matrix groups, which are (generally) non-abelian.

Groups from Familiar Number Systems

We have already treated the additive group of the integers, $\mathbb{Z}$ as well as its subgroups $n\mathbb{Z}$. We have also used the integers modulo n, which we can now identify as the quotient group of $\mathbb{Z}$ by its subgroup $n\mathbb{Z}$. As pointed out in Section 1.2, the additive group of the fields $\mathbb{Q}$ (rational numbers), $\mathbb{R}$ (real numbers) and $\mathbb{C}$ (complex numbers) are abelian groups. They are complicated as groups because they are not finitely generated. The next exercise shows that the quotient $\mathbb{Q}/\mathbb{Z}$ is interesting; every element has finite order, but the group is not finitely generated.

Exercises 2.9.1. The group $\mathbb{Q}/\mathbb{Z}$

(a) Show that every element in $\mathbb{Q}/\mathbb{Z}$ has finite order.

(b) On a number line, sketch a region that contains one element for each equivalence class of $\mathbb{Q}/\mathbb{Z}$.

(c) Show that for any integer n there is an element of order n in $\mathbb{Q}/\mathbb{Z}$.

(d) How many elements of order n are there in $\mathbb{Q}/\mathbb{Z}$?

(e) Show that for any finite set $\{r_1, r_2, \ldots, r_t\}$ of rational numbers,

$$\langle r_1 + \mathbb{Z}, r_2 + \mathbb{Z}, \ldots, r_t + \mathbb{Z}\rangle \neq \mathbb{Q}/\mathbb{Z}$$

This shows that the group $\mathbb{Q}/\mathbb{Z}$ is not finitely generated.

We can also consider the multiplicative groups from familiar number systems.

We noted in Section 1.1 there is both an additive and a multiplicative structure $\mathbb{Z}/n$, to the integers modulo n. For an integer a that is comprime to n there are integers u, v such that $ua + nv = 1$ by the GCD theorem. Then u and a are multiplicative inverses of each other in $\mathbb{Z}/n$. We call them units. In Proposition 2.1.8 we showed that U_n, the set of units in $\mathbb{Z}/n$, forms a group under multiplication: 1 is the identity element, every element has an inverse by definition, the product of two units is also a unit (with $(ab)^{-1} = a^{-1}b^{-1}$), and multiplication is associative and commutative.

For a prime number p, every nonzero element in $\mathbb{Z}/p$ is a unit, so $\mathbb{Z}/p$ is a field. When considering it as a field we will write it $\mathbb{F}_p$.

Exercises 2.9.2. Automorphism groups of $\mathbb{Z}_n$

(a) Prove that $\mathrm{Aut}(\mathbb{Z}_n) \cong U_n$, the group of units in $\mathbb{Z}_n$.

(b) We will prove in Section 3.2 that every finite abelian group is isomorphic to a cyclic group or a direct product of such. For each of $n = 8, 9, 10, 11, 12$ find the product of cyclic groups that is isomorphic to U_n.

Returning to the fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, we now consider their multiplicative groups, which are written $\mathbb{Q}^*$, $\mathbb{R}^*$, and $\mathbb{C}^*$.

Example 2.9.3. In $\mathbb{Q}^*$, there is only one element of finite order $\mathbb{Q}^*$, other than the identity element, namely -1. It has order 2. Similarly the only non-identity element of $\mathbb{R}^*$ which has finite order is -1. To get elements of order n we are, in effect, looking for solutions of $x^n - 1$, that is nth roots of unity. These live in the complex number field $\mathbb{C}$.

Exercises 2.9.4. Subgroups of $\mathbb{Q}^*$

(a) For any rational number $r \neq \pm 1$ show that $\left\{r^i : i \in \mathbb{Z}\right\}$ is a subgroup of $\mathbb{Q}^*$ that is isomorphic to $\mathbb{Z}$.

(b) Show that the subgroup generated by 2 and 2 is $\langle 2, 3\rangle = \left\{2^i 3^j : i, j \in \mathbb{Z}\right\}$ and that it is isomorphic to $\mathbb{Z} \times \mathbb{Z}$. Generalize.

(c) Show that the positive rational numbers $\mathbb{Q}^{**} = \{a \in \mathbb{Q} : a > 0\}$ form a subgroup of $\mathbb{Q}^*$.

(d) Show that $\mathbb{Q}^*$ is isomorphic to the direct product of $\mathbb{Q}^{**}$ and $\langle -1\rangle$.

(e) Extend this result to the multiplicative group of the real numbers, $\mathbb{R}^*$.

Exercises 2.9.5. An additive group isomorphic to a multiplicative group.

(a) Show that there is a homomorphism from $\mathbb{Q}, +$ to $\mathbb{C}^*, *$, namely $a \longmapsto e^{a2\pi i}$.

(b) Show that the image is the set of all nth roots of unity (for $n \in \mathbb{N}$) and that this forms a subgroup of $\mathbb{C}^*$ under multiplication.

(c) What is the kernel?

Matrix Groups

We will work primarily with matrix groups over the fields, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{F}_p$, but the general results below are true for any field, so we express them for a general field F. We denote the multiplicative group of F by F^*.

Definition 2.9.6. Let F be a field and let n be an integer. The set of $n \times n$ matrices over F with nonzero determinant is called the **General Linear Group** and is written $\mathrm{GL}_n(F)$. The subgroup consisting of the matrices with determinant 1 is the **Special Linear Group** and is written $\mathrm{SL}_n(F)$. The next proposition shows that these are indeed groups with the identity matrix, I_n as identity element.

Proposition 2.9.7. *Let F be a field and let n be an integer. The set of $n \times n$ matrices over F with nonzero determinant forms a group. The determinant function*

$$\det : \mathrm{GL}_n(F) \longrightarrow F^*$$

is a homomorphism, and its kernel is $\mathrm{SL}_n(F)$.

Proof. Matrix multiplication is associative: One can show that for $n \times n$ matrices A, B, C, the i, j component of the product of $A(BC)$ and of $(AB)C$ is

$$\sum_{s=1}^{n}\sum_{t=1}^{n} a_{is}b_{st}c_{tj}$$

Thus $A(BC) = (AB)C$. (Associativity holds for any product of matrices that is well defined. We are treating the special case where they are all square of the same dimension.)

The result from linear algebra (which we assume here) that the determinant of a product of two matrices is the product of their determinants shows, in particular, that the product of two matrices with nonzero determinant also has nonzero determinant. So $\mathrm{GL}_n(F)$ is closed under multiplication. The identity matrix, I_n, and the usual formula for the inverse of a matrix perform the expected roles to make $\mathrm{GL}_n(F)$ a group. The determinant function respects products, so it gives a homomorphism to F^*. The kernel is the subgroup of matrices with determinant 1, that is $\mathrm{SL}_n(F)$. □

Exercises 2.9.8. INTERESTING SUBGROUPS OF THE GENERAL LINEAR GROUP

Show that the general linear group has these subgroups:

(a) The matrices of the form aI_n for $a \in F$ are called the constant diagonal matrices. Taking a nonzero we get the subgroup F^*I_n of $\mathrm{GL}_n(F)$. Show that F^*I_n is the center of $\mathrm{GL}_n(F)$ when $n > 1$. [Show first that F^*I_n is in the center of $(\mathrm{GL}_n(F))$. Then show no other matrices are in the center.]

(b) The diagonal matrices with nonzero entries on the diagonal.

(c) The upper triangular matrices with nonzero entries on the diagonal.

(d) The **orthogonal group** $\mathrm{O}(n, F)$ is the group of matrices Q such that Q^{-1} is the transpose of Q.

(e) For any subgroup H of F^* the set of all matrices with determinant in H.

There are two other matrix groups of particular interest. In the exercises above, you showed that for $n > 1$, the constant diagonal matrices, F^*I_n form the center of $\mathrm{GL}_n(F)$. In particular F^*I_n is normal in $\mathrm{GL}_n(F)$.

Definition 2.9.9. The quotient group $\mathrm{GL}_n(F)/F^*I_n$ is called the **Projective General Linear Group** and is written $\mathrm{PGL}_n(F)$. The quotient group $\mathrm{SL}_n(F)/\Big(F^*I_n\cap \mathrm{SL}_n(F)\Big)$ is called the **Projective Special Linear Group** and is written $\mathrm{PSL}_n(F)$.

Exercises 2.9.10. Upper and lower triangular matrices

(a) Show that the subgroup of upper triangular 2×2 matrices is conjugate to the group of lower triangular matrices. [Hint: $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.]

(b) Show that the set of matrices with nonzero determinant of the form $\begin{bmatrix} 0 & a \\ b & c \end{bmatrix}$ is a coset of the upper triangular matrices.

Exercises 2.9.11. Another version of the quaternions

In $\mathrm{GL}(2,\mathbb{C})$ consider the matrices

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$

(a) Show that the set of matrices $\{\pm I, \pm A, \pm B, \pm C\}$ forms a subgroup of $\mathrm{GL}_n(\mathbb{C})$. It is called the quaternion matrix group.

(b) Show that this group is isomorphic to the quaternions as defined in Exercise 2.6.2.

Exercises 2.9.12. Another version of the dihedral group D_4

(a) Show that D_4 is isomorphic to the matrix group with elements $\{\pm I, \pm A, \pm B, \pm C\}$ where

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(b) Draw the lattice diagram for this matrix group (it looks just like D_4, but use the elements here).

(c) More generally find a subgroup of $\mathrm{GL}_2(\mathbb{R})$ that is isomorphic to D_n. (Remember your trigonometry.)

Consider the matrix P^σ that has a single 1 in each column with the other entries being 0, specifically, $P^\sigma_{\sigma(i),i} = 1$. Notice that P^σ can be considered as a matrix over any field F. For $v \in F^n$, and $i \in \{1, \ldots, n\}$ the $\sigma(i)$ component of the vector $P^\sigma(v)$ is v_i. So, P^σ permutes the components of v. Another way to say this is that the ith component of $P^\sigma(v)$ is $v_{\sigma^{-1}(i)}$.

In particular, the null space of P^σ is trivial, so $P^\sigma \in \mathrm{GL}_n(F)$.

Proposition 2.9.13. *The function from S_n to $\mathrm{GL}_n(F)$ taking σ to P^σ is an injective homomorphism.*

Proof. We must check that $P^\pi P^\sigma = P^{\pi\sigma}$, which we do by verifying that for any $v \in F^n$, the ith components of $P^\pi\big(P^\sigma(v)\big) = P^{\pi\sigma}(v)$ are the same. The ith component of $P^{\pi\sigma}(v)$ is $v_{(\pi\sigma(i)}$. The ith component of $P^\pi\big(P^\sigma(v)\big)$ is the $\pi(i)$ component of $P^\sigma(v)$, which is the $\sigma\big(\pi(i)\big)$ component of v. Since $\pi\sigma = \sigma^{-1}\pi^{-1}$, the two matrices $P^\pi P^\sigma$ and $P^{\pi\sigma}$ are giving the same answer.

Injectivity is clear because the only permutation σ such that P^σ takes each basis vector to itself is the identity permutation.

□

2.10 Structure in the Quotient Group: The Third Isomorphism Theorem and the Correspondence Theorem

The next step is to understand the structure of a quotient group. The two main results—the third isomorphism theorem and the correspondence theorem—have fairly simple statements, which obscure some subtle issues. The proof of the third isomorphism theorem is a consequence of the first isomorphism theorem.

Theorem 2.10.1 (Third Isomorphism)**.** *Let N and K be normal subgroups of G with K contained in N. Then N/K is a normal subgroup of G/K and*

$$\frac{(G/K)}{(N/K)} \cong G/N.$$

Proof. We have two well defined quotient groups of G: G/K and G/N. I claim that there is a well-defined function from G/K to G/N taking gK to gN. To prove this, we have to check that if two cosets aK and bK are equal then the cosets aN and bN are also equal. Suppose $aK = bK$. Then $a^{-1}b \in K$ and since $K \subseteq N$ we have $a^{-1}b \in N$. Consequently $aN = bN$. Thus, there is a function taking aK to aN.

We now do the straightforward verification that the function $\varphi : G/K \longrightarrow G/N$ defined above is surjective and a homomorphism. Given any gN there is an element, namely gK, that clearly maps to it, $\varphi(gK) = gN$, so we get surjectivity. Also, φ respects multiplication: $\varphi(gK * g'K) = \varphi(gg'K) = gg'N = gN * g'N = \varphi(gK) * \varphi(g'K)$

The kernel of φ is $\{gK : gN = eN\}$. But $gN = eN$ if and only if $g \in N$. So the kernel is N/K. Applying the First Isomorphism Theorem 2.8.9 to $\varphi : G/K \longrightarrow G/N$,

$$\frac{(G/K)}{(N/K)} \cong G/N.$$

□

The more powerful theorem is the Correspondence Theorem, which we may be seen as a strengthening of the First Isomorphism Theorem. Recall what we said about isomorphic groups at the end of Section 2.3: If G and H are isomorphic the isomorphism sets up a correspondence between subgroups of G and subgroups of H, so the lattices of G and H have the same structure. Furthermore, normal subgroups of G are paired with normal subgroups of H under this correspondence by Exercise 2.8.6. The Correspondence Theorem gives similar information in the more general setting of a *surjective homomorphism* $\varphi : G \longrightarrow H$.

Before stating it, let us recall some simple facts about functions. Let $f : X \to Y$ and let $A \subseteq X$ and $B \subseteq Y$. Then $A \subseteq f^{-1}\big(f(A)\big)$ because for any $a \in A$, $a \in f^{-1}\big(f(a)\big)$. On the other hand, for an element $b \in B$, if x is a preimage of b then $f(x) = b$, but there may be no preimage for b, so we know only that $f\big(f^{-1}(B)\big) \subseteq B$.

If f is surjective then for each $b \in B$ there is some $x \in X$ such that $f(x) = b$. Thus for f surjective, $f\big(f^{-1}(B)\big) = B$. Respecting containment is also immediate: If $A \subseteq A' \subseteq X$ then $f(A) \subseteq f(A')$ and similarly if $B \subseteq B' \subseteq Y$ then $f^{-1}(B) \subseteq f^{-1}(B')$.

Theorem 2.10.2 (Correspondence)**.** *Let $\varphi : G \longrightarrow H$ be a surjective homomorphism with kernel K. There is a one-to-one correspondence, given by φ, between subgroups of H and subgroups of G that contain K.*

$$\begin{aligned} G &\longrightarrow H \\ A &\longleftrightarrow \varphi(A) \\ \varphi^{-1}(B) &\longleftrightarrow B \end{aligned}$$

The correspondence respects containment, normality, and quotients as follows. For A, A' containing K,

(1) $K \leq A \leq A'$ if and only if $\varphi(A) \leq \varphi(A')$.

(2) A is normal in G if and only if $\varphi(A)$ is normal in H.

(3) When A is normal in G, the map φ induces an isomorphism $G/A \cong H/\varphi(A)$.

Proof. Let A be a subgroup of G containing K and let B be a subgroup of H. From Proposition 2.3.5 we know that $\varphi(A)$ is a subgroup of H and $\varphi^{-1}(B)$ is a subgroup of G. Based on the above discussion, we know $\varphi(\varphi^{-1}(B)) = B$ and $A \subseteq \varphi^{-1}(\varphi(A)$ so we need to show that $\varphi^{-1}(\varphi(A)) \subseteq A$ to get the one-to-one correspondence. Let $g \in \varphi^{-1}(\varphi(A))$. Then $\varphi(g) = \varphi(a)$ for some $a \in A$. Consequently, $\varphi(ga^{-1}) = e_H$ and therefore $ga^{-1} \in \ker(\varphi) = K$. Since $K \subseteq A$, $ga^{-1} \in A$ so $g \in A$. Thus $\varphi^{-1}(\varphi(A)) = A$. We have established the one-to-one correspondence.

We have also shown in a problem in Exercise 2.8.6 that if B is normal in H then $\varphi^{-1}(B)$ is normal. These results are true for an arbitrary homomorphism. Let's now show that when φ is surjective, if A is normal in G then $\varphi(A)$ is normal in H.

Let $h \in H$. We need to show $h\varphi(A)h^{-1} = \varphi(A)$, or equivalently, $h\varphi(a)h^{-1} \in \varphi(A)$ for all $a \in A$. Since φ is surjective, there is some $g \in G$ such that $\varphi(g) = h$.

$$h\varphi(a)h^{-1} = \varphi(g)\varphi(a)\varphi(g)^{-1} = \varphi(gag^{-1}) \in \varphi(A)$$

The last step holds because A is normal in G, so $gag^{-1} \in A$.

Now we apply the first isomorphism theorem. Let B be normal in H. We have a composition of surjective homomorphisms

$$G \longrightarrow H \longrightarrow H/B$$

whose kernel is $\varphi^{-1}(B)$. Letting $A = \varphi^{-1}(B)$, the first isomorphism theorem says that $G/A \cong H/\varphi(A)$. □

If $\varphi : G \longrightarrow H$ is a surjective homomorphism the First Isomorphism Theorem says that H is isomorphic to $G/\ker(\varphi)$. So the statements in the Correspondence Theorem can be rewritten to say that there is a one-to-one correspondence between sbgroups of G/K and subgroups of G containing K. In particular, we can derive the Third Isomorphism Theorem as a corollary of the Correspondence Theorem.

Corollary 2.10.3 (Third Isomorphism Theorem)**.** *Let K and N be normal subgroups of G with $K \leq N$. Then $G/N \cong (G/K)\Big/(N/K)$.*

Proof. Apply the correspondence theorem to $G \longrightarrow G/K$. The subgroup N of G corresponds to the subgroup N/K of G/K. Thus $G/N \cong (G/K)\Big/(N/K)$. □

The Third Isomorphism Theorem gives a framework for understanding the lattices of subgroups

ADD EXAMPLES

Exercises 2.10.4. Lattices and the Correspondence Theorem

For each of the following groups G and for each of the normal subgroups $N \trianglelefteq G$, identify the sublattice of the G that has the same structure as the lattice of the quotient group G/N.

(a) D_4

(b) A_4

(c) $\mathbb{Z}_{45}$

(d) $\mathbb{Z}_4 \times \mathbb{Z}_4$

2.11 Problems

Exercises 2.11.1. Weakened group axioms

Define a *hemigroup* to be a set G with an operation $*$ that is associative, has an identity element, and such that each element has a *right* inverse.

(a) Show that the right inverse of a is also a left inverse of a, so that a hemigroup is actually a group.

Exercises 2.11.2. Unit groups of $\mathbb{Z}_n$

Let $(U_n, *)$ be the group of invertible elements of $\mathbb{Z}/n$. Find all n such that $(U_n, *)$ is isomorphic to the following.

(a) $(\mathbb{Z}_2, +)$;

(b) $(\mathbb{Z}_4, +)$;

(c) $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$.

Exercises 2.11.3. Some subgroups of abelian groups.

Let A be an abelian group and let m be an integer.

(a) Show that multiplication by m gives a homomorphism of A:

$$\begin{aligned} \varphi_m : A &\longrightarrow A \\ a &\longrightarrow ma \end{aligned}$$

(b) Show that the image and kernel are the groups mA and $A[m]$ from Exercise 2.2.6.

(c) If A is a finite group that has no elements of order m then multiplication by m gives an isomorphism of A.

(d) If m and n are coprime show that $A[m] \cap A[n] = \{0\}$.

Exercises 2.11.4. The torsion subgroup of an abelian group

Let A be an infinite abelian group. Let $\operatorname{Tor}(A)$ be the set of elements with finite order, which is called the **torsion subgroup** of A.

(a) Show that $\operatorname{Tor}(A)$ is, indeed, a subgroup of A and that it is normal.

(b) Show that $\operatorname{Tor}(A) = \bigcup_{m \in \mathbb{N}} A[m]$. (Note that, even inside an abelian group, the union of subgroups is not usually a group!)

(c) Show that $\operatorname{Tor}(A/\operatorname{Tor}(A))$ is trivial. That is, letting $T = \operatorname{Tor}(A)$, the only element of finite order in A/T is the identity element, $e + T$.

(d) Give an example of a finitely generated abelian group in which the identity element together with the elements of infinite order do *not* form a subgroup. (As opposed to the torsion subgroup.)

Exercises 2.11.5. "Almost" abelian groups

A group is **metabelian** when it has a normal subgroup N such that N

and G/N are both abelian. A group is **metacyclic** when it has a normal subgroup N such that N and G/N are both cyclic.

(a) Show that S_3 is metacyclic.

(b) Show that A_4 is metabelian but not metacyclic.

(c) Prove that any subgroup of a metabelian group is also metabelian.

(d) Prove that any quotient group of a metabelian group is metabelian. [Look carefully at the proof of the 2nd isomorphism theorem and adapt it to this question.]

Exercises 2.11.6. THE HEISENBERG GROUP

Let $H = H(F)$ be the set of 3 by 3 upper triangular matrices over a field F with 1s on the diagonal.

(a) Give a brief explanation of why this is indeed a subgroup of $\mathrm{GL}(3, F)$.

(b) (HW) Show that the center $Z(H)$ consists of all matrices of the form $\begin{bmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Furthermore $Z(H) \cong (F, +)$.

(c) (HW) Show that $H/Z(H)$ is isomorphic to $F \times F$.

(d) (HW) Conclude that H is metabelian.

(e) Show that the following 3 types of matrices generate the Heisenberg group.

$$\begin{bmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}$$

(f) Suppose that $F = \mathbb{F}_p$. Explain why H is then generated by 3 matrices, those in the form above with $a = b = c = 1$.

(g) Show that $H(\mathbb{F}_2) \cong D_4$.

Exercises 2.11.7. UPPER TRIANGULAR MATRICES

(a) Let F be a field and let F^* be its multiplicative group. Show that there is a homomorphism

$$\begin{aligned} \{\text{upper triangular matrices in } \mathrm{GL}(2, F)\} &\longrightarrow (F^*)^2 \\ \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} &\longmapsto (a, c) \end{aligned}$$

(b) Show that kernel is isomorphic to $(F, +)$ the additive group of F.

Exercises 2.11.8. SOME NORMAL SUBGROUPS

(a) Show that the intersection of two normal subgroups of G is normal in G.

(b) Let G be a group, possibly infinite. Let I be some indexing set and for each $i \in I$ let H_i be a subgroup of G. Prove that for any $a \in G$,

$$a\Big(\bigcap_{i\in I} H_i\Big)a^{-1} = \bigcap_{i\in I} aH_ia^{-1}$$

(c) Let H be a subgroup of G and let $N = \bigcap_{g\in G} g^{-1}Hg$. Prove that N is normal in G.

(d) Let $n \in \mathbb{N}$ and let K be the intersection of all subgroups of G of order n. Prove that K is normal in G.

Exercises 2.11.9. Normal subgroups and index

Exercise 2.8.6 showed that every group of index 2 is normal. Here is a generalization due to Lam [MAA Monthly Mar. 2004 p. 256].

Theorem 2.11.10. *Let H be a subgroup of G with $[G : H] = p$ a prime number. The following are equivalent.*

(1) H is normal in G.

(2) For any $a \in G \setminus H$, $a^p \in H$.

(3) For any $a \in G \setminus H$, $a^n \in H$ for some positive integer n that has no prime divisor less than p.

(4) For any $a \in G - H$, $a^2, a^3, \ldots, a^{p-1} \notin H$.

(a) Prove Lam's theorem by showing $(1) \implies (2) \implies (3) \implies (4) \implies (1)$. The last step is the one requiring some cleverness.

(b) Show that Lam's theorem implies that any group whose index is the smallest prime dividing $|G|$ is normal in G.

Exercises 2.11.11. Inner automorphisms of a group

For $a \in G$ let φ_a be the **inner automorphism** defined by a and consider the function $\varphi : a \longmapsto \varphi_a$.

$$\begin{aligned} \varphi_a : G &\longrightarrow G \\ g &\longmapsto aga^{-1} \end{aligned} \qquad\qquad \begin{aligned} \varphi : G &\longrightarrow \mathrm{Aut}(G) \\ a &\longmapsto \varphi_a \end{aligned}$$

Clearly $\mathrm{im}(\varphi) = \mathrm{Inn}(G)$.

(a) Show that $\mathrm{Inn}(G)$ is a normal subgroup of $\mathrm{Aut}(G)$.

(b) Show that φ is a homomorphism and that $\mathrm{im}(\varphi) \cong G/Z(G)$.

Exercises 2.11.12. Computing some simple automorphism groups.

(a) Compute $\mathrm{Aut}(Q)$ for Q the quaternion matrix group.

(b) Show that $\mathrm{Aut}(D_4) \cong D_4$

Exercises 2.11.13. Classification of the groups of order 8

Let G be a group of order 8. Prove each of the following.

(a) If G has an element of order 8 then $G \cong \mathbb{Z}_8$.

(b) If every nonzero element of G has order 2 then G is abelian and isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$.

(c) Suppose G has no element of order 8 and some element $a \in G$ has order 4.

- If G is abelian then it is isomorphic to $\mathbb{Z}_4 \times \mathbb{Z}_2$.
- Suppose G is not abelian. Let $b \notin \langle a \rangle$. Show that if b has order 2 then $G \cong D_4$. If b has order 4 then $G \cong Q$. [See Exercise 2.6.2

Exercises 2.11.14. Counting in S_n

(a) How many k-cycles are there in S_n?

(b) How many product of disjoint transpositions are there in S_n?

Exercises 2.11.15. The normalizer and centralizer of a subgroup

Let K be a subgroup of G and define

$$N_G(K) = \{g \in G : gKg^{-1} = K\}$$
$$C_G(K) = \{g \in G : gkg^{-1} = k \text{ for all } k \in K\}$$

These are called the **normalizer** of K in G and the **centralizer** of K in G.

(a) Show that $N_G(K)$ is a subgroup of G.

(b) Show that K is a normal subgroup of $N_G(K)$.

(c) If $H \leq G$ and K is a normal subgroup of H show that $H \leq N_G(H)$. So, $N_G(K)$ is the largest subgroup of G in which K is normal.

(d) Show that $C_G(K)$ is a normal subgroup of $N_G(K)$.

(e) Show that $N_G(K)/C_G(K)$ is isomorphic to a subgroup of $\text{Aut}(K)$.

Exercises 2.11.16. The commutator subgroup

In a group G, the **commutator** of a, b is $aba^{-1}b^{-1}$. Notice that this is e_G iff a and b commute. The **commutator subgroup** of a group G is the group G' generated by the commutators.

$$G' = \langle aba^{-1}b^{-1} : a, b \in G \rangle$$

(a) Compute the commutator subgroup of D_n (two cases: n odd and n even). Think of D_n as generated by r, t with $r^n = t^2 = e$ and $tr = r^{n-1}t$.

(b) Write down the commutator of the conjugation of a by x and the conjugation of b by x.

(c) Prove that G' is a normal subgroup of G. It is enough to show that the conjugation of any commutator is another commutator.

(d) Prove that G/G' is abelian.

(e) Prove that G/N abelian implies $G' \leq N$. So, the commutator subgroup of G is the smallest normal subgroup N group such that the quotient G/N is abelian.