Lecture Notes for Math 620
Groups, Rings, and Fields

Michael E. O'Sullivan
mosullivan@sdsu.edu

January 19, 2024

Contents

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 $b > a$ if and only if $b - a > 0$;
- 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 $K = \mathbb{N}$.

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 know 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 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 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 $0 \leq r < d$. This shows $r = 0$, 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 divisible by 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. 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 more efficient algorithms, but understanding this approach is sufficient here.)

Exercises 1.1.7. Use the Euclidean algorithm to express the greatest common divisor as a linear combination of the given integers.

(a) 89, 24

(b) 24, 40, 30

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 1.1.8. 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 1.1.9 (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.

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 $p(1 - xb) = 0$. This shows that $1 - xb = 0$, so $x = b = \pm 1$. □

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

Theorem 1.1.10 (Unique Factorization)**.** *Let a be a positive integer. 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. TBD □

Exercises 1.1.11. The following are consequences 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 $u_i \in \mathbb{Z}$.

(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?

Exercises 1.1.12. The following are quick consequences of unique factorization

(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. Generalize this to a statement about rational numbers.

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

(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. 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 $Ba_i/(Ab_i)$.]

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 $(a-b)+(b-c) = a-c$, 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 $\mathbb{Z}/n$)

$$b(n - 1) = (-1)b = n - b.$$

Exercises 1.1.17.

(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 $[a][r] = [1]$.

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

$$\begin{array}{ll} -\frac{n-1}{2}, -\frac{n-3}{2}, \ldots, \frac{n-3}{2}, \frac{n-1}{2} & \text{for } n \text{ odd, and,} \\ -\frac{n-2}{2}, -\frac{n-4}{2}, \ldots, \frac{n-4}{2}, \frac{n-2}{2}, \frac{n}{2} & \text{for } n \text{ even.} \end{array}$$

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 $1{:}00, 2{:}00, \dots, 12{:}00$ rather than using $0{:}00, 1{:}00, \dots, 11{:}00$). 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 focusing 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. A complex number is of the form $a+bi$ in which $a, b \in \mathbb{R}$ and i is the square root of -1. 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 $\mathbb{Q}(i)$.

There is one other field that is accessible to those who have learned "clock arithmetic." If our clock had a prime number p of positions, we would get $\mathbb{Z}/p$, in which every nonzero element has a multiplicative inverse, as was shown in Exercise 1.1.17. 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 by looking at the multiplicative group of a field. 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 $\mathbb{C}$).

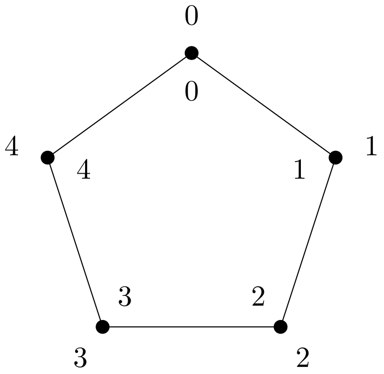


Figure 1.1: The Pentagon

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 2.2.)

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? First, there are some ambiguities to clear up, so we adopt the following conventions:

- The rotation r is rotation of the vertex i 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 rt_i means reflect then rotate as is customary using functional notation; we apply the function on the right first.

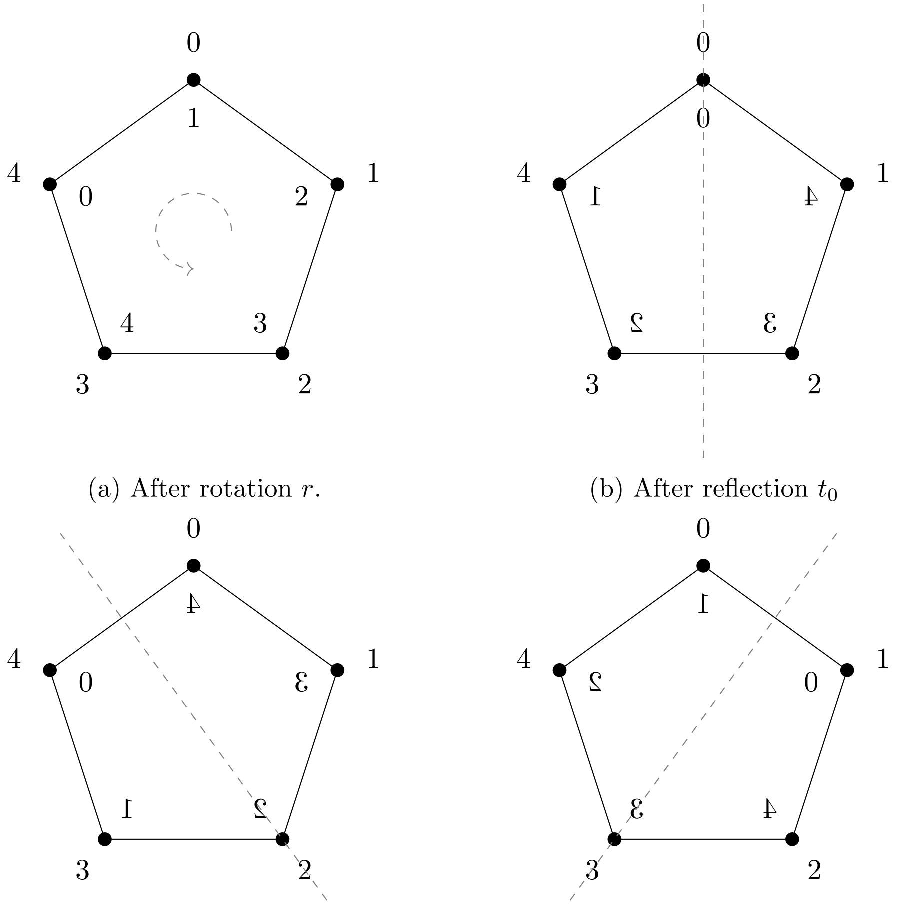
Figure 1.2 shows the two computations, $t_0 r = t_2$ and $rt_0 = t_3$. These are unequal, so these computations show that we have our first example of a nonabelian group. It is called the dihedral group of order 5, and written D_5.

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.

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. Continuing with the D_5 example above, the goal of this problem is to find formulas for the product of two arbitrary elements of D_5. We will use arithmetic in $\mathbb{Z}/5$ with the system of representatives $0, 1, 2, 3, 4$.



(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

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

(c) 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 $t_i r^j$.

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

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).

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

The symmetric groups will be discussed in detail in Section 2.4.

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 2.2) 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.

NEED SOME IMAGES: CUBE, CRYSTAL, FRIEZE DIAGRAMS, TILINGS,

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 previous section. By $F[x]$ we mean the polynomial ring in the indeterminate x. The key fact about polynomial rings is that all the theorems we discussed for integers in Section 1.1 also hold—with appropriate modifications—for polynomial rings 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 as

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

- 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.

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

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

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

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 $r_k(x) \neq 0$)

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

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. Use the Euclidean algorithm to 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.

□

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. TBD □

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.

(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 ahve 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 operation, 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. 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.

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.

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

Example 2.1.3. 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 2.1.4. 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.

Example 2.1.5. The Cartesian product $\mathbb{Z} \times \mathbb{Z}$ is a group under coordinatewise addition with identity element $(0, 0)$ and inverse operator $(a, b) \longmapsto (-a, -b)$. It is not possible to generate all elements by repeated addition of a single element, but 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 2.1.6.

(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) Show that $\mathbb{Z} \times \mathbb{Z}$ is not generated by a single element.

(c) 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 elaborates on Exercise 1.1.17 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 2.1.7. *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 $[a][u] = [1]$. 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 **??**.

Definition 2.1.8. 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 2.1.9. 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).

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.

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

The Order of an Element

Definition 2.1.10. 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 $\operatorname{ord}(g)$ 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 $\exp(G) = \operatorname{lcm}\{\operatorname{ord}(g) : g \in G\}$. If no such element exists one can say the exponent is infinite.

Only the identity element of a group has order 1. Every nonzero element of $\mathbb{Z}$ has infinite order. In $\mathbb{Z}_n$ some elements have order n, but other non-identity elements may have a different order. For any finite group there is a well defined exponent, but an infinite group may or may not have one.

Theorem 2.1.11 (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) $\operatorname{ord}(g^t) = \frac{n}{\gcd(t,n)}$.

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. For an 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 shows claim (2a). 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 2.1.12. The reflections in D_n all have order 2. The rotation by $2\pi/n$ has order n, but some of the other rotations has 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 2.1.13. 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 2.1.14. 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 $\operatorname{ord}(a) = \exp(A)$.

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 2.1.15. A nonempty subset H of a group G is a **subgroup**, when H is a group using the operation $*_G$ on G. This means that for any $h \in H$ the inverse h^{-1} must also be in H and for $h, h' \in H$, the product $h *_G h'$ must also be an element of H. We will say H is **closed under inversion and multiplication**. 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 2.1.16. *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.)

Proposition 2.1.17 (Subgroup Properties)**.** *If K is a nonempty subset of G that is closed under inversion and closed under multiplication in G then K is a subgroup of G (i.e. it also contains e_G).*

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

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

Suppose K is a nonempty subset of G such that $h *_G k^{-1} \in K$ for all $h, k \in K$. For any $k \in K$, setting $h = k$ gives $k *_G k^{-1} = e_G \in K$. Letting $h = e_G$ gives $e_G * k^{-1} = k^{-1} \in K$, so K is closed under inversion. Now for any $h, k \in K$ we know $k^{-1} \in K$, so $h *_G (k^{-1})^{-1} = h *_G k \in K$. This shows K 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 2.1.18.

- **(a)** Let G be a group and 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 2.1.19. 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 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 2.1.20. 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.)

Example 2.1.21. 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 be also 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 2.4).

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 the 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 case of subgroups.

Definition 2.1.22. 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. So any minimal subgroup must be cyclic. This suggests a general approach to work up from the trivial subgroup $\{e_G\}$ to construct the lattice of subgroups.

(1) Find all subgroups generated by 1 element.

(2) Find all subgroups generated by 2 elements by adding an 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.

$$\begin{tikzcd}
 & \mathbb{Z}_2 \times \mathbb{Z}_2 \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\langle (1,0) \rangle \arrow[dr, dash] & \langle (1,1) \rangle \arrow[d, dash] & \langle (0,1) \rangle \arrow[dl, dash] \\
 & \langle (0,0) \rangle &
\end{tikzcd}$$

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

Example 2.1.23. 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, $n\mathbb{Z}$, for 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

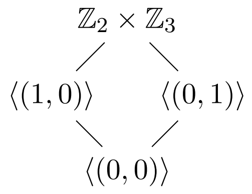


Figure 2.2: The lattice diagram for $\mathbb{Z}_2 \times \mathbb{Z}_3$.

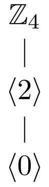


Figure 2.3: The lattice diagram for $\mathbb{Z}_4$.

subgroups of $\mathbb{Z}$ and that the subgroups for distinct $n \in \mathbb{Z}$ 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 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 $\mathbb{Z}$ 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 2.1.24. 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 $\mathbb{Z}_3 \times \mathbb{Z}_4$.

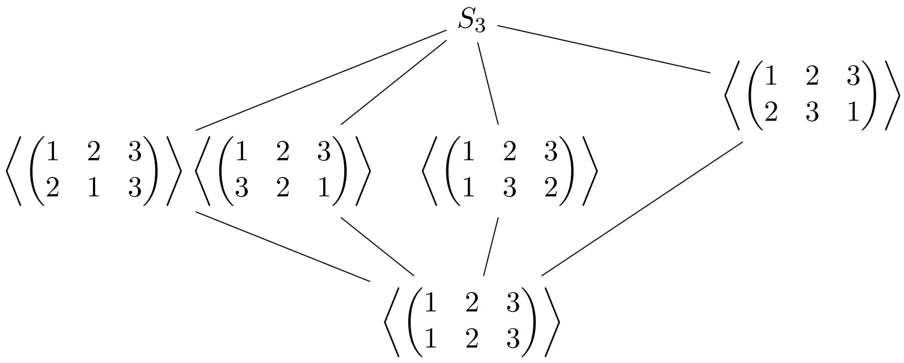

S_3
$\left\langle\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}\right\rangle$
$\left\langle\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}\right\rangle$
$\left\langle\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}\right\rangle$
$\left\langle\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}\right\rangle$
$\left\langle\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix}\right\rangle$

Figure 2.4: The lattice diagram for S_3.

(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.

2.2 Group Homomorphisms

In any field of algebra a key aspect of the field is to identify the functions that are appropriate to study. These functions are typically called homomorphisms (from classical languages: hom meaning “same” and morph “shape”) because they relate to the algebraic structures on which they act.

Definition 2.2.1. 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 2.2.2 (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 2.2.3. 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).

Proposition 2.2.4. *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. □

Proposition 2.2.5. *Let $G \longrightarrow H$ 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 $\varphi(g^{-1})$ and 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 2.2.6. 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 2.2.7. *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 2.7 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 2.2.8. *A homomorphism $\varphi : G \longrightarrow$ 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 and that $\varphi(a) = \varphi(b)$ for some $a, b \in G$. Then $\varphi(ab^{-1}) = \varphi(a)\varphi(b^{-1}) = \varphi(a)(\varphi(b))^{-1} = e_H$. This forces $ab^{-1} = e_G$, and multiplying on the right by b we get $a = b$. Since φ is injective. □

Isomorphisms

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 2.2.9 (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 enough 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 2.2.5 we know that the composition of homomorphisms is a homomorphism. Thus the composition of two isomorphisms is an isomorphism.

Clearly any group is isomorphic to itself under the identity map. The first part of this proposition shows that if G is isomorphic to H then H is also isomorphic to G. The second part establishes transitivity. □

If there is an isomorphism between groups G and H then G and H have the same algebraic structure, so we consider them equivalent.

Exercises 2.2.10.

(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?

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

Exercises 2.2.11.

(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 2.2.12.

(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 2.2.13. A Perverse Group

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

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

Exercises 2.2.14. 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 $\operatorname{ord}(\varphi(g))$ divides $\operatorname{ord}(g)$.

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

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 $\operatorname{ord}(g)$.

Exercises 2.2.15.

(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.

2.3 Some Constructions of Groups

In this section we show three ways to construct new groups from ones that we already have. All three have been touched on briefly; we give more detail here. The first construction really has two aspects, given some subgroups of a group we can form the intersection of the subgroups and also the join of the subgroups. More broadly, given a subset of a group we can define the subgroup generated by the subset. The second 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. Finally, for a given group G the isomorphisms of G to itself have a group structure that is useful in understanding the properties of G.

Intersections and the Join of Subgroups

Proposition 2.3.1. *Let $H_1, \dots, H_t$ be subgroups of G. The intersection $\bigcap_{i=1}^t H_i$ is a subgroup of G.*

More generally, if $\mathcal{H}$ is a set of subgroups of G then $\bigcap_{H \in \mathcal{H}} H$ is a subgroup of G.

Proof. Suppose h and h' are in $\bigcap_{H\in\mathcal{H}} H$. Since each $H \in \mathcal{H}$ is a subgroup of G, $h'h^{-1} \in H$ for each $H \in \mathcal{H}$. Thus $h'h^{-1} \in \bigcap_{H\in\mathcal{H}} H$, and Proposition 2.1.17 shows $\bigcap_{H\in\mathcal{H}} H$ is a subgroup of G. □

Let S be an arbitrary subset of a group G. Let $\mathcal{H}$ be the set of all subgroups of G containing S. Then $\bigcap_{H\in\mathcal{H}} H$ is a subgroup of G, and it contains S, since each $H \in \mathcal{H}$ contains S. Furthermore, any subgroup K of G containing S is in $\mathcal{H}$ so $\bigcap_{H\in\mathcal{H}} H \subseteq K$. This argument justifies the following definition.

Definition 2.3.2. Let G be a group and let S be a subset of G. By $\langle S \rangle$ we mean the **smallest subgroup** of G containing S. It is the intersection of all subgroups of G containing S. We say $\langle S \rangle$ is the subgroup of G **generated** by S.

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

One interesting problem is to find minimal size sets that generate a group. For example, the elements 1 and -1 both generate $\mathbb{Z}$. The element 1 generates $\mathbb{Z}_n$ as does any $a \in \mathbb{Z}_n$ that is coprime to n. (Exercise 2.1.6.)

The Direct Product

Definition 2.3.3. 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 2.3.4 (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 2.3.5.

(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 2.3.6 (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 2.3.6 we get

Corollary 2.3.7. *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 2.2.11 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 (which says a bit more than our corollary does). The following exercise broadens the perspective

Exercises 2.3.8. 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 2.3.6 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 an isomorphism?

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

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

Automorphism Groups

Proposition 2.3.9 (Automorphisms). *Let G be a group. The set of all isomorphisms from G to itself is a group. This new group is called* $\mathrm{Aut}(G)$*, 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 2.2.9 shows that the inverse of an isomorphism is an isomorphism and the composition of two isomorphisms is an isomorphism. Thus $\mathrm{Aut}(G)$ is a subgroup of the group of bijections of G. □

Exercises 2.3.10.

(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.

2.4 Permutation Groups

In this section we delve more deeply into the structure of the **symmetric group** S_n, the group of permutations of $\{1, \dots, n\}$. 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 $\{1, \dots, n\}$ 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 2.4.1.

(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 2.4.2. 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 2.4.3.

(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 2.4.4. 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 2.4.5.

(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 2.4.6. 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 2.4.7. *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 2.4.8. *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, $\operatorname{orb}(a) = \left\{a, \pi(a), \ldots, \pi^{r-1}(a)\right\}$. 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. □

Definition 2.4.9. We will call the list of cycle lengths, in decreasing order, the **signature** of the permutation.

We will include one-cycles in the definition of the cycle decomposition, although we will not write them unless it is needed for clarity. For example, the permutation

π in S_5 from Exercise 2.4.1 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 $3, 2, 1$.

Exercises 2.4.10. 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 2.4.11. A sanity check: 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) What is the exponent of S_n?

(c) 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.

Transpositions and the Alternating Group

There is another factorization that is important.

Proposition 2.4.12. *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

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

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 2.4.13. Generators for S_n.

(a) Show that S_n is generated by the $n-1$ elements $(1,k)$ for $k = 2, \ldots, n$. [Show that you can get an arbitrary transposition by conjugating $(1,k)$ by some $(1,j)$, see Definition 2.6.9.]

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

Exercises 2.4.14.

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

Consequently, any permutation can be written as a product of an even number of transpositions, or an odd number of transpositions, but not both.

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 $\mathrm{id} = \tau_1 \cdots \tau_n$ with $\tau_i = (b_{2i-1}, b_{2i})$. The b_i are not necessarily distinct, except $b_{2i-1} \neq b_{2i}$ 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 $\tau_1 \cdots \tau_n(a) = \tau_1(b_1) = b_2 \neq a$ 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 n 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 cancelation, or we arrive at a factorization $\tau_1\sigma_2 \cdots \tau_{k+1}\tau_{k+2} \cdots \tau_n$ 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 $n-2$ 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 2.4.16. *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.

Exercises 2.4.17.

(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 2.4.18.

(a) Show that there is a homomorphism from S_n to $\mathbb{Z}_2$ whose kernel is A_n.

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

(c) What is the intersection of A_4 and D_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.

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 2.4.19. *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 $g = g'$.

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, it should be clear that S_T is isomorphic to S_n. Nevertheless, it is worth explicitly giving a construction of an embedding of G into S_n, for $|G| = n$. Suppose G has order n and enumerate 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 2.4.20. Cayley's Theorem

- **(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 $a_ia_1 = a_{\sigma_i(1)}$, $a_ia_2 = a_{\sigma_i(2)}$, $a_ia_3 = a_{\sigma_i(3)}$, $a_ia_4 = a_{\sigma_i(4)}$.
- **(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

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

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 2.4.21.

(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 2.4.22.

(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 $n/2$.

2.5 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 2.5.1.

(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 $c \rightarrow ab$.

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

$$(a^i)(a^j b^k) = a^{i+j} b^k$$
$$(a^i b)(a^j b^k) = a^{i-j} b^{k+1}$$

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 2.5.2. Here is another group. It is called the **quaternion group**.

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

(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) Show that Q has 1 element of order 2 and 6 of order 4.

(c) Draw the lattice diagram for this group.

Exercises 2.5.3. 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 these diagrams have that are not captured by D_∞?

2.6 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 2.6.1. 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, taking note that for three sets, $(ST)U = S(TU)$, by associativity, so we may just call this set STU.

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

Notice that ST and TS are not necessarily equal when the group G is not abelian.

Exercises 2.6.2. R

evisiting the properties of a subgroup.

(a) Let T be a 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 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 show that $TT = T$ is not sufficient to ensure T is a group of G.

Definition 2.6.3. 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 2.6.4. *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 2.6.5. *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 $gh = akh$ and 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 2.6.6 (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 $a_1, \ldots, a_t$ such that G is the disjoint union of $a_1H, a_2H, \ldots, a_tH$. The cosets of H all have the same number of elements by Lemma 2.6.4. 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 $\operatorname{ord}(a)$. So $\operatorname{ord}(a)$ divides $|G|$. □

Definition 2.6.7. 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 2.6.8. Recall that the exponent of a group G is the lcm of the orders of the elements (if this is finite).

(a) For a finite group G with finite exponent t show that t divides the order of G.

(b) Give an example to show that there may not be an element in G whose order is the exponent of G.

Conjugation

Now we consider conjugation.

Definition 2.6.9. 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 2.6.10. *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 $a = g^{-1}b\big(g^{-1}\big)^{-1}$ 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

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

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

Exercises 2.6.11. Let $C(a) = \{g \in G : ga = ag\}$ be the **centralizer** of a.

(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 2.6.12. Conjugates and subgroups.

(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 2.6.13. 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}.

Exercises 2.6.14. Inner automorphisms.

(a) Define a function $\varphi_a : G \longrightarrow G$ by $\varphi(g) = aga^{-1}$. Show that φ_a is an automorphism of G.

(b) Show that $\varphi : G \longrightarrow \mathrm{Aut}(G)$ defined by $\varphi : a \longmapsto \varphi_a$ is a homomorphism. The image, $\{\varphi_a : a \in G\}$, is therefore a subgroup of $\mathrm{Aut}(G)$. It is called $\mathrm{Inn}(G)$, the group of **inner automorphisms** of G.

(c) What is the kernel of φ?

Proposition 2.6.15. *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 2.6.16. Consider conjugation by $\pi \in S_n$.

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

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

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

(b) If A and B are disjoint subsets of $\{1, \ldots, n\}$ 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.

2.7 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 2.2.5 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.

This section and Section 2.9 are focused on surjective homomorphisms, which are intimately related to subgroups that have a special property, treated in the next proposition.

Theorem 2.7.1 (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) $\Longrightarrow$ (2): $(aN)(bN) = a(Nb)N = a(bN)N = (ab)N$. Here we have used asociativity, then the assumption in (1), and finally, $NN = N$ since N is a subgroup of G.
(2) $\Longrightarrow$ (3): Set $b = a^{-1}$. Then, using (2), $aNa^{-1}N = aa^{-1}N = eN$. In particular, this shows that $aNa^{-1} \subseteq N$.
(3) $\Longrightarrow$ (4): 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): 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 2.7.2. A group satisfying the conditions of the theorem is called **normal**. We write $N \trianglelefteq G$ for N a normal subgroup of G.

Exercises 2.7.3.

(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 2.6.5, $arN = bsN$. 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). The next theorem shows that this gives a group structure on the cosets of N in G.

Theorem 2.7.4. *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 2.7.5. 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 2.7.6. 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 2.7.7. The **center** of G is 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.

Exercises 2.7.8. 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 2.7.9 (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. First we prove that K is normal by showing $aKa^{-1} \subseteq K$ for all $a \in G$. For any $k \in K$,

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

Thus $aka^{-1} \in K$.

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 $aK = e_GK$. Thus the kernel of $\tilde{\varphi}$ just contains just the identity element of G/K. □

Here is a typical snappy use of the First Isomorphism theorem.

Theorem 2.7.10. *Let $G_1, G_2, \dots G_r$ be groups and let $N_1, N_2, \dots, N_r$ be normal subgroups, $N_i \trianglelefteq G_i$. There is a well defined map*

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

and it is an isomorphism.

Proof. By the Direct Product Theorem 2.3.4, 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 2.3.6 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. $(g_1, g_2, \dots, g_r)$ and $(g_1', g_2', \dots, g_r')$ be elements of $G_1 \times G_2 \times \cdots \times G_r$. Then

$$\begin{aligned}\varphi\Big((g_1, g_2, \dots, g_r) * (g_1', g_2', \dots, g_r')\Big) &= \varphi\Big(\,(g_1g_1', g_2g_2', \dots, g_rg_r')\,\Big)\\ &= (g_1g_1'N_1, g_2g_2'N, \dots, g_rg_r'N_r)\\ &= (g_1N_1, g_2N_2, \dots, g_rN_r) * (g_1'N_1, g_2'N_2, \dots, g_r'N_r)\\ &= \varphi(g_1, g_2, \dots, g_r) * \varphi(g_1', g_2', \dots, g_r')\end{aligned}$$

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 $(g_1, \dots, g_r)$ such that $g_1N_1, g_2N_2, \dots, g_rN_r = N_1 \times N_2 \times \cdots \times N_r$. 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 2.7.11 (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 $\varphi(\widetilde{gN}) = e_H$. 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 $n\mathbb{Z}$ (well isomorphic to it). 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 2.2.4 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.

2.8 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 2.8.1.

(a) Consider the group $\mathbb{Q}/\mathbb{Z}$. Show that every element 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.

Example 2.8.2. As 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 is the multiplicative inverse of a modulo n. The converse is also true, if $a + n\mathbb{Z}$ has multiplicative inverse $u + n\mathbb{Z}$ then ua differs from a multiple of n by 1, so the GCD of a and n is 1. The set of **units** in $\mathbb{Z}_n$, is the set of those elements with a multiplicative inverse. One can check that this set forms an abelian group under multiplication, written U_n: 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 2.8.3.

(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 $\mathbb{Q}^*$, $\mathbb{R}^*$, and $\mathbb{C}^*$.

Example 2.8.4. In $\mathbb{Q}^*$, the subgroup generated by 2 is $\{2^i : i \in \mathbb{Z}\}$. The subgroup generated by 2 and 2 is $\langle 2, 3\rangle = \{2^i 3^j : i, j \in \mathbb{Z}\}$. There is only one element of finite order in $\mathbb{Q}^*$, other than the identity element, namely -1, which 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 2.8.5. An isomorpism between an additive group and 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?

Exercises 2.8.6.

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

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

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

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 2.8.7. 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 2.8.8. *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 2.8.9. There are many interesting subgroups of the general linear group.

(a) Show that the general linear group has these subgroups:

- The diagonal matrices with nonzero entries on the diagonal.
- 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 $n > 1$.
- The upper triangular matrices with nonzero entries on the diagonal.
- The **orthogonal group** $\mathrm{O}(n, F)$ is the group of matrices Q such that Q^{-1} is the transpose of Q.

(b) For any subgroup H of F^* the set of all matrices with determinant in H is a subgroup of $\mathrm{GL}(n, F)$.

There are two other matrix groups of particular interest. In the exercises above, it is claimed 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 2.8.10. 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 2.8.11.

(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.

Example 2.8.12. In $\mathrm{GL}(2, \mathbb{C})$ consider the matrices

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

The set of matrices $Q = \{\pm I, \pm A, \pm B, \pm C\}$ is called the quaternion matrix group.

Exercises 2.8.13.

(a) Show by brute force that the quaternion matrix group is indeed a group.

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

(c) 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.]

Exercises 2.8.14.

(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.)

Another interesting class of matrices is permutation matrices. Let $\sigma \in S_n$. Consider the matrix P^σ that has a single 1 in each column with the other entries being 0, specifically, $P^\sigma_{i,\sigma(i)} = 1$. 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 2.8.15. *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 $v_{(\pi\sigma)^{-1}(i)}$. The ith component of $P^\pi\big(P^\sigma(v)\big)$ is the $\pi^{-1}(i)$ component of $P^\sigma(v)$, which is the $\sigma^{-1}\big(\pi^{-1}(i)\big)$ 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.

□

2.9 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 2.9.1 (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 $aN = bN$, so there is a function taking aK to aN.

It is easy to check 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. Finally, 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 2.7.9 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. 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 $A \subseteq f^{-1}(f(A))$ because for any $a \in A$, $a \in f^{-1}(f(a)$. 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 $f(f^{-1}(B)) \subseteq B$. But, 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, $f(f^{-1}(B)) = B$. 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 2.9.2 (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 G/K and subgroups of G containing 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,

- *$K \leq A \leq A'$ if and only if $\varphi(A) \leq \varphi(A')$.*
- *A is normal in G if and only if $\varphi(A)$ is normal in H.*
- *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 2.2.5 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 $\varphi^{-1}(\varphi(A)) = A$. Thus, we have established the one-to-one correspondence.

We have also shown in a problem in Exercise 2.7.6 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 $G/A \cong H/\varphi(A)$. □

We can derive the Third Isomorphism Theorem as a corollary of the Correspondence Theorem.

Corollary 2.9.3 (Third Isomorphism Theorem)**.** *Let KN 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

$\mathbb{Z}_n$

$\mathbb{Z}_4 \times \mathbb{Z}_4$

Exercises 2.9.4. 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

2.10 Problems

Exercises 2.10.1. 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 2.10.2. Let $(U_n, *)$ be the group of invertible elements of $\mathbb{Z}_n$. Find all n such that $(U_n, *)$ is isomorphic to

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

(b) $(\mathbb{Z}_4, +)$;

(c) $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$.

Exercises 2.10.3. 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 2.1.20.

(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 2.10.4. 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 2.10.5. “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 2.10.6. A matrix group. Let G be the group of all matrices of the form with $a, b, c \in \mathbb{Q}$.

$$\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$$

- **(a)** Find the center C of G and show that C is isomorphic to the additive group $\mathbb{Q}$.
- **(b)** Show that G/C is isomorphic to $\mathbb{Q} \times \mathbb{Q}$.
- **(c)** Conclude that G is metabelian.

Exercises 2.10.7. 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)** Show that the following 3 types of matrices generate this group.

$$\begin{bmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}$$

- **(c)** Let $F = \mathbb{F}_p$. Explain why H is then generated by 3 matrices, those in the form above with $a = b = c = 1$.
- **(d)** Show that $H(\mathbb{F}_2) \cong D_4$.
- **(e)** (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 $Z(H) \cong (F, +)$.
- **(f)** (HW) Show that $H/Z(H)$ is isomorphic to $F \times F$.
- **(g)** (HW) Conclude that H is metabelian.

Exercises 2.10.8. 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 2.10.9. General linear group.

(a) Which has more elements, unit group of $\mathrm{Mat}(2,\mathbb{Z}/4)$ or $\mathrm{GL}(2,\mathbb{F}_4)$? Find the number of elements in each and characterize $U(\mathrm{Mat}(2,\mathbb{Z}_4)$.

(b) Try to generalize: How many elements are there in $\mathrm{GL}(n,\mathbb{F}_p)$? In $\mathrm{SL}(n,\mathbb{F}_p)$?

Exercises 2.10.10. 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.

An earlier exercise showed that every group of index 2 is normal. Here is a generalization due to Lam [MAA Monthly Mar. 2004 p. 256].

Theorem 2.10.11. *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 $a \in G - H$, $a^2, a^3, \ldots, a^{p-1} \notin H$.

Exercises 2.10.12.

(a) Prove Lam's theorem by showing (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (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 2.10.13. 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 2.10.14. 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 2.10.15. Automorphism group of $\mathbb{Z}/p^n$.

(a) Prove that $\mathrm{Aut}(\mathbb{F}_p^2) \cong \mathrm{GL}(2, \mathbb{F}_p)$ for p prime.

(b) Find an element of $\mathrm{Aut}(\mathbb{F}_4 \times \mathbb{F}_4)$ that is not in $\mathrm{GL}(2, \mathbb{F}_4)$.

Exercises 2.10.16. 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 $\mathbb{Z}_8$.

(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 $\mathbb{Z}_4 \times \mathbb{Z}_2$.
- NEED TO FINISH! If G

Exercises 2.10.17. Symmetries of a solid.

(a) Enumerate the faces of a tetrahedron as follows. Each rigid motion from the tetrahedron to itself defines a permutation of the vertices. There are 12 such permutations. Write them down in an organized fashion, and briefly describe each.

(b) Enumerate the faces of a cube as follows. Each rigid motion from the cube to itself defines a permutation of the vertices. There are 24 such permutations. Write them down in an organized fashion, and briefly describe each.

Exercises 2.10.18. 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 2.10.19.

(a) Let A_1, A_2, B_1, B_2 be normal subgroups of a group G with $B_1 \leq A_1$ and $B_2 \leq A_2$ and finally $A_1 \cap A_2 = \{1\}$. Then B_1B_2 is normal subgroup in A_1A_2 and there holds

$$(A_1A_2)/(B_1B_2) \cong (A_1/B_1) \oplus (A_2/B_2).$$

v

Exercises 2.10.20. 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 $\mathrm{Aut}(K)$.

Exercises 2.10.21. 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.

$$G' = \langle aba^{-}b^{-1} : a, b \in G\rangle$$

(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.

Exercises 2.10.22. Let G be a group. For $a, b \in G$ define the commutator $[a, b] := aba^{-1}b^{-1}$ of a and b. For arbitrary subgroups U, V of G define $[U, V] := \langle [u, v] \mid u \in U, v \in V \rangle$. Now show the following:

(a) If U, V are normal subgroups of G, then so is $[U, V]$.

(b) $[G, G]$ is the smallest normal subgroup of G for which the quotient group is abelian.

(c) Setting $G^{(0)} := G$ and $G^{(i)} := [G^{(i-1)}, G^{(i-1)}]$ for all $i \in \mathbb{N}$, we find that G is solvable if and only if there exists $n \in \mathbb{N}$ such that $G^{(n)} = \{1\}$.

(d) For $n \geq 5$ let U be a subgroup of S_n and N a normal subgroup of U for which U/N is abelian. Show that if U contains all 3-cycles of S_n, then also N will contain these.

Hint: If $a, b, c, d, e \in \{1, \ldots, n\}$ are distinct elements, then there holds the equation

$$(a, b, c) = (a, b, d)(c, e, a)(d, b, a)(a, e, c).$$

(e) Show that this implies that the symmetric group S_n is not solvable for $n \geq 5$.

Chapter 3

Classification and Structure of Groups

3.1 Interaction between Two Subgroups: The Second Isomorphism Theorem and Semi-Direct Products

We now consider two subgroups of a group G and prove several results about the interaction between them. At first we make no additional assumptions on the two subgroups groups, then we assume that one is normal in G, and finally that both are. The main result is the second isomorphism theorem. But,we also get two key corollaries that introduce the notion of an internal direct product (as opposed to the external direct product that we have been using), and the more general notion of a semi-direct product (both internal and external).

Lemma 3.1.1. *Let K, H be subgroups of G. The following are equivalent:*

(1) $G = KH$ and $K \cap H = \{e_G\}$

(2) Every element of G can be uniquely written as kh for $k \in K$ and $h \in H$.

Proof. $G = KH$ is equivalent to saying that every element of G can be written in the form kh. We'll next show $K \cap H = \{e_G\}$ if and only if any expression for $g \in G$ as a product kh, with $k \in K$ and $h \in H$, is unique.

Suppose $K \cap H = \{e_G\}$ and $k_1 h_1 = k_2 h_2$. Then $k_1^{-1} k_2 = h_1 h_2^{-1}$. Since this is in both K and in H, it must be the identity. Therefore, $h_1 = h_2$ and $k_1 = k_2$, which proves uniqueness.

Now suppose that $K \cap H \neq e_G$; say $g \in K \cap H$ is not equal to e_G. Setting $h = g$ and $k = e_G$ gives one way to express g in the form kh, while setting $h = e_G$ and $k = g$ gives a different way. Thus we have non-uniqueness. □

As another prelude to the second isomorphism theorem we have the following lemma. We will use a concise argument to prove a subset B of a group G is a subgroup. We show that $B^{-1} \subseteq B$ (closure under inversion) and $BB \subseteq B$ (closure under products).

Lemma 3.1.2. *Let H, K be subgroups of G.*

$$HK = KH \qquad \Longleftrightarrow \ KH \textit{ is a subgroup of } G$$

Proof. Suppose $HK = KH$ we will show KH is a subgroup of G. We see KH is closed under inversion: $(kh)^{-1} = h^{-1}k^{-1} \in HK = KH$. We can show that KH is also closed under products with an element-wise argument, but let's use the associativity identified in Notation 2.6.1,

$$(KH)(KH) = K(HK)H = K(KH)H = (KK)(HH) = KH.$$

Since KH is closed under inversion and under products, it is a subgroup of G.

Suppose KH is a subgroup of G. Since KH is closed under inversion, $KH = (KH)^{-1} = H^{-1}K^{-1} = HK$. This gives the reverse implication of the lemma. □

Suppose now that H, N are subgroups of G with N *normal in* G. We can conclude the following.

- $HN = NH$ since $gN = Ng$ for any $g \in G$.
- HN is therefore a subgroup of G by the lemma.
- N is normal in HN, since it is normal in any subgroup of G that contains it.

Theorem 3.1.3 (Second Isomorphism)**.** *Let N be normal in G and H a subgroup of G. Then $HN \cap N$ is normal in H and $H/(H \cap N) \cong HN/N$.*

When these groups are finite we may take cardinalities to get

$$|H||N| = |HN||H \cap N|$$

Proof. Consider $G \xrightarrow{\pi} G/N$ restricted to the subgroup H, and call the restricted homomorphism $\pi' : H \longrightarrow G/N$. The kernel of π' is $H \cap N$. The image is $HN/N = \{hN : h \in H\}$ and HN is a subgroup of G as we noted above. By the first isomorphism theorem, $H/(H \cap N) \cong HN/N$. □

The following special case is of interest. It combines the assumptions of Theorem 3.1.3 with those of Lemma 3.1.1.

Corollary 3.1.4. *Let H and N be subgroups of G with $N \trianglelefteq G$. Suppose $G = HN$ and $H \cap N = \{e_G\}$. Then $G/N \cong H$.*

Proof. One simply substitutes G for HN and notes $H \cong H/\langle e_G \rangle$. □

Definition 3.1.5. In the situation of the corollary we say that G is the **internal semi-direct product** of N by H and we write $G \cong N \rtimes H$.

Note that the order is important: $N \rtimes H$ and $H \rtimes N$ mean two different things. The first assumes N is normal in G and the second assumes H is normal in G.

If both H and N are normal then the two semidirect products are isomorphic to each other and to the direct product, as the following corollary shows.

Corollary 3.1.6. *Suppose $K \trianglelefteq G$ and $N \trianglelefteq G$ and $G = KN$ and $K \cap N = \{e_G\}$. Then elements of K and N commute: for any $k \in K$ and $n \in N$, $kn = nk$. Furthermore, $G \cong K \times N$.*

Proof. To prove that elements of K and N commute with each other it is sufficient to show that $knk^{-1}n^{-1} = e$. Since N is normal, $knk^{-1} \in N$ and therefore $knk^{-1}n^{-1} \in N$. Similarly, since K is normal, $nk^{-1}n^{-1} \in K$ so $knk^{-1}n^{-1} \in K$. Now $K \cap N = \{e\}$ gives the result.

Consider the map $K \times N \xrightarrow{\varphi} G$ defined by $(k, n) \longmapsto kn$. The map is well defined. It is injective since $kn = e$ gives $k = n^{-1} \in K \cap N = \{e\}$. It is surjective since $G = KN$. It respects multiplication (so is a homomorphism):

$$\begin{aligned}
\varphi\big((k_1, n_1)\big)\varphi\big((k_2, n_2)\big) &= (k_1n_1)(k_2n_2)\\
&= k_1(n_1k_2)n_2\\
&= k_1(k_2n_1)n_2\\
&= (k_1k_2)(n_1n_2)\\
&= \varphi\big((k_1k_2, n_1n_2)\big)\\
&= \varphi\big((k_1, n_1)(k_2, n_2)\big)
\end{aligned}$$

Thus φ is an isomorphism. □

Definition 3.1.7. In the situation of the last corollary, G is often called the **internal direct product** of K and N.

The distinction between internal direct product and the usual (external) direct product of two arbitrary groups G and H is subtle as the next examples show. The first example shows that the external direct product of two groups is also the internal direct product of two of its subgroups (in a way that seems perhaps pedantic). The second example shows the real motivation for distinguishing internal direct products, they can be rather hidden, and give insight into the structure of a group.

Example 3.1.8. Let G and H be two groups and consider the external direct product $G \times H$. Let $\overline{G} = G \times \{e_H\}$ and similarly $\overline{H} = \{e_G\} \times H$. The interesection of $\overline{G}$ and $\overline{H}$ is the identity element of $G \times H$ and it is easy to see that every element of $G \times H$ may be written as a product of something in $\overline{G}$ and $\overline{H}$. Thus $G \times H$ is the internal direct product of $\overline{G}$ and $\overline{H}$.

Example 3.1.9. Consider $\mathbb{Z}_6$. It has two proper subgroups $K = \{0, 3\}$ and $N = \{0, 2, 4\}$ both of which are normal since $\mathbb{Z}_6$ is abelian. It can be seen that everything in $\mathbb{Z}_6$ can be written as a sum of something in K and something in N, and clearly $K \cap N = \{0\}$. Thus $\mathbb{Z}_6$ is the internal direct product of H and K. Of course, in $\mathbb{Z}_6$, the subgroup $\{0, 3\}$ is isomorphic to $\mathbb{Z}_2$, and $\{0, 2, 4\}$ is isomorphic to $\mathbb{Z}_3$ and we know from Corollary 2.3.7 that $\mathbb{Z}_6$ is isomorphic to the external direct product $\mathbb{Z}_2 \times \mathbb{Z}_3$. More generally, for m and n coprime, $\mathbb{Z}_{mn}$ is the internal direct product of its subgroups $\langle m \rangle$ and $\langle n \rangle$.

One can also define the external semi-direct product of of two groups.

Definition 3.1.10. Let N, H be two groups and let $\varphi : H \longrightarrow \mathrm{Aut}(N)$ be a homomorphism. Write $\varphi(h)$ as φ_h. Define a new group with elements $N \times H$ and multiplication defined by

$$(n_1, h_1) * (n_2, h_2) = (n_1 \varphi_{h_1}(n_2), h_1, h_2)$$

This is the **external semi-direct product** of N and H defined by φ and is written $N \rtimes_\varphi H$.

The relationship between the internal and external semi-direct product is even more subtle than that for the internal and external direct product. Consider a group G with two subgroups N, H with N normal and H not normal and such that $NH = G$ and $N \cap H = \{e\}$. There is a bijective map from the Cartesian product $N \times H$ to G taking (n, h) to nh. It is not a homomorphism. But it is true that in G,

$$\begin{aligned} n_1 h_1 n_2 h_2 &= n_1 h_1 n_2 h_1^{-1} h_1 h_2 \\ &= n_1 \varphi_{h_1}(n_2) h_1 h_2 \end{aligned}$$

Where φ_{h_1} is conjugation by h_1. Because N is normal in G, φ_h is an automorphism of N and, indeed, we have a homomorphism $\varphi : H \longrightarrow \mathrm{Aut}(N)$ that takes h to φ_h. Thus, if we use φ to to define the external semi-direct product $N \trianglelefteq_\varphi H$ we get an isomorphism with G.

Exercises 3.1.11. The definition of external semi-direct product makes sense!

(a) Verify that (e_H, e_K) is the identity element.

(b) Show that each element does have an inverse.

(c) Show that the associative law holds.

Exercises 3.1.12. Verify the following are semi-direct products.

(a) $D_n \cong C_n \rtimes_\varphi C_2$ where $\varphi : C_2 \longrightarrow \mathrm{Aut}(C_n)$ takes the non-identity element of C_2 to the automorphism of C_n taking n to n^{-1}.

(b) $S_n = A_n \rtimes \langle (1,2) \rangle$. What is the map φ?

(c) $S_4 = V \rtimes S_3$ where V is Klein-4 subgroup with elements of the form $(a,b)(c,d)$ with a, b, c, d distinct elements of $\{1,2,3,4\}$. What is the map φ?

(d) In $\mathrm{GL}_n(F)$, for F a field, let T be the upper triangular matrices with nonzeros on the diagonal; let U be the upper triangular matrices with 1's on the diagonal and let D be the diagonal matrices with nonzero elements on the diagonal. For $n = 2$, show that $T = U \rtimes D$. Describe the map $\varphi : D \longrightarrow \mathrm{Aut}(U)$.

(e) Do the previous problem for arbitrary n.

Proposition 3.1.13. *Let N be a normal subgroup of G and let $\pi : G \longrightarrow G/N$ be the quotient homomorphism. Suppose that there is a homomorphism $\alpha : G/N \longrightarrow G$ such that $\pi \circ \alpha$ is the identity map on G/N. Then G is the internal direct product $N \rtimes \alpha(G/N)$.*

Proof. Let the image of α be $H = \alpha(G/N)$, which is a subgroup of G. By Corollary 3.1.4, we need only show that $HN = G$ and that $H \cap N$ is trivial.

Let $g \in G$. Let $h = \alpha \circ \pi(g)$. This is an element of H, since it is in the image of α. I claim $gh^{-1} \in N$. This is because

$$\begin{aligned}\pi(gh^{-1}) &= \pi(g)\pi(h^{-1}) = \pi(g)\pi\big(\alpha \circ \pi(g^{-1})\big) = \pi(g)\big(\pi \circ \alpha \circ \pi\big)(g^{-1})\big) \\ &= \pi(g)(\pi(g^{-1})\big) = \pi(g)\pi(g^{-1}) = eN\end{aligned}$$

Consequently $gh^{-1} = n$ for some $n \in N$ and therefor $g = hn$. Since g was an arbitrary element of G we have shown $G = HN$.

Now suppose that $h \in H \cap N$. Since $h \in H$, there is some $gN \in G/N$ such that $h = \alpha(gN)$. We know that $\pi \circ \alpha$ is the identity on H, so

$$\alpha\Big(\pi\big(\alpha(g)\big)\Big) = \big(\alpha \circ \pi\big)\big(\alpha(g)\big) = \alpha(g) = h.$$

On the other hand, since $h \in N$,

$$\alpha\Big(\pi\big(\alpha(g)\big)\Big) = \alpha\big(\pi(h)\big) = \alpha(eN) = e.$$

Consequently, $h = e$ and we have shown $H \cap N = \{e\}$. □

Exercises 3.1.14. Let F be a field. Let $\mathrm{GL}_n(F)$ be the general linear group: $n \times n$ matrices over F with nonzero determinant. Let $\mathrm{SL}_n(F))$ be the special linear group: matrices with determinant 1. Let F^*I be the nonzero multiples of the identity matrix. In this problem we investigate the finite fields F and values of n for which $\mathrm{GL}_n(F) \cong \mathrm{SL}_n(F) \times F^*I$.

(a) For the fields $F = \mathbb{F}_3$ and $F = \mathbb{F}_5$, show that $\mathrm{GL}_n(F)$ is a direct product as above for n odd, but not for n even.

(b) For the field $F = \mathbb{F}_7$, show that $\mathrm{GL}_n(F)$ is a direct product as above for n coprime to 6, and is not otherwise.

(c) (Challenge) For which fields $\mathbb{F}_q$ and which n is $\mathrm{GL}_n(\mathbb{F}_q)$ a direct product as above?

Exercises 3.1.15. External semi-direct products of cyclic groups.

(a) Use the definition of external semi-direct product to create the other non-abelian group of order 12 (besides D_6 and A_4), $\mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_4$ where φ is the only possible map $\mathbb{Z}_4 \longrightarrow \mathrm{Aut}(\mathbb{Z}_3)$ that is not trivial. Let a be the generator for $\mathbb{Z}_3$ and b the generator for $\mathbb{Z}_4$. Show the following:

(1) Every element can be represented uniquely as a^ib^j for $i \in \{0, 1, 2\}$ and $b \in \{0, 1, 2, 3\}$

(2) The group can be presented as $\langle a, b | a^3 = b^4 = 1, ba = a^2b\rangle$

(3) Find the inverse of a^ib^j.

(4) Find a general formula for $a^ib^j * a^mb^n$. It may be useful to break this into cases.

(b) Use the definition of external semi-direct product to create the only non-abelian group of order 21 (the smallest non-abelian group of odd order), $\mathbb{Z}_7 \rtimes \mathbb{Z}_3$. Let a be the generator for $\mathbb{Z}_7$ and b the generator for $\mathbb{Z}_3$. Show how to represent, invert, and multiply elements of this group as you did in the previous problem.

(c) (Challenge Problem) Use the definition of external semi-direct product to construct semi-direct products $\mathbb{Z}_m \rtimes \mathbb{Z}_n$. You will need to start with a homomorphism $\varphi : \mathbb{Z}_n \longrightarrow \mathrm{Aut}(\mathbb{Z}_m)$. See how many of the small non-abelian groups you can find in the table of small abelian groups on Wikipedia.

3.2 Finitely Generated Abelian Groups

In this section we show that the structure of finitely generated abelian groups is fairly simple. Any finitely generated abelian group is isomorphic to a direct product of cyclic groups that can be put in a standard, uniquely determined, format. We proceed in several steps, each subsection below gives a complete story about a particular class of abelian groups; each extends the result of the previous subsection to a broader class of abelian groups.

Our first step is to show that a direct product of cyclic groups can be put into a standard format that elucidates its structure. There are actually two such formats, one using *elementary divisors* and the other using *invariant factors*. In particular, two groups are isomorphic if and only if their standard formats are the same. Our next step is to show that any *finite abelian* group is actually a direct product of cyclic groups, and it therefore can be placed in the two standard formats. This result has one very technical lemma whose proof we sketch. Finally, we state and prove some aspects of the more general result that any *finitely generated abelian* group can be written as a direct product of a finite group (with standard formats above) and a group that is isomorphic to $\mathbb{Z}^r$ for some integer r.

We will write the group operation additively. For A an abelian group, $a \in A$, and m an integer, we write mA for $a + \cdots + a$ with m summands. Think of ma as repeated addition, not multiplication. The order of a is the smallest positive integer m such that $ma = 0$. One can check that $ma + na = (m + n)a$ and $(mn)a = m(na)$. If B is a subgroup of A (it is normal since A is abelian) we write a coset as $a + B$ and the identity element of A/B is $0 + B$.

A key tool in this chapter is Corollary 2.3.7, which says that for coprime integers m and n the group $\mathbb{Z}_{mn}$ is isomorphic to $\mathbb{Z}_m \times \mathbb{Z}_n$. The corollary says more, that there is a unique isomorphism that takes $[1]_{mn}$ to $\big([1]_m, [1]_n\big)$, but we only need the existence of the isomorphism in this section. An easy induction argument establishes the following result

Proposition 3.2.1. *Let $m_1, m_2, \ldots, m_t$ be pairwise coprime positive integers and let $m = \prod_{i=1}^{t} m_i$, then*

$$\mathbb{Z}_m \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_t}$$

Products of Cyclic Groups

Let's start with abelian groups that we understand well, cyclic groups, and direct products of cyclic groups. The notation in the theorems below is a bit heavy, so we start with an example.

Example 3.2.2. Consider the group $\mathbb{Z}_{60} \times \mathbb{Z}_{12} \times \mathbb{Z}_8 \times \mathbb{Z}_{25}$. Using Proposition 3.2.1,

$$\begin{aligned}
\mathbb{Z}_{60} &\cong \mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \\
\mathbb{Z}_{12} &\cong \mathbb{Z}_4 \times \mathbb{Z}_3 \\
\mathbb{Z}_8 &\cong \mathbb{Z}_8 \\
\mathbb{Z}_{75} &\cong \mathbb{Z}_3 \times \mathbb{Z}_{25}
\end{aligned}$$

Let's take the direct product of all these factors ordering them by the prime involved (2, 3,or 5) and for each prime, the highest power of that prime first.

$$\begin{aligned}
\mathbb{Z}_{60} \times \mathbb{Z}_{12} \times \mathbb{Z}_8 \times \mathbb{Z}_{25} \cong &\quad \mathbb{Z}_8 \times \mathbb{Z}_4 \times \mathbb{Z}_4 \\
&\times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \\
&\times \mathbb{Z}_{25} \times \mathbb{Z}_5
\end{aligned} \tag{3.1}$$

Now, we regroup by combining the highest powers of each prime.

$$\begin{aligned}
\cong &\quad \mathbb{Z}_8 \times \mathbb{Z}_3 \times \mathbb{Z}_{25} \\
&\times \mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \\
&\times \mathbb{Z}_4 \times \mathbb{Z}_3
\end{aligned}$$

Finally, we have

$$\mathbb{Z}_{60} \times \mathbb{Z}_{12} \times \mathbb{Z}_8 \times \mathbb{Z}_{25} \cong \mathbb{Z}_{600} \times \mathbb{Z}_{60} \times \mathbb{Z}_{12} \tag{3.2}$$

Both factorizations are of interest: one (3.1) into cyclic groups of prime power order, the other (3.2) combining the factors prime power factors in a greedy fashion.

The following proofs are just adaptations of the computations in the example to deal with the general context.

Theorem 3.2.3. *Let $m_1, \ldots, m_t$ be positive integers and $A = \mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_t}$. Let $P = \{p_1, \ldots, p_s\}$ be the set of all primes dividing $m_1 m_2 \cdots m_t$ and let the m_j have factorizations $m_j = \prod_{i=1}^{s} p_i^{e_{ij}}$ (allowing some $e_{ij} = 0$). Then*

$$A \cong A_1 \times \cdots \times A_s$$

where $A_i = \mathbb{Z}_{p_i}^{e_{i1}} \times \mathbb{Z}_{p_i}^{e_{i2}} \times \cdots \mathbb{Z}_{p_i}^{e_{it}}$.

Furthermore $|A_i| = p^{e_i}$ where $e_i = \sum_{j=1}^{t} e_{ij}$.

Proof. By Proposition 3.2.1, $\mathbb{Z}_{m_j} \cong \mathbb{Z}_{p_1}^{e_{1j}} \times \cdots \times \mathbb{Z}_{p_s}^{e_{sj}}$. Thus

$$\begin{aligned} A = & \quad \mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_t} \\ \cong & \quad \mathbb{Z}_{p_1}^{e_{11}} \times \cdots \times \mathbb{Z}_{p_s}^{e_{s1}} \\ & \times \mathbb{Z}_{p_1}^{e_{12}} \times \cdots \times \mathbb{Z}_{p_s}^{e_{s2}} \\ & \qquad \cdots \\ & \times \mathbb{Z}_{p_1}^{e_{1t}} \times \cdots \times \mathbb{Z}_{p_s}^{e_{st}} \end{aligned}$$

Rearranging terms so that the jth column of factors becomes the jth row, and the ith row becomes the ith column, we have

$$\begin{aligned} \cong & \quad \mathbb{Z}_{p_1}^{e_{11}} \times \cdots \times \mathbb{Z}_{p_1}^{e_{1t}} \\ & \times \mathbb{Z}_{p_2}^{e_{21}} \times \cdots \times \mathbb{Z}_{p_2}^{e_{2t}} \\ & \qquad \cdots \\ & \times \mathbb{Z}_{p_s}^{e_{s1}} \times \cdots \times \mathbb{Z}_{p_s}^{e_{st}} \\ \cong & \quad A_1 \times \cdots \times A_s \end{aligned}$$

The cardinality of A_i is just the product of the cardinalities of its factors. So, letting $e_i = \sum_{j=1}^{t} e_{ij}$, we have $|A_i| = p^{e_i}$. □

Definition 3.2.4. The multiset $\left\{p_i^{e_{ij}} : i = 1, \ldots, s; \text{ and } j = 1, \ldots t\right\}$ is the set of **elementary divisors** of A.

Theorem 3.2.5. *With the notation of the previous theorem, for each i let $f_{i1} \geq f_{i2} \cdots \geq f_{it}$ be a permutation of the exponents $e_{i1}, \ldots e_{it}$ putting them in decreasing order. For $j = 1, \ldots, t$, let $n_j = \prod_{i=1}^{s} p_i^{f_{ij}}$. Then $n_t \mid n_{t-1} \mid \cdots \mid n_1$ and $A \cong \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_t}$.*

Proof. The fact that $n_j \mid n_{j-1}$ follows from $f_{ij} \leq f_{i,j-1}$. Revisiting the previous proof, we enter after the point where we rearranged the factors. In each line we then permute the e_{ij} to have them in decreasing order $(f_{i1}, \ldots, f_{it})$. The final step is to rearrange again by combining all the largest prime power factors to create $\mathbb{Z}_{n_1}$ and proceeding iteratively with the next largest prime power factors. As with the previous theorem this is just an application of Proposition 3.2.1.

$$\begin{aligned}
A = \quad & \mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_t} \\
\cong \quad & \mathbb{Z}_{p_1}^{e_{11}} \times \cdots \times \mathbb{Z}_{p_1}^{e_{1t}} \\
& \times \mathbb{Z}_{p_2}^{e_{21}} \times \cdots \times \mathbb{Z}_{p_2}^{e_{2t}} \\
& \qquad \cdots \\
& \times \mathbb{Z}_{p_s}^{e_{s1}} \times \cdots \times \mathbb{Z}_{p_s}^{e_{st}} \\
\cong \quad & \mathbb{Z}_{p_1}^{f_{11}} \times \cdots \times \mathbb{Z}_{p_1}^{f_{1t}} \\
& \times \mathbb{Z}_{p_2}^{f_{21}} \times \cdots \times \mathbb{Z}_{p_2}^{f_{2t}} \\
& \qquad \cdots \\
& \times \mathbb{Z}_{p_s}^{f_{s1}} \times \cdots \times \mathbb{Z}_{p_1}^{f_{st}} \\
\cong \quad & \mathbb{Z}_{p_1}^{f_{11}} \times \cdots \times \mathbb{Z}_{p_s}^{f_{s1}} \\
& \times \mathbb{Z}_{p_1}^{f_{12}} \times \cdots \times \mathbb{Z}_{p_s}^{f_{s2}} \\
& \qquad \cdots \\
& \times \mathbb{Z}_{p_1}^{f_{1t}} \times \cdots \times \mathbb{Z}_{p_s}^{f_{st}} \\
\cong \quad & \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_t}
\end{aligned}$$

□

Definition 3.2.6. The n_j (that are not 1) in the previous theorem are called the **invariant factors** of A.

Exercises 3.2.7.

(a) Find the elementary divisors and the invariant factors for $\mathbb{Z}_5 0 \times \mathbb{Z}_7 5 \times \mathbb{Z}_1 36 \times \mathbb{Z}_2 1000$.

(b) Let $n_1, n_2, \ldots, n_r$ be integers larger than 1. Under what conditions will $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_r}$ have r invariant factors?

Exercises 3.2.8. There is a well defined homomorphism $\mathbb{Z}_{2400} \longrightarrow \mathbb{Z}_d$ that takes $[1]_{2400}$ to $[1]_d$ for any d that divides 2400. Given several divisors of 2400 we can use the universal property of a direct product 2.3.6 to get a homomorphism into the direct product of several such groups, for example

$$\mathbb{Z}_{2400} \xrightarrow{\varphi} \mathbb{Z}_{40} \times \mathbb{Z}_{30} \times \mathbb{Z}_{16}$$

(a) What is the kernel?

(b) Find the elementary divisors of (i) $\mathbb{Z}_{2400}$, (ii) $\mathbb{Z}_{40} \times \mathbb{Z}_{30} \times \mathbb{Z}_{16}$, and (iii) the kernel of the homomorphism φ.

(c) Find the invariant factors of (i) $\mathbb{Z}_{2400}$, (ii) $\mathbb{Z}_{40} \times \mathbb{Z}_{30} \times \mathbb{Z}_{16}$, and (iii) the kernel of the homomorphism φ.

Finite Abelian Groups

In the previous section we showed that a product of cyclic groups can be written in two different forms that illuminate the structure better. One form uses cyclic groups of prime power order (and gives the elementary divisors of the group) and the other uses a format that identifies the largest cyclic component and, after splitting off that component, the next largest cyclic component, and so forth. This gives the invariant factors of the group. We now want to show that this classification applies to any finite abelian group.

The first step is to split a group into pieces that are, in a sense, coprime. We then apply induction to write the group as a direct product of groups that have prime power order. The difficult step is to show that a group of prime power order is actually a product of cyclic groups (whose orders are a power of the same prime).

Definition 3.2.9. Let A be an abelian group. For $m \in \mathbb{N}$ let

$$\begin{aligned} mA &= \{ma : a \in A\} \\ A[m] &= \{a : ma = 0\} \end{aligned}$$

For p a prime define the p**-torsion subgroup** of A to be

$$A(p) = \left\{a \in A : \operatorname{ord}(a) = p^k \text{ for some } k\right\}$$

An abelian group such that $A = A(p)$ is called a p-group.

Exercises 3.2.10.

(a) Prove that mA, $A[n]$ and $A(p)$ are all subgroups of A.

(b) Prove that $A(p) = \cup_{i=0}^{\infty} A[p^i]$, and that, for A finite, $A(p) = A[p^k]$ for some large enough k.

Proposition 3.2.11. *Suppose that A is abelian with $|A| = mn$ and m, n coprime. Then*

(1) $mA = A[n]$

(2) A is the internal direct product of $A[m]$ and $A[n]$

Proof. Let $u, v \in \mathbb{Z}$ be such that $um + vn = 1$. Let $a \in A[n]$. Then $a = (um + vn)a = u(ma) + v(na) = m(ua)$, since we assume $na = 0$. This shows that $A[n] \subseteq mA$. On the other hand, an arbitrary element of mA can be written ma for $a \in A$. Since $|A| = mn$, $n(ma) = (nm)a = 0$, and this shows $mA \subseteq A[n]$ (we have used Lagrange's Theorem that the order of a divides the order of the group).

For the second claim of the proposition, we show that $A[m] \cap A[n] = \{0\}$ and that $A[m] + A[n] = A$. Then, by Corollary 3.1.6, $A \cong A[m] \times A[n]$.

Let $a \in A$. Since $a = (mu + nv)a = m(ua) + n(va)$ we see that $a \in mA + nA$, which by the previous paragraph is equal to $A[n] + A[m]$. Thus $A[m] + A[n] = A$. On the other hand, if $a \in A[m] \cap A[n]$ then and $a = (um+vn)a = u(ma)+v(na) = 0+0$. Thus $A[m] \cap A[n] = \{0\}$. We have shown that A is the internal direct product of $A[m]$ and $A[n]$. □

The next proposition shows that our decomposition is uniquely determined.

Proposition 3.2.12. *Let A_1, A_2, B_1 and B_2 be finite groups. Suppose that $A_1 \times B_1 \cong A_2 \times B_2$ where everything in A_i has order dividing m and everything in B_i has order dividing n, with m and n coprime. Then $A_1 \cong A_2$ and $B_1 \cong B_2$.*

Proof. Assume $A_1 \times B_1 \cong A_2 \times B_2$.

$$\begin{aligned} m(A_i \times B_i) &= mA_i \times mB_i \\ &= \{(0, mb) : b \in B_i\} \\ &= \{0\} \times B_i \end{aligned}$$

The first step because $m(a, b) = (ma, mb)$ and the last step because multiplication by m (coprime to n) gives an automorphism of B_i. Since $m(A_1 \times B_1) \cong m(A_2 \times B_2)$ we get $B_1 \cong B_2$. Similarly we show $A_1 \cong A_2$. □

Corollary 3.2.13. *Let $|A| = p_1^{e_1} \dots p_s^{e_s}$ then*

$$A \cong A[p_1^{e_1}] \times \cdots \times A[p_s^{e_s}] = A(p_1) \times \cdots \times A(p_s)$$

This factorization is unique up to reordering.

Proof. Existence of the factorization follows from Proposition 3.2.11 by induction:

$$\begin{aligned} A &\cong A[p_1^{e_1}] \times A[p_2^{e_2} p_3^{e_3} \cdots p_s^{e_s}] \\ &\cong A[p_1^{e_1}] \times A[p_2^{e_2}] \times A[p_3^{e_3} p_4^{e_4} \cdots p_s^{e_s}] \end{aligned}$$

and so forth. There is one subtlety though; we have used $A[p_2^{e_2}] = \Big(A[p_2^{e_2} p_3^{e_3} \cdots p_s^{e_s}]\Big)[p_2^{e_2}]$. This is easily verified. Any nonzero element of A whose order is a power of p_2 is in each of these groups, and nothing else is.

Uniqueness follows from a similar inductive application of Proposition 3.2.12 □

The previous corollary is the first step in the classification of finite abelian groups. The next step is to classify groups satisfying the following definition.

Definition 3.2.14. Let p be a prime number. An abelian **p-group** is an abelian group A in which every element of A has order that is a power of p.

The key lemma follows. Its proof is quite technical and not very illuminating, so I sketch the proof in [Hun12][Sec 8.2].

Lemma 3.2.15. *Let A be an abelian p-group and let a be an element of maximal order. Then $A = K + \langle a \rangle$ and $K \cap \langle a \rangle = \{0\}$ for some subgroup K of A. Thus A is isomorphic to the direct product of K and $\langle a \rangle$.*

Proof. Let a be an element of maximal order in the abelian p-group A; this order is a power of p. Let K be as large as possible such that $K \cap \langle a \rangle = \{0\}$. We want to show that $K + \langle a \rangle = A$. Then Corollary 3.1.6 says that $A \cong K \times \langle a \rangle$.

Suppose $b \in A \setminus (K + \langle a \rangle)$. Do some tricks to show:

(1) There is a $c \in A \setminus (K + \langle a \rangle)$ such that $pc \in K + \langle a \rangle$. [Take the minimal r such that $p^r b \in K + \langle a \rangle$, then let $c = p^{r-1}b$.]

(2) There is a $d \in A \setminus (K + \langle a \rangle)$ such that $pd \in K$. [Let $pc = k + ma$, argue that $m = pm'$, for some integer m' using that a has maximal degree in A and $K \cap \langle a \rangle = \{0\}$. Then set $d = c - m'a$.]

By assumption on K, $(K + \langle d \rangle) \cap \langle a \rangle \neq \{0\}$, so there is some $k \in K$, and nonzero $r, s \in \mathbb{Z}$ such that $k + rd = sa$.

Now we consider two cases: If $p \mid r$ then $rd \in K$ and consequently $sa \in K$. This contradicts $K \cap \langle a \rangle = \{0\}$. If $p \nmid r$ then there are u, v such that $up + vr = 1$. Then $d = u(pd) + v(rd)$. The first term is in K and the second in $K + \langle a \rangle$, so $d \in K + \langle a \rangle$, which is a contradiction.

Summarizing, we assumed K maximal such that $K \cap \langle a \rangle = \{0\}$. Supposing the existence of some $b \in A \setminus (K + \langle a \rangle)$, we showed there was some $d \in A \setminus (K + \langle a \rangle)$ such that $pd \in K$. From the maximality assumption on K, there is some element of $(K + \langle d \rangle) \cap \langle a \rangle$, which we can write as $k + rd = sa$ for $k \in K$ and $r, s \in \mathbb{Z}$. There are two possibilities, p divides r or not. Both lead to a contradiction. Thus A must be equal to $K + \langle a \rangle$. □

Theorem 3.2.16. *Let $A = A(p)$ be an abelian p-group. Then A is the direct product of cyclic groups each of which has order a power of p. Consequently, the order of A is also a power of p.*

The decomposition is unique (up to reordering). Put another way, two abelian p-groups are isomorphic if and only if their decompositions have the same number of factors for each power of p.

Proof. Note first that a non-trivial abelian p-group must have order divisible by p since Lagrange's Theorem 2.6.6 says that the order of an element of the group (which, by assumption, is a power of p) must divide the order of the group. We will show that the order of an abelian p-group must actually be equal to a power of p.

We proceed by induction on the largest power of p that divides the cardinality of the abelian group. Our induction hypothesis for $t \geq 0$ is that any abelian p-group with order that is divisible by p^t but not by p^{t+1} is isomorphic to the direct product of cyclic groups whose orders are a power of p. An immediate consequence is that the order of such a group is a power of p, so the order is exactly p^t. We noted above that the induction hypothesis is true for $t = 0$, in which case the abelian p-group is trivial.

Let $t \geq 1$. Let A be an abelian p-group of cardinality $p^t m$ with m not divisible by p. Using the lemma we can write A as a direct sum $A = K + \langle a \rangle$ with $K \cap \langle a \rangle = \{0\}$. The subgroup $\langle a \rangle$ is cyclic of order p^s for some $s > 0$ and therefore $|K| \leq p^{t-s}m$. Applying the induction hypothesis to K, shows that $m = 1$ and K is isomorphic to the direct product of cyclic groups whose order is a power of p. Since $\langle a \rangle$ is also cyclic of order p^s, we have that A is the direct product of cyclic groups of order a power of p.

To prove uniqueness (up to reordering) we note first that if two groups have the same number of factors for each power of p they are isomorphic. We will write factorizations by writing the factors in increasing powers of p as follows

$$A \cong (\mathbb{Z}_p)^{k_1} \times (\mathbb{Z}_p^2)^{k_2} \times \cdots \times (\mathbb{Z}_p^r)^{k_r}$$

We now show that we can recover the k_i by operating on A. In other words, two factorizations, one with $k_1, k_2 \ldots$ and one with $m-1, m_2, \ldots$ are isomorphic if and only if $k_i = m_i$.

We can recover the k_i iteratively. Since $\log_p(|\mathbb{Z}_p^n|) = n$, we have $\log_p(|A|) = \sum_{i=1}^r ik_i$. Notice that $p^{n-1}\mathbb{Z}_p^n \cong \mathbb{Z}_p$ and $p^k\mathbb{Z}_p^n$ is trivial for $k \geq n$. Thus the subgroup $p^{r-1}A$ is isomorphic to

$$p^{r-1}A \cong (\mathbb{Z}_p)^{k_r}$$

Thus we have $\log_p(|p^{r-1}A|) = k_r$. Similar computations for p^iA with $i = r-2, r-3, \ldots, 1$ allows one to recover the other k_i. (Try it as an exercise!)

□

From Corollary 3.2.13 and the previous theorem we obtain the fundamental theorem for finite abelian groups..

Theorem 3.2.17 (Fundamental Theorem of Finite Abelian Groups)**.** *Let A be an abelian group of order $p_1^{e_1} \dots p_r^{e_r}$. Then A is a direct product of cyclic groups, each having order a power of one of the p_i. If we write*

$$A(p_i) \cong \mathbb{Z}_{p_i^{e_{i,1}}} \times \mathbb{Z}_{p_i^{e_{i,2}}} \times \dots \mathbb{Z}_{p_i^{e_{i,s_i}}}$$

then for each i, $\sum_{\ell=1}^{s_i} e_{i,\ell} = e_i$. The decomposition is unique, up to reordering.

Exercises 3.2.18. Consider the following problems for $n = 72000$ and $n = 84000$.

(a) Classify all abelian groups of order 84,000.

(b) Let $n = 72,000$. How many abelian groups are there of order n?

(c) How many of these abelian groups have 2 invariant factors?

(d) How many abelian groups are there of order $p^6q^5r^4$ where p, q, r are distinct primes?

(e) How many have k invariant factors, for $k = 1, 2, 3, 4, 5, 6$? Check your answer against the response to the previous question.

Exercises 3.2.19. Let p, q and r be prime and let $n = p^6q^2r^3$.

(a) How many abelian groups are there of order n?

(b) How many of these groups have exactly *two* invariant factors?

Exercises 3.2.20. Here is another approach to proving uniqueness in the classification of finite abelian groups.

(a) Show that $p^k\mathbb{Z}_{p^n} \cong \mathbb{Z}_{p^{n-k}}$ for $k \leq n$. Seen another way, there is an exact sequence

$$0 \longrightarrow \mathbb{Z}_{p^m} \xrightarrow{\cdot p^k} \mathbb{Z}_{p^{k+m}} \longrightarrow \mathbb{Z}_{p^k} \longrightarrow 0$$

(b) Show that $p^{k-1}\mathbb{Z}_{p^n} \Big/ p^k\mathbb{Z}_{p^n} \cong \mathbb{Z}_p$ for $k \leq n$.

(c) Suppose that $A \cong (\mathbb{Z}_p)^{k_1} \times (\mathbb{Z}_{p^2})^{k_2} \times \cdot \times (\mathbb{Z}_{p^n})^{k_n}$. Show that $p^{t-1}A/p^tA \cong (\mathbb{Z}_p)^{k_t+\dots+k_n}$.

(d) Conclude the uniqueness part of the classification of finite abelian groups: If

$$(\mathbb{Z}_p)^{k_1} \times (\mathbb{Z}_{p^2})^{k_2} \times \cdot \times (\mathbb{Z}_{p^n})^{k_n} \cong (\mathbb{Z}_p)^{m_1} \times (\mathbb{Z}_{p^2})^{m_2} \times \cdot \times (\mathbb{Z}_{p^n})^{m_n}$$

then $k_i = m_i$.

Finitely Generated Abelian Groups

Our final step is to generalize the results on classification to finitely generated abelian groups. The strategy is simple. Let A be a finitely generated group. The **torsion subgroup** of A is the set of elements of finite order. The first proposition below shows that it is indeed a subgroup of A. The torsion subgroup has to be a finite group, since it is finitely generated, so it is classified by the results above. The next big step is to show that A has a subgroup that is isomorphic to $\mathbb{Z}^r$ such that A is the direct product of that subgroup and $\operatorname{Tor}(A)$. Thus, a finitely generated abelian group is isomorphic to a finite direct product of cyclic groups that are either infinite of prime power order.

Proposition 3.2.21. *Let A be an abelian group. Let*

$$\operatorname{Tor}(A) = \{a \in A : a \textit{ has finite order}\}$$

(1) $\operatorname{Tor}(A)$ *is a normal subgroup of A.*

(2) All elements of $A/\operatorname{Tor}(A)$ (except the identity) have infinite order.

Proof. This was Exercise 2.10.4. Let $T = \operatorname{Tor}(A)$. Clearly $0 \in T$, so T is nonempty. If $a \in T$ has order m then so does $-a = (m-1)a$. If b is another element in T and it has order n, then $a + b$ has order at most mn since

$$mn(a + b) = (mn)a + (mn)b = n(ma) + n(mb) = 0$$

Thus T is closed under inversion and multiplication, so it is a subgroup of A. Normality is immediate since A is abelian.

If $b + T$ has finite order m in A/T then

$$mb + T = m(b + T) = 0 + T$$

This shows that $mb \in T$, so mb has some finite order n in A. Then $(nm)b = n(mb) = 0$, so b itself has finite order. Thus $b \in T$ and $b + T = 0 + T$. So the only element of finite order in $A/\operatorname{Tor}(A)$ is the identity element. □

Definition 3.2.22. A group that has no elements of finite order, other than the identity, is said to be **torsion free**.

Let A be an abelian group and let $S = \{a_1, a_2, \ldots, a_r\}$ be a set of elements in A. We say that the elements of S are **independent** when for any integers $m_1, \ldots m_r$ that are not all zero, $m_1a_1 + m_2a_2 + \cdots + m_ra_r \neq 0$.

A **free abelian group of rank** r is a group that is isomorphic to $\mathbb{Z}^r$. It will have r elements that are independent and also generate A. We will generally use

e_i for the element of $\mathbb{Z}^r$ that is 1 in the ith component and 0 elsewhere. Borrowing from the language of vector spaces, we say that the e_i are the standard basis for $\mathbb{Z}^r$.

Proposition 3.2.23. *The rank of a finitely generated free abelian group is unique.*

Proof. Let $A \cong \mathbb{Z}^r$. Then $A/2A \cong \mathbb{Z}^r/(2\mathbb{Z}^r) \cong (\mathbb{Z}_2\mathbb{Z})^r$. The final isomorphism comes from Theorem 2.7.10 and the observation that $2\mathbb{Z}^r = 2\mathbb{Z} \times 2\mathbb{Z} \times \cdots \times 2\mathbb{Z}$ (both containments are easy to show). Since $(\mathbb{Z}_2\mathbb{Z})^r$ has 2^r elements, we may recover the rank by computing $\log_2(|A/2A|)$. □

Exercises 3.2.24.

- **(a)** Let A be an abelian group. Suppose $f : A \to \mathbb{Z}$ is a surjective homomorphism with kernel K. Show that A has an element a such that A is the internal direct product $K \times \langle a \rangle$.
- **(b)** In the previous problem, suppose f is not surjective but $f(A) = n\mathbb{Z}$ for some $n \in \mathbb{N}$. Show that it still holds that there is an element $a \in A$ such that A is the internal direct product $K \times \langle a \rangle$.
- **(c)** Suppose that A is torsion free and $mA \cong \mathbb{Z}^r$. Show that $A \cong \mathbb{Z}^r$.

We need two results before proving that a finitely generated torsion free abelian group is actually isomorphic to $\mathbb{Z}^r$ for some r.

Proposition 3.2.25. *Let $A \leq \mathbb{Z}^r$. Then A is isomorphic to $\mathbb{Z}^s$ for some integer $s \leq r$.*

Proof. We proceed by induction on r. For $r = 1$ we already know the subgroups of $\mathbb{Z}$. They are the trivial group (rank 0) and $n\mathbb{Z}$, which is isomorphic to $\mathbb{Z}$.

Assume the statement of the theorem is true for integers less than r. Let $A \leq \mathbb{Z}^r$ and let e_i be the element of $\mathbb{Z}^r$ that is 1 in the ith component and 0 elsewhere. Consider projection onto the rth component $\mathbb{Z}^r \xrightarrow{\pi} \mathbb{Z}$. The kernel of this map is $\mathbb{Z}^{r-1}$ with generators $\{e_1, \ldots, e_{r-1}\}$. Let $A \xrightarrow{\iota} \mathbb{Z}^r$ be the embedding of A in $\mathbb{Z}^r$ and consider the composite $\pi \circ \iota$. The kernel of this map is $B = A \cap \mathbb{Z}^{r-1}$. By the induction hypothesis, B is isomorphic to $\mathbb{Z}^s$ for some integer $s \leq r - 1$.

If $B = A$ we are done. Otherwise, $\pi \circ \iota$ has image $n\mathbb{Z}$ for some $n > 0$. Let $a \in A$ be a preimage of n. By Exercise 3.2.24(b) we have $A \cong B \times \langle a \rangle \cong \mathbb{Z}^s \times \mathbb{Z} = \mathbb{Z}^{s+1}$. □

Proposition 3.2.26. *Let A be a finitely generated, torsion-free abelian group. Then A is a free abelian group of finite rank.*

Proof. Let $b_1, \ldots, b_r$ be a maximal independent set in A. These elements are independent and there is no set of $r + 1$ elements in A that is independent.

Let B be the subgroup of A generated by $b_1, \ldots, b_r$, that is $B = \{\sum_{i=1}^{r} m_i b_i : m_i \in \mathbb{Z}\}$. Consider A/B. I claim that it is a torsion group; every element has finite order.

For any $a \in A$, there are integers $m_0, \ldots, m_r$ such that $m_0 a + m_1 b_1 + m_2 b_2 + \cdots + m_r b_r = 0$, for otherwise $\{a, b_1, \ldots b_r\}$ would be an independent set, contradicting maximality of $\{b_1, \ldots b_r\}$. Furthermore m_0 is not zero, since $\{b_1, \ldots b_r\}$ is independent, so $m_0 a$ is a nonzero element of B. Thus for any $a \in A$ there is an m_0 such that $m_0(a + B) = m_0 a + B = 0 + B$.

We have shown every element of A/B has finite order. Since it is also finitely generated it is a finite group. Thus the exponent of A/B (the lcm of the orders of elements of A/B) is some finite $m \in \mathbb{N}$. Then, $m(A/B)$ is the trivial subgroup $\{0 + B\}$ inside A/B. This shows mA is a subgroup of the free abelian group B. Thus by Proposition 3.2.25 mA is free abelian of rank $s \leq r$, where r is the rank of B. Applying Exercise 3.2.24(c) we have that A itself is free of rank s. This concludes the proof.

It is worth noting that, since $B \leq A$, Proposition 3.2.25 says that the rank of B is at most the rank of A, so $r \leq s$. We already showed $s \leq r$, so A and B have the same rank. □

The proof of the following theorem is similar to Proposition 3.1.13. Exercise 3.2.24(a) was a special case (with $r = 1$).

Proposition 3.2.27. *Let A be a finitely generated abelian group. and let $\varphi : A \longrightarrow \mathbb{Z}^r$ be a surjective homomorphism with kernel K. There exists a subgroup $B \leq A$ such that φ restricted to B is an isomorphism. Furthermore $A = K \times B$ (We use $=$ rather than $\cong$ because A is the internal direct product of the two subgroups).*

Proof. Let $b_1, \ldots, b_r \in A$ map to the elements e_i of $\mathbb{Z}^r$. Let $B = \langle b_1, \ldots, b_r \rangle = \{m_1 b_1 + \cdots + m_r b_r : m_i \in \mathbb{Z}\}$. We know that the b_i are independent, because their images in $\mathbb{Z}^r$ are independent. In other words, since φ is a homomorphism, $\sum_{i=1}^{r} m_i b_i = 0$ would imply that $\sum_{i=1}^{r} m_i e_i = 0$. The latter is only true if all m_i are zero. Thus $\varphi|_B$ is injective. Furthermore $\varphi : B \longrightarrow \mathbb{Z}^r$ is surjective since the e_i are in the image and they generate $\mathbb{Z}^r$. Thus $\varphi|_B$ is an isomorphism.

The conclusion that now follows from Proposition 3.1.13, but we will prove it directly by using Corollary 3.1.6. The argument above shows that $K \cap B = \{0\}$. We will show $K + B = A$, which gives $A \cong K \times B$.

Let $a \in A$ and let $\varphi(a) = \sum_{i=1}^{r} m_i e_i$. Let $b = \sum_{i=1}^{r} m_i b_i$ and consider $a - b$. It is easy to see that $\varphi(a - b) = 0$ so $a - b = k$ for some $k \in K$. Thus $a \in K + B$. Since a was arbitrary $A = K + B$.

□

Theorem 3.2.28. *Let A be a finitely generated abelian group. There is a unique integer $r \geq 0$ such that A has a subgroup B that is a free abelian group of rank r. For any such B, A is the internal direct product of* $\operatorname{Tor}(A)$ *and B. Furthermore* $\operatorname{Tor}(A)$ *is a finite abelian group, so it is the direct product of cyclic groups of prime power order.*

Proof. Proposition 3.2.21 shows that $A/\operatorname{Tor}(A)$ is torsion free. Proposition 3.2.26 shows that $A/\operatorname{Tor}(A)$ must then be free of some uniquely defined rank r. Proposition 3.2.27 shows that A has a subgroup B that maps isomorphically to $A/\operatorname{Tor}(A)$ and that $A = \operatorname{Tor}(A) \times B$.

Finally, we note that A is finitely generated and the quotient group A/B is isomorphic to $\operatorname{Tor}(A)$. Thus $\operatorname{Tor}(A)$ is finitely generated and a torsion group. Therefore it is finite and is classified by Theorem 3.2.17, $\operatorname{Tor}(A)$ is also finitely generated. □

Exercises 3.2.29. Infinitely generated abelian groups can be more complicated than finite ones. Consider the group $\mathbb{Q}/\mathbb{Z}$.

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

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

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

(d) Show that every element has finite order.

(e) Show that every nontrivial cyclic subgroup is generated by $\frac{1}{n}$ for some integer $n > 1$.

(f) Show that $\mathbb{Q}/\mathbb{Z}$ is not finitely generated as an abelian group.

(g) Show that $\mathbb{Q}/\mathbb{Z}$ cannot be written as a direct product of $\langle a \rangle$ and another group H for any nonzero $a \in \mathbb{Q}/\mathbb{Z}$.

3.3 Simple Groups and the Classification of Finite Groups

In the previous section we saw that finite abelian groups have a very simple structure; they are direct products of cyclic groups, each having order a power of a prime. In this section we take steps to understand the the classification of arbitrary finite groups.

The model for classification is unique factorization of integers: Every positive integer is the product of prime numbers in a unique way. Finite abelian groups have a somewhat more complicated factorization because the constituents of the unique factorization may involve $\mathbb{Z}_{p^r}$ for arbitrary r. So, $\mathbb{Z}_p \times \mathbb{Z}_p$ and $\mathbb{Z}_{p^2}$ are distinct even though they have the same number of elements.

The classification of finite groups is vastly more complicated than the classification of finite abelian groups. We would like to say that every finite group is "built" from a set of groups that are analogous to the prime numbers, indivisible themselves. The "building" process is much more complicated then simply forming a direct product. As we have seen, for example, S_3 is the semi-direct product $\mathbb{Z}_3 \rtimes \mathbb{Z}_2$ (with the action of $\mathbb{Z}_2$ on $\mathbb{Z}_3$ being $a \mapsto -a$), while $\mathbb{Z}_6$ is the direct product $\mathbb{Z}_3 \times \mathbb{Z}_2$. We will consider S_3 as built from the groups $\mathbb{Z}_2$ and $\mathbb{Z}_3$, just in a different way than $\mathbb{Z}_6$ is built, so S_3 and $\mathbb{Z}_6$ have the same constituent parts, but a different pasting together of the parts.

As another example consider the quaternion group Q, which is not even a semi-direct product. We will say that it is built from 3 copies of $\mathbb{Z}_2$ (as are all groups of order 8). The reasoning is this: $Q = \{\pm 1, \pm i, \pm j, \pm k\}$ has a normal subgroup generated by i with 4 elements. The quotient $Q/\langle i \rangle$ is isomorphic to $\mathbb{Z}_2$. That normal subgroup generated by i is isomorphic to $\mathbb{Z}_4$, and has a proper normal subgroup generated by i^2 that is isomorphic to $\mathbb{Z}_2$. The quotient $\langle i \rangle / \langle i^2 \rangle$ is isomorphic to $\mathbb{Z}_2$. We have what is called a composition series

$$\langle 1 \rangle \trianglelefteq \langle i^2 \rangle \trianglelefteq \langle i \rangle \trianglelefteq Q$$

Each subgroup is normal in the next in the sequence. Furthermore, the quotients in this case are $\mathbb{Z}_2$ at each step, and $\mathbb{Z}_2$ is a group that has no normal subgroup except for the group itself and the trivial group.

Definition 3.3.1. A group G is **simple** when the only normal subgroups of G are $\langle e_G \rangle$ and G.

A **composition series** for a group G is a sequence of subgroups $G_0 = \langle e_G \rangle$, G_1, G_2, $\ldots$, $G_n = G$ such that G_i is a normal subgroup of G_{i+1} and G_{i+1}/G_i is simple. The simple quotients are called **composition factors** of G. We will write

$$G = G_0 \trianglelefteq G_1 \trianglelefteq G_2 \trianglelefteq \ldots G_{n-1} \trianglelefteq G_n = G$$

The **length** of the composition series is n.

Proposition 3.3.2. *The only simple finite abelian groups are $\mathbb{Z}_p$ for p prime.*

Proof. We have already shown that every finite abelian group is isomorphic to the direct product of cyclic groups of prime power order. But $\mathbb{Z}_{p^r}$ is not simple. It is contructed from r copies of $\mathbb{Z}_p$. There is in fact a unique composition series

$$\langle 0 \rangle \trianglelefteq \langle p^{r-1} \rangle \trianglelefteq \langle p^{r-2} \rangle \trianglelefteq \ldots \langle p^2 \rangle \trianglelefteq \langle p \rangle \trianglelefteq \langle 1 \rangle = \mathbb{Z}_{p^r}$$

□

Here are the two big theorems on classifying finite groups. The first is essentially a uniqueness theorem about the composition factors of a group.

Theorem 3.3.3 (Jordan Holder). *Suppose G has two composition series*

$$\begin{aligned} \langle e_G \rangle = G_0 \trianglelefteq G_1 \trianglelefteq G_2 \trianglelefteq G_3 \ldots G_{n-1} \trianglelefteq G_n = G \\ \langle e_G \rangle = G'_0 \trianglelefteq G'_1 \trianglelefteq G'_2 \trianglelefteq G'_3 \ldots G'_{m-1} \trianglelefteq G'_m = G \end{aligned}$$

Then $m = n$ and the lists of simple groups from the two series, $G_1/G_0, \ldots, G_n/G_{n-1}$ and $G'_1/G'_0, \ldots, G'_m/G'_{m-1}$ are the same up to reordering.

The proof takes several steps and is a bit technical, so we refer to skip it here.

The second big theorem identifies all the simple groups. It was a massive project in the late 20th century. See the Wikipedia article.

Theorem 3.3.4 (Finite Simple Groups). *Every finite simple group is isomorphic to one of the following:*

(1) a cyclic group of prime order, $\mathbb{Z}_p$,

(2) an alternating group A_n, for $n \geq 5$,

(3) a group of Lie type,

(4) one of 27 "sporadic groups" (including the Tits group).

We have already shown that $\mathbb{Z}_p$ is simple. In the rest of this section we get halfway through the identification of simple groups ;-) by proving that A_n is simple for $n \geq 5$. We start with two simple lemmas, then prove A_5 is simple, then extend by induction to A_n for $n > 5$.

Lemma 3.3.5. *Let $n \geq 4$. If $N \trianglelefteq A_n$ and N contains a 3-cycle then $N = A_n$.*

Proof. Suppose for simplicity $(1,2,3) \in N$. Let $a \in \{4, \ldots, n\}$. Conjugate with $(1,2)(3,a)$ to get another element of N.

$$(1,2)(3,a)(1,2,3)(1,2)(3,a) = (2,1,a) \in N$$

Now for $b \neq 1,2,a$ conjugate with $(1,a)(2,b)$

$$(1,a)(2,b)(1,a,2)(1,a)(2,b) = (a,1,b) \in N$$

Finally, the same trick can be used to give an arbitrary $(a,b,c) \in N$. We know from Exercise 2.4.17 that the 3-cycles generate A_n, so $N = A_n$. □

Lemma 3.3.6. *Let $N \trianglelefteq A_n$ with $n \geq 5$. If N contains a product of two distinct transpositions then $N = A_n$.*

Proof. Let $\sigma = \tau_1\tau_2 \in N$. If τ_1 and τ_2 do not have disjoint support then their product is a 3-cycle, and we can apply the last lemma. Suppose they have disjoint support, $\sigma = (a, b)(c, d)$. Since $n \geq 5$ there is another element in $\{1, \dots, n\}$, call it x. Since N is normal, conjugating gives another element of N,

$$(a, b, x)\big((a, b)(c, d)\big)(a, x, b) = (x, a)(c, d) \in N.$$

Now take the product of the two elements of N that we have identified,

$$\Big((a, b)(c, d)\Big)\Big(x, a)(c, d)\Big) = (a, x, b) \in N$$

Since N has a 3-cycle, $N = A_n$ by the previous lemma. □

Proposition 3.3.7. *A_5 is simple.*

Proof. Let N be a non-trivial normal subgroup in A_5. Let σ be an element of N that is not the identity and consider its signature. The possibilities are $2, 2, 1$ or $3, 1, 1$ or 5. In the lemmas above, we have shown that if σ is a product of disjoint transpositions, or if it is a 3-cycle, then $N = A_n$. Suppose the 5-cycle $\sigma = (a, b, c, d, f)$ is in N and let $\delta = (a, b, c)$. Note that $\sigma^{-1}(\delta\sigma\delta^{-1}) \in N$ because the conjugation of σ by an element of A_n lands in N, and N is closed under multiplication. On the other hand,

$$\begin{aligned}(\sigma^{-1}\delta\sigma)\delta^{-1} &= (b, c, d)(c, b, a) \\ &= (a, d, b)\end{aligned}$$

This shows N contains a 3-cycle, so $N = A_n$. □

Theorem 3.3.8. *A_n is simple for $n \geq 5$.*

Proof. We proceed by induction, the case $n = 5$ has been established. Assume that A_i is simple for $i < n$; we'll prove that A_n is simple. We start with several observations about $G_i = \{\sigma \in A_n : \sigma(i) = i\}$. First, G_i is a subgroup and $G_i \cong A_{n-1}$. Second, the G_i are all conjugate subgroups in A_n since $G_i = (1, i, 2)G_1(1, 2, i)$ (check!). Finally, we show in the next paragraph that $A_n = \langle G_1, \dots, G_n\rangle$.

Any $\sigma \in A_n$ can be written as a product of an even number of transpositions. Since $n \geq 5$, the product of a pair of transpositions must fix some i, and is therefore in G_i. For example $(1, 2)(3, 4) \in G_5$. Pairing off consecutive terms in the factorization of σ we see that σ can be written as a product of elements in the groups G_i. Thus $A_n = \langle G_1, \dots, G_n\rangle$.

The strategy now is to show that if N is normal in A_n with $N \neq \{\text{id}\}$ then $N \cap G_i \neq \{\text{id}\}$ for some i. This in turn, by the following argument, implies that $N = A_n$. By the induction hypothesis, each G_i is simple. Since $N \cap G_i$ is a nontrivial normal subgroup of G_i, we have $N \cap G_i = G_i$. We noted above that the G_i are conjugate, so for any j there is some $\pi \in A_n$ such that $G_j = \pi G_i \pi^{-1}$. But then

$$\begin{aligned} G_j &= \pi G_i \pi^{-1} \\ &= \pi(N \cap G_i)\pi^{-1} \\ &= (\pi N \pi^{-1}) \cap (\pi G_i \pi^{-1}) \\ &= N \cap G_j \end{aligned}$$

Since $N \cap G_j = G_j$ for all j and the G_j generate A_n we have $N = A_n$.

Finally, to complete the proof, we will show that for a nontrivial $N \trianglelefteq A_n$ we must have $N \cap G_i \neq \{\text{id}\}$ for some i. Suppose that N contains an element σ whose cycle decomposition has a cycle of length at least 3; say $\sigma(a) = b$, $\sigma(b) = c$ with a, b, c distinct. Let d, f be different from a, b, c (we are using $n \geq 5$) and let $\tau = (a, d, f)$. Then $(\tau\sigma\tau^{-1})\sigma^{-1} \in N$ and straightforward computation shows $\tau\sigma\tau^{-1}\sigma^{-1}(c) = c$. This shows that $N \cap G_c \neq \{\text{id}\}$ and therefore, by the previous paragraph, that $N = A_n$.

If N contains no element whose cycle decomposition has a cycle of length at least 3, then all elements of N are products of disjoint transpositions. Suppose σ is such a nontrivial element of N. We may assume that σ doesn't fix anything, for we have already shown this would imply $A_n = G$. We are assuming $n \geq 6$, so σ has at least 3 transpositions $\sigma = (a, b)(c, d)(f, g) \cdots$. Conjugate by $\tau = (a, b)(c, f)$ and multiply by σ^{-1} and we have $(\tau\sigma\tau^{-1})\sigma^{-1} \in N$ and $\tau\sigma\tau^{-1}\sigma^{-1}(b) = b$. As above this implies that $N = A_n$.

Thus for $n \geq 5$, any normal subgroup of A_n is either trivial or A_n itself. □

Exercises 3.3.9. This problem fleshes out some details in the proof of Theorem 3.3.8: A_n is simple for $n > 5$.

- **(a)** Within the alternating group A_n for each $i = 1, \ldots, n$, let $G_i = \{\sigma \in A_n : \sigma(i) = i\}$. Show that G_i is a subgroup of A_n.
- **(b)** Find a $\pi \in A_n$ such that each $G_i = \pi G_j \pi^{-1}$.
- **(c)** Justify the statement in the fourth paragraph of the proof of 3.3.8 "Then $(\tau\sigma\tau^{-1})\sigma^{-1} \in N$ and straightforward computation shows $\tau\sigma\tau^{-1}\sigma^{-1}(c) = c$."
- **(d)** Justify the statement in the fifth paragraph of the proof of 3.3.8 "$(\tau\sigma\tau^{-1})\sigma^{-1} \in N$ and $\tau\sigma\tau^{-1}\sigma^{-1}(b) = b$."

3.4 Free groups, generators and relations

Chapter 4

Rings

Of the three main topics in this book—groups, rings, and fields—the study of rings will probably be the most familiar to the reader. The first section of Chapter 1 concerned the protypical ring, the integers. The main properties examined—quotients and remainders, divisibility, prime numbers, factorization—are familiar from grade school. At the end of that chapter, these properties are extended to the polynomial ring $F[x]$ for F a field. Polynomials over the rational numbers, $\mathbb{Q}$, are a major portion of the secondary school curriculum, although the focus is on the geometry of solutions to polynomial equations more than algebraic properties. Still, the division of polynomials and factorization do play an important role in the curriculum.

In this chapter, we deal with general rings, with the caveat that multiplication is assumed to be commutative. The special roles of the integers and of polynomial rings will be evident, and the properties of divisibility and factorization are major themes. The two GCD Theorems 1.1.4 and 1.3.3 involve a construction that also plays a major role: the GCD of two integers is a linear combination of the two and the GCD of two polynomials is a polynomial combination of the two. Ideals are subsets of a ring that generalize the construction of linear/polynomial combinations of two elements. Ideals are used to define quotient rings in Section 4.4. That following section presents the relatively straightforward generalizations of the isomorphism theorems for groups 2.7.9 2.9.1 3.1.3 and of the related Factor Theorem 2.7.11 and Correspondence Theorem 2.9.2. Key properties of ideals and the relationship to properties of the quotient rings are studied in Section 4.6. The culminating topic of this Chapter is rings of fractions 4.7, which extends the techniques used to derive the rational numbers from the integers.

4.1 Rings

In this section we recall the definition of a ring and introduce three particular types of rings: fields, integral domains, reduced rings. The integers and $F[x]$ provide ways to construct examples and non-examples of these three types of rings using the modular arithmetic introduced in Sections 1.1 and 1.3.

Definition 4.1.1. A **ring** is a set R, with two operations $+$ and $*$ that satisfy the following properties.

(1) $+$ and $*$ are both associative. That is $(a+b)+c = a+(b+c)$ and $a*(b*c) = (a*b)*c$.

(2) $+$ and $*$ are both commutative. That is $a+b = b+a$ and $a*b = b*a$.

(3) $+$ and $*$ both have identity elements. 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 $a*1 = a$.

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

(5) $*$ distributes over $+$. That is $a*(b+c) = a*b+a*c$.

A few comments are in order. Strictly speaking, the definition above is for a **commutative ring with identity**. The modifier "commutative" is referring to commutativity of multiplication, and "with identity" is referring to the multiplicative identity. There is a rich study of rings where multiplication is non-commutative (for example the ring of $n \times n$ matrices over $\mathbb{R}$, see Section 4.3), but treating the subject would spread our efforts too thinly. It is common when treating only commutative rings to simplify the terminology at the outset as we do here: A ring for us is assumed to be commutative and have an identity.

One may also say that a ring R is a commutative group under $+$ and that $R^* = R \setminus \{0\}$ is a commutative *monoid* (look it up!) under $*$, with the additional property that $*$ distributes over $+$.

One can show by induction that a sum of several terms (or a product of several terms) may be computed in any order and that $r*(a_1+a_2+\cdots+a_n) = r*a_1 + r*a_2+\cdots+r*a_n$.

There are a bunch of little results one should verify. Since any ring R is a group under $+$ we already know that the additive identity 0 is unique and that the additive inverse of $r \in R$ is unique (we write it as $-r$). We know that $-(-r) = r$ and that for $s \in R$, $-(r+s) = (-r)+(-s)$ (using commutativity of $+$). Additionally we have the following properties. Proofs are left as an exercise.

Proposition 4.1.2. *Let R be a ring and $r, s \in R$*

(1) For an integer m, we can make sense of mr via

$$\Big(\underbrace{1 + 1 + \cdots + 1}_{m\ terms}\Big) * r = \underbrace{r + r + \cdots + r}_{m\ terms}$$

(2) For any $r \in R$, $r * 0 = 0$.

(3) The multiplicative identity element 1 is unique.

(4) The additive inverse and multiplication operate as expected.

- $r * (-s) = -(r * s)$
- $(-r) * (-s) = r * s$

We haven't excluded the possibility that $1 = 0$. In this case for any $r \in R$, $r = r * 1 = r * 0 = 0$. Thus we have a unique situation, a ring that has just one element. We call it the **trivial ring**.

We will write the product without the multiplication symbol when there is no concern about ambiguity, that is rs instead of $r * s$. But for clarity and emphasis on the basic properties of a ring, we will continue to explicitly show the product symbol in this section.

There are three special types of elements in a ring, and, based on their existence or not, three special types of rings.

Definition 4.1.3. An element u of a ring R is a **unit** when there is another element v such that $u * v = 1$. An element a of a ring R is a **zero-divisor** when $a \neq 0$ and there is some $b \neq 0$ in R such that $a * b = 0$. An element a of a ring R is **nilpotent** when there exists some positive integer n such that $a^n = 0$.

Definition 4.1.4. A **field** is a nontrivial ring in which every nonzero element is a unit. An **integral domain** is a nontrivial ring that has no zero-divisors. A nontrivial ring is **reduced** if it has no nilpotent elements other than 0.

Exercises 4.1.5. Prove the results in Proposition 4.1.2, and in addition prove the following.

(a) The inverse of a unit is unique.

(b) The inverse of a unit is also a unit.

(c) A unit cannot be a zero divisor.

(d) A nilpotent element is either 0 or a zero-divisor.

Exercises 4.1.6. Cancellation in integral domains.

(a) Let R be an integral domain. Show that the cancellation law holds: If $ar = as$ then $r = s$.

(b) Let R be an integral domain that is finite. Show that R is a field. (For a nonzero $a \in R$, consider the function $R \longrightarrow R$ that takes r to $a * r$. Use the cancellation law to show injectivity.)

The Integers and $F[x]$ for F a Field

We can now fully appreciate the integers $\mathbb{Z}$, having ignored multiplication when we studied the integers as a group. The integers, the number system we learn in elementary school, form the first example of a ring. One of the key properties of the integers (used in solving a quadratic equation!) is that $ab = 0$ implies $a = 0$ or $b = 0$. In terms defined above, $\mathbb{Z}$ has no zero-divisors, so it is an integral domain.

Let us now turn to modular arithmetic, which we introduced in Section 1.1. We will use $[a]_n$ for the equivalence class of a modulo n and we will omit the subscript n when the modulus is obvious. The following expands on Exercise 1.3.10.

Theorem 4.1.7 (Units, Zero Divisors in $\mathbb{Z}/n$)**.** *Let $n \geq 2$ be an integer and let a be an integer.*

(1) $[a]_n$ *is a unit iff* $\gcd(a, n) = 1$.

(2) $[a]_n$ *is a zero divisor iff* $1 < \gcd(a, n) < n$.

In particular, an element of $\mathbb{Z}/n$ is either 0, or a unit, or a zero-divisor, and these are mutually exclusive.

Proof. Let $d = \gcd(a, n)$. There are three mutually exclusive cases, $d = 1$, $d = n$ and $1 < d < n$.

If $d = 1$ then, by the GCD Theorem 1.1.4 there are integers u, v such that $ua + vn = 1$. Reducing modulo n we have

$$\begin{aligned} [u]_n * [a]_n + [v]_n * [n]_n &= [1]_n \\ [u]_n * [a]_n &= [1]_n \end{aligned}$$

so $[u]_n$ is the multiplicative inverse of $[a]_n$ and $[a]_n$ is a unit in $\mathbb{Z}/n$.

If $d = n$, then $[a]_n = [0]_n$.

If $1 < d < n$ then $a = db$ and $n = dc$ for some integers b and c with $c < n$. Then $[c]_n * [a]_n = [c]_n * [db]_n = [cd]_n * [b]_b = [n]_n * [b]_n = [0]_n$. But $[a]_n \neq [0]_n$ and, since $c < n$, $[c]_n \neq [0]_n$. Thus $[a]_n$ is a zero divisor. □

We are particularly interested in the following special case, to which we will return in depth later.

Corollary 4.1.8. *Let p be a prime number. Then $\mathbb{Z}/p$ is a field; every nonzero element has an inverse.*

When working with the integers modulo a prime we usually use the notation $\mathbb{F}_p$ instead of $\mathbb{Z}/p$ to emphasize that we have a field.

A general theme emphasized in these notes is the similarity between the integers and a polynomial ring over a field $F[x]$. We have similar results when working modulo a polynomial as we did for modular arithmetic.

Theorem 4.1.9 (Units, Zero Divisors in $F[x]/m(x)$)**.** *Let F be a field and let $m(x)$ be a polyomial of degree $\delta > 0$. Let $a(x) \in F(x)$ and let $[a(x)]$ be its congruence class modulo $m(x)$.*

(1) $[a(x)]$ *is a unit iff* $\gcd(a(x), m(x)) = 1$.

(2) $[a(x)]$ *is a zero divisor iff* $\gcd(a(x), m(x))$ *has degree greater than 0 and less than d.*

In particular, an element of $F[x]/m(x)$ is either 0, or a unit, or a zero-divisor, and these are mutually exclusive.

Proof. Let $d(x) = \gcd(a(x), m(x))$. There are three mutually exclusive cases, $d(x) = 1$, $d(x) = m(x)$ and $0 < \deg(d(x)) < \deg(m(x))$. These lead to $[a(x)]$ a unit, $[a(x)] = [0]$ and $[a(x)]$ a zero divisor, respectively. The proof for each case is entirely similar to that for the integers. □

As with the integers, the following case is of special interest and we will return to it in depth later.

Corollary 4.1.10. *Let $m(x)$ be an irreducible polynomial in $F[x]$ for F a field. Then every nonzero element in $F[x]/m(x)$ has an inverse, so $F[x]/m(x)$ is a field. Conversely, if $m(x)$ is reducible, $F[x]/m(x)$ is not even an integral domain.*

Exercises 4.1.11.

(a) Find the nilpotent elements of $\mathbb{Z}/8$, of $\mathbb{Z}/12$, and of $\mathbb{Z}/30$.

(b) Under what conditions on n does $\mathbb{Z}/n$ have nonzero nilpotent elements?

(c) Identify the nilpotent elements of $\mathbb{Z}/n$ using the unique factorization of n.

(d) Let F be a field and $m(x) \in F[x]$. Under what conditions on $m(x)$ does $F[x]/m(x)$ have nonzero nilpotent elements? Identify the nilpotents $a(x) \in F[x]/m(x)$ using unique factorization into irreducibles of $m(x)$ and $a(x)$.

4.2 Ring Homomorphisms

As with groups, the functions that preserve structure on rings are of primary interest. We first treat subrings, then introduce homomorphisms.

Definition 4.2.1. Let R be a ring. A subset $T \subseteq R$ is a **subring** of R when T is an additive subgroup of R, T is closed under multiplication, and T contains 1_R.

Example 4.2.2. The ring of integers $\mathbb{Z}$ is a subring of $\mathbb{Q}$. The rings $\mathbb{Z}/n$ have no proper subrings because of the requirement that the unitary element, 1, be contained in a subring. Adding 1 to itself will give all of $\mathbb{Z}_n$.

The polynomial ring $F[x]$ for F a field has many subrings: F itself, any subring of F, and subrings generated by a polynomial. For example, $F[x^2]$ would contain all polynomials in which each term has even degree. One can check that it is indeed a subring of $F[x]$.

Exercises 4.2.3. Suppose R, S are subrings of a ring T.

(a) Show that $R \cap S$ is also a subring of T.

(b) If R and S are integral domains, show that $R \cap S$ is also an integral domain.

(c) If R and S are fields, show that $R \cap S$ is also a field.

(d) More generally, show that for any subset A of T, that the intersection of all rings containing A is a subring of T.

(e) Similarly, show that for any subset A of T, that the intersection of all fields containing A is a subring of T.

(f) Give an example to show that $R \cup S$ may not be a ring.

Definition 4.2.4. Let R, S be rings. A funtion $\varphi : R \longrightarrow S$ is a *ring homomorphism* when

(1) φ is a homomorphism of the additive groups $R, +_R$ and $S, +_S$, and

(2) $\varphi(1_R) = 1_S$, and

(3) for $r_1, r_2 \in R$,

$$\varphi(r_1 *_R r_2) = \varphi(r_1) *_S \varphi(r_2)$$

Notice that the operation on the left-hand side is in R and the operation $*$ on the right-hand side is in S. When we want to be careful we specify the ring for the operation as we did here, but generally this is left to the reader to infer. We say φ *respects addition, multiplication and the identity element.* That is, it *respects* the ring structure. Recall from group theory that it is sufficient to check that a function φ respects the group operation to ensure that φ is a homomorphism. Thus to check if $\varphi : R \longrightarrow S$ is a ring homomorphism one verifies that $\varphi(1_R) = 1_S$, $\varphi(r_1 +_R r_2) = \varphi(r_1) +_S \varphi(r_2)$, and $\varphi(r_1 *_R r_2) = \varphi(r_1) *_S \varphi(r_2)$.

If S is a subring of R then there is a homomorphism, the **inclusion homomorphism**, from S to R.

Proposition 4.2.5. *Let R, S, T be rings. If $\varphi : R \longrightarrow S$ and $\theta : S \longrightarrow T$ are ring homomorphisms then the composition $\theta \circ \varphi$ is also a ring homomorphism.*

The proof is left as an exercise.

Definition 4.2.6. The **kernel** of a ring homomorphism $\varphi : R \longrightarrow S$ is the preimage of 0_S; that is $\{r \in R : \varphi(r) = 0_S\}$. A homomorphism that is injective is called an **embedding**. A homomorphism that is a bijection (injective and surjective) is called an **isomorphism**.

From group theory we know that the kernel is a normal subgroup (since addition in rings is commutative, any subgroup is normal). The kernel has an additional important property, which is item (3) in the following theorem.

Theorem 4.2.7. *Let $\varphi : R \longrightarrow S$ be a homomorphism of rings and let K be the kernel.*

(1) The image of R is a subring of S.

(2) φ is injective if and only if $K = \{0_R\}$.

(3) For any $r \in R$ and any $k \in K$, $r *_R k \in K$.

Proof. We know that $\varphi(R)$ is a subgroup of S. From the requirement that $\varphi(1_R) = 1_S$ we have $1_S \in \varphi(R)$. To show that $\varphi(R)$ is closed under multiplication, let $\varphi(r_1)$ and $\varphi(r_2)$ be arbitrary elements of $\varphi(R)$. Then $\varphi(r_1) *_S \varphi(r_2) = \varphi(r_1 *_R r_2)$, and this is in $\varphi(R)$.

For the properties of the kernel, note first that if φ is injective there can only be one element that maps to 0_S, and that is 0_R. Conversely, suppose $K = \{0_R\}$, and suppose $\varphi(r) = \varphi(r')$. Then $\varphi(r - r') = \varphi(r) - \varphi(r') = 0_S$. Since the kernel is trivial, $r - r' = 0_R$ so $r = r'$. This establishes injectivity.

Let $k \in K$ and $r \in R$. We have

$$\varphi(r *_R k) = \varphi(r) *_S \varphi(k) = \varphi(r) *_S 0_S = 0_S$$

Thus $r *_R k \in K$.

□

As with isomorphisms of groups, the isomorphisms between rings set up an equivalence relation. Every ring is clearly isomorphic to itself via the identity map. The following theorem establishes symmetry and transitivity.

Theorem 4.2.8. *If $\varphi : R \longrightarrow S$ is an isomorphism then φ^{-1} is also an isomorphism. The composition of two isomorphisms is an isomorphism.*

Proof. Let $\varphi : R \longrightarrow S$ be an isomorphism of rings. Then φ is a bijection, so φ^{-1} is also a bijection from S to R. We must show it is a homomorphism. Since φ is a homomorphism, $\varphi(1_R) = 1_S$ and therefore $\varphi^{-1}(1_S) = 1_R$ (since φ is injective, this is the sole preimage).

To show φ^{-1} respects addition and multiplication, let s, s' be arbitrary elements of S. Since φ is a bijection, there are unique $r, r' \in R$ such that $\varphi(r) = s$ and $\varphi(r') = s'$.

$$\begin{aligned}\varphi^{-1}(s *_S s') &= \varphi^{-1}\big(\varphi(r) *_S \varphi(r')\big)\\ &= \varphi^{-1}\big(\varphi(r *_R r')\big)\\ &= r *_R r'\\ &= \varphi^{-1}(s) *_R \varphi^{-1}(s')\end{aligned}$$

This shows φ^{-1} respects multiplication. A completely analogous proof is used for addition. □

Definition 4.2.9. Two rings R, S are **isomorphic** if there is an isomorphism from R to S (and therefore, by the theorem, also an isomorphism from S to R).

Exercises 4.2.10. Let $\varphi : R \longrightarrow S$ be a ring homomorphism.

(a) Show that for any subring R' in R, the image $\varphi(R')$ is a subring of S.

(b) Show that any subring S' of S, the preimage $\varphi^{-1}(S')$ is a subring of R.

The Integers and $F[x]$ for F a Field

The following theorem shows that the Integers are the "original (or initial) ring."

Theorem 4.2.11 (The Initial Ring)**.** *For any ring R there is a unique homomorphism from $\mathbb{Z}$ to R. The kernel is the set of multiples of some integer m. If $m = 0$ then R has a subring isomorphic to $\mathbb{Z}$. If $m > 0$, there is an isomorphism of $\mathbb{Z}/m$ with a subring of R.*

Proof. A homomorphism $\varphi : \mathbb{Z} \longrightarrow R$, if it exists, would have to take $1_{\mathbb{Z}}$ to 1_R. Applying the requirement that a homomorphism respects addition we see inductively that we must have, for $m > 0$,

$$\varphi(m) = \varphi(\underbrace{1_{\mathbb{Z}} + \cdots + 1_{\mathbb{Z}}}_{m \text{ terms}}) = \underbrace{1_R + \cdots + 1_R}_{m \text{ terms}} = m1_R$$

We also must have $\varphi(-m) = -\varphi(m)$. Thus, there is at most one way to define a homomorphism from $\mathbb{Z}$ to R. This function respects addition:

$$\begin{aligned}
\varphi(m+n) &= (m+n)1_R \\
&= \underbrace{1_R + 1_R + \cdots + 1_R}_{m+n \text{ terms}} \\
&= \underbrace{1_R + \cdots + 1_R}_{m \text{ terms}} + \underbrace{1_R + \cdots + 1_R}_{n \text{ terms}} \\
&= m1_R + n1_R \\
&= \varphi(m) + \varphi(n)
\end{aligned}$$

It also respects multiplication:

$$\begin{aligned}
\varphi(mn) &= mn1_R \\
&= \varphi\big(\underbrace{1_R + 1_R + \cdots + 1_R}_{mn \text{ terms}}\big) \\
&= \big(\underbrace{1_R + \cdots + 1_R}_{m \text{ terms}}\big) *_R \big(\underbrace{1_R + \cdots + 1_R}_{n \text{ terms}}\big) \\
&= (m1_R) * (n1_R) \\
&= \varphi(m) *_R \varphi(n)
\end{aligned}$$

If m generates the kernel the first isomorphism theorem for groups says that $\mathbb{Z}/m$ is isomorphic to a subgroup of R. Since multiplication is just repeated addition, this is also an isomorphism of rings.

□

The integer m in the theorem is called the **characteristic** of R.

The Arithmetic Modulo n Theorem 1.1.16 shows that the function taking $a \in \mathbb{Z}$ to $[a]_n \in \mathbb{Z}/n$ is a homomorphism. We actually define addition and multiplication in $\mathbb{Z}/n$ via addition and multiplication in $\mathbb{Z}$, so it is an immediate consequence that the map is a homomorphism.

If d divides m there is a well-defined function from $[a]_m$ to $[a]_d$— this is because $d|m$ and $m|(b-a)$ implies $d|(b-a)$—so any two integers that are congruent modulo m are also congruent modulo d. On the other hand if $d \nmid m$, there is no well-defined homomorphism from $\mathbb{Z}/m$ to $\mathbb{Z}/d$: we would have to take $[1]_m$ to $[1]_d$ but

$m[1]_m = [0]_m$ and $m[1]_d \neq [0]_d$. Since the arithmetic on $\mathbb{Z}/m$ is "inherited" from $\mathbb{Z}$, we have:

Theorem 4.2.12 (Homomorphism mod m)**.** *The function $\mathbb{Z} \longrightarrow \mathbb{Z}/n$ taking a to $[a]_n$ is a homomorphism.*

There is a homomorphism from $\mathbb{Z}/m$ to $\mathbb{Z}/d$ if and only if d divides m. The homomomorphism is unique (since it takes $[1]_m$ to $[1]_d$).

We have a similar result for polynomial rings.

Theorem 4.2.13 (Homomorphism mod $m(x)$)**.** *Let F be a field and $m(x) \in F[x]$ The function $F[x] \longrightarrow F[x]/m(x)$ taking $a(x)$ to its equivalence class $[a(x)]$ is a homomorphism.*

There is a homomorphism from $F[x]/m(x)$ to $F[x]/r(x)$ if and only if $r(x)$ divides $m(x)$.

4.3 Constructions

In this section we introduce three ways to construct new rings: using direct products, using an indeterminate to create a polynomial ring, and using matrices to create a noncommutative ring.

Direct Products

In Section 2.3 we showed that the Cartesian product of groups has the structure of a group. Not surprisingly, we have the same situation with rings, but there is one subtle difference, discussed below.

Definition 4.3.1. Let R and S be rings. The Cartesian product $R \times S$, along with the operations below form the **direct product** of R and S.

$$\begin{aligned} -(r, s) &= (-r, -s) \\ (r_1, s_1) +_{R\times S} (r_2, s_2) &= (r_1 +_R r_2, s_1 +_S s_2) \\ (r_1, s_1) *_{R\times S} (r_2, s_2) &= (r_1 *_R r_2, s_1 *_S s_2) \end{aligned}$$

The additive identity and multiplicative identies are of course $(0_R, 0_S)$ and $(1_R, 1_S)$. The following proposition shows that the direct product of rings is in fact a ring and gives other important properties.

Proposition 4.3.2 (Direct Product)**.** *Let R and S be rings.*

(1) The above definition does, indeed, make $R \times S$ a ring.

(2) The associative law for products of several rings holds: $R_1 \times (R_2 \times R_3) \cong (R_1 \times R_2) \times R_3$.

(3) If R' is a subring of R and S' is a subring of S then $R' \times S'$ is a subring of $R \times S$.

(4) The projection maps $p_R : R \times S \longrightarrow R$ and $p_S : R \times S \longrightarrow S$ are surjective homomorphisms.

(5) The construction and the observations above can be generalized to the direct product of any set of rings $\{R_i : i \in I\}$ indexed by a finite set I. (It extends with some modification due to subtle issues when I is infinite.)

The subtle difference between the product of groups and the product of rings is that there does not exist a natural homomorphism $R \longrightarrow R \times S$. The choice that one might expect would be to send r to $(r, 0)$ but this violates the requirement that the multiplicative identity on R should map to the multiplicative identity on $R \times S$.

Exercises 4.3.3. Let R and S be rings and consider $R \times S$.

(a) Identify all the units in $R \times S$.

(b) Identify all of the zero-divisors in $R \times S$.

The following property of the direct product of rings is analogous to the one for groups, Proposition 2.3.6. The proof is easily adapted from the proof for groups.

Proposition 4.3.4 (Universal Property of the Product)**.** *Let R, S and T be rings, and let $\varphi : T \longrightarrow R$ and $\psi : T \longrightarrow S$ be homomorphisms. The function $\alpha : T \longrightarrow R \times S$ defined by $t \longmapsto (\varphi(t), \psi(t))$ is a homomorphism. It is the unique homomorphism such that $p_R \circ \alpha = \varphi$ and $p_S \circ \alpha = \psi$.*

Exercises 4.3.5. An element e in a ring R is **idempotent** when $a^2 = a$. Evidently, both 0_R and 1_R are idempotents. If R and S are rings, then $R \times S$ has two additional idempotents $(1, 0)$ and $(0, 1)$.

(a) let R be a ring with an idempotent e.

(1) Prove that the set $Re = \{re : r \in R\}$ with the operations inherited from R has the structure of a ring, with identity e. It is not a subring of R because the multiplicative identity element is different.

(2) Prove that $(1 - e)$ is also an idempotent in R.

(3) Prove that every element in R may be uniquely expressed as the sum of an element in Re and an element in $R(1 - e)$.

(b) Find the idempotents in $\mathbb{Z}/12$ and comment on the decomposition above.

(c) Find the idempotents in $\mathbb{Z}/30$ and comment on the decomposition above.

Polynomial Rings

We have already discussed polynomial rings that have coefficients in a field, such as $\mathbb{Q}[x]$, $\mathbb{F}_p[x]$. The construction generalizes to any ring. For R a ring, the polynomial ring $R[x]$ is the set of elements of the form

$$a_0 + a_1x + a_2x^2 + \cdots + a_\delta x^\delta$$

with the $a_i \in R$. The sum of two elements and product of two elements are familiar formulas. For example,

$$\begin{aligned}
&(a_0 + a_1x + a_2x^2 + a_3x^3) + (b_0 + b_1x + b_2x^2 + b_3x^3)\\
&\qquad = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + (a_3 + b_3)x^3\\
&(a_0 + a_1x + a_2x^2)(b_0 + b_1x + b_2x^2)\\
&\qquad = a_0b_0 + (a_1b_0 + a_0b_1)x + (a_2b_0 + a_1b_1 + a_0b_2)x^2
\end{aligned}$$

Some care is in order to stipulate that, for example, $a_0 + a_1x = a_0 + a_1x + 0x^2$. I'll introduce the following formal definition, but your instincts should be your guide.

Definition 4.3.6. Let R be a ring. A **polynomial** in x over R is a sum $a(x) = \sum_{i=0}^{\infty} a_ix^i$ in which only a finite number of the a_i are nonzero. The **support** of $a(x)$ is the set of powers of x for the nonzero terms $\{x^i : a_i \neq 0\}$, or depending on the context, the indices of those terms, $\{i : a_i \neq 0\}$. For $a(x)$ nonzero, the **degree** is the maximal index with a nonzero term, $\deg(a(x)) = \max_{i\in\mathbb{N}_{\geq 0}} \{i : a_i \neq 0\}$. We set the degree of the 0 polynomial to be $-\infty$.

The **polynomial ring over R with indeterminate** x is the set of all polynomials.

$$\left\{ \sum_{i=0}^{\infty} a_ix^i : a_i \in R, \text{ and } \{i : a_i \neq 0\} \text{ is finite} \right\}$$

The additive inverse of $\sum_{i=0}^{\infty} a_ix^i$ is $\sum_{i=0}^{\infty}(-a_i)x^i$. The sum is defined by

$$\sum_{i=0}^{\infty} a_ix^i + \sum_{i=0}^{\infty} b_ix^i = \sum_{i=0}^{\infty}(a_i + b_i)x^i$$

and the product is defined by $(ax^i) * (bx^j) = abx^{i+j}$ and applying distributivity, commutativity and distributivity. Consequently,

$$(\sum_{i=0}^{\infty} a_ix^i)(\sum_{j=0}^{\infty} b_jx^j) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_ib_jx^{i+j}$$

and we can rearrange the sum using $k = i + j$ and gathering terms in x^k,

$$= \sum_{k=0}^{\infty} x^k \sum_{i=0}^{k} (a_i b_{k-i})$$

As an aside, if we don't require the support to be finite we still get a valid ring, which is called the ring of formal power series.

One should check that all the properties of a ring hold, a somewhat tedious exercise. As an example, for associativity one can show that

$$\Big(a(x)b(x)\Big)c(x) = \sum_m \sum_{\substack{i,j,k \geq 0 \\ i+j+k=m}} a_i b_j c_k = a(x)\Big(b(x)c(x)\Big)$$

Proposition 4.3.7. *For any ring R and $a(x), b(x) \in R$,*

$$\deg\big(a(x)b(x)\big) \leq \deg\big(a(x)\big) + \deg\big(b(x)\big)$$

If R is an integral domain then equality holds, the degree of a product of polynomials is the sum of the degrees of the factors:

$$\deg\big(a(x)b(x)\big) = \deg\big(a(x)\big) + \deg\big(b(x)\big)$$

In particular, if R is an integral domain, then $R[x]$ is also an integral domain.

Proof. Let $\gamma = \deg(a(x))$ and $\delta = \deg(b(x))$. From the formula for the product the degree k term in $a(x)b(x)$ is $\left(\sum_{i=0}^{k} a_i b_{k-i}\right)$. When $k > \gamma + \delta$, the kth term will be 0 because for $i > \gamma$, $a_i = 0$, and for $i \leq \gamma$, $k - i > \delta$ so $b_{k-i} = 0$. For $k = \gamma + \delta$ the kth term in the product is $a_\gamma b_\delta$. This may be 0 in a ring with zero-divisors, hence the degree of a product may be less than the sum of the degrees of the factors over a ring R with zero-divisors.

In an integral domain, $a_\gamma b_\delta \neq 0$ since we assume a_γ and b_δ are nonzero. In particular, a product of nonzero polynomials over an integral domain cannot be zero. This proves the proposition. □

The key lemma (1.3.1) that was used to prove the Quotient Remainder Theorem (1.3.2) in $F[x]$ does not apply over an arbitrary ring.

Exercises 4.3.8. Polynomial rings and zero divisors.

(a) Give an example to show the analogue of Lemma 1.3.1 does not hold in the polynomial ring over $\mathbb{Z}/4$.

(b) Show that $\mathbb{Z}/4[x]$ has a polynomial of degree 1 that is a unit (in fact its square is 1).

(c) Give an example of two polynomials of degree 2 in $\mathbb{Z}/8[x]$ such that their product has degree 1. Show that this is not possible in $\mathbb{Z}/4[x]$ and $\mathbb{Z}/6[x]$.

Here is an additional result about polynomial rings that will be useful. It is similar to The Initial Ring Theorem 4.2.11 for the integers.

Theorem 4.3.9 (Universal property of polynomial rings)**.** *Let R, S be rings and let $\varphi : R \longrightarrow S$ be a ring homomorphism. For any $s \in S$ there is a unique homomorphism from $R[x]$ to S that agrees with φ on R and takes x to s, namely*

$$\begin{aligned} \overline{\varphi} : R[x] &\longrightarrow S \\ \big(\sum_i r_i x^i\big) &\longmapsto \sum_i \varphi(r_i) s^i \end{aligned}$$

Proof. To simplify the notation the summations in the text below are implicitly over the nonnegative integers unless expressed otherwise. If there is a homomorphism $\overline{\varphi}$ taking x to s and agreeing with φ on R then we must have

$$\overline{\varphi}\Big(\sum_i r_i x^i\Big) = \sum_i \overline{\varphi}(r_i x^i) = \sum_i \overline{\varphi}(r_i)\overline{\varphi}(x)^i = \sum_i \varphi(r_i) s^i$$

So there is only one possible way to define $\overline{\varphi}$. The key observation is that $\overline{\varphi}$ is well defined because there is a unique way to write each element of $R[x]$ and we have used this unique formulation to define $\overline{\varphi}$. We also note that the sums are all finite sums.

To show this function is indeed a homomorphism we check that it respects the operations. Here we check just products. As we saw above commutativity, associativity and distributivity in the polynomial ring give

$$(\sum_i a_i x^i)(\sum_j r_j x^j) = \sum_k x^k \sum_{i=0}^{k} (a_i r_{k-i})$$

A similar derivation shows that for $b_i, t_i \in S$

$$(\sum_i b_i s^i)(\sum_i t_i s^i) = \sum_k s^k \sum_{i=0}^{k} (b_i t_{k-i})$$

Thus we have

$$\begin{aligned}
\overline{\varphi}\Big(\Big(\sum_i a_i x^i\Big)\Big(\sum_i b_i x^i\Big)\Big) &= \overline{\varphi}\Big(\sum_k x^k\Big(\sum_{i=0}^{k} a_i b_{k-i}\Big)\Big) \\
&= \sum_k s^k\Big(\sum_{i=0}^{k} \varphi(a_i b_{k-i})\Big) \\
&= \Big(\sum_i \varphi(a_i)s^i\Big)\Big(\sum_j \varphi(b_j)s^j\Big) \\
&= \overline{\varphi}\Big(\sum_i a_i x^i\Big)\overline{\varphi}\Big(\sum_i b_i x^i\Big)
\end{aligned}$$

This shows $\overline{\varphi}$ respects products. □

In the first encounter a student has with polynomials they are treated as functions. For a polynomial over $\mathbb{Q}$, one substitutes a rational number for x and computes a rational number as output. Up to this point we have treated polynomials algebraically by adding and multiplying. The previous theorem has applications in which we treat polynomials as functions.

Exercises 4.3.10.

(a) Consider $R = S = \mathbb{Q}$ and $s = 0$. Explain how to apply the theorem to show that evaluating all polynomials in $\mathbb{Q}[x]$ at 0 yields a homomorphism from $\mathbb{Q}[x]$ to $\mathbb{Q}$.

(b) More generally, apply the theorem to the situation where $S = R$ and s is some particular element of R. Interpret the theorem for this situation as saying that "evaluating at a fixed $s \in R$" determines a homomorphism.

(c) Let's apply the theorem to the situation where $S = R[x]$ and the particular element $s(x) \in R[x]$. The theorem says that there is a homomorphism

$$\begin{aligned}
\varphi : R[x] &\longrightarrow R[x] \\
x &\longrightarrow s(x)
\end{aligned}$$

Show that an arbitrary $f(x) \in R[x]$ maps to $f(s(x))$. Show that, as a function this is the composition of the function defined by $f(x)$ and the function defined by $s(x)$.

By iteratively applying the polynomial ring construction we can create a polynomial ring in several indeterminates over a ring R.

Definition 4.3.11. The polynomial ring over the ring R in indeterminates $x_1, x_2, \ldots, x_n$, which we write as $R[x_1, x_2, \ldots, x_{n-1}, x_n]$, is defined inductively as $R[x_1, \ldots, x_{n-1}][x_n]$.

An element $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$ is called a **monomial**. We will say that $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is the **multidegree** of this monomial. The elements of $R[x_1, x_2, \ldots, x_{n-1}, x_n]$ are finite sums of **terms** of the form $ax_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$.

This has the following important property.

Theorem 4.3.12 (Universal property of polynomial rings)**.** *Let R, S be rings and let $\varphi : R \longrightarrow S$ be a ring homomorphism. For any $s_1, \ldots, s_n \in S$ there is a unique homomorphism from $R[x_1, x_2, \ldots, x_n]$ to S that agrees with φ on R and takes x_i to s_i.*

Matrix Rings

Although we are focusing on rings that are commutative, and moreover, have defined a ring to be commutative, it is worth presenting one very important example of a noncommutative ring.

Definition 4.3.13. Let R be a ring and let n be a positive integer. The $n \times n$ **matrix ring** over R, written $M_n(R)$, is the set of $n \times n$ matrices using the usual formulas for addition and multiplication.

Even for a field, the matrix ring is noncommutative, as you may recall from your experience with linear algebra.

Exercises 4.3.14. Consider $M_2(\mathbb{Q})$ and recall the various spaces associated to a matrix (e.g. rowspace, nullspace).

(a) Show that $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ is a zero divisor in $M_2(\mathbb{Q})$. Find a B such that $AB = 0$ (the zero matrix). Find a C such that $CA = 0$.

(b) What characterizes the zero divisors in $M_2(\mathbb{Q})$?

(c) Show that every nonzero element of $M_2(\mathbb{Q})$ (or more broadly $M_n(F)$ for F a field) is either a zero divisor or a unit.

In linear algebra, matrices over a field arise as functions that map one vector space to another. A square matrix maps a vector space to itself. Matrices over a ring can also be treated as functions: an $m \times n$ matrix over R maps a "vector" of length n over R to a "vector" of length m over R, using the familiar formulas for the product of a matrix and vector. I put vector in quotes because R^n is not a vector space when R is not a field. Yet, R^n, with rules for addition and scalar multiplication analogous to those for vector spaces, is an interesting object to study. Modules over a ring are the generalization of vector spaces. The subject is a bit more complex because not every nonzero element is a unit, and, even more challenging, there may be zero-divisors in the ring.

4.4 Ideals and Quotient Rings

For a group $G, *$ the operation can be extended to cosets of a subgroup, provided the subgroup N is normal. This allows for the construction of the quotient group G/N. For rings, we want both addition and multiplication to extend to cosets, and the appropriate subsets of a ring that allow for this are ideals.

Definition 4.4.1. An **ideal** of a ring R is a nonempty subset $I \subseteq R$ which is closed under addition and closed under multiplication by an arbitrary element of R:

$$a + b \in I \qquad \text{if } a, b \in I \tag{4.1}$$

$$ra \in I \qquad \text{if } a \in I \text{ and } r \in R \tag{4.2}$$

We will say that I **absorbs products**.

Proposition 4.4.2. *Let R be a ring.*

If an ideal I of R contains a unit, then $I = R$.

For any $a_1, a_2, \ldots, a_n \in R$, the following set is an ideal of R.

$$I = \{r_1a_1 + r_2a_2 + \cdots + r_na_n : r_i \in R\}$$

Proof. Let R be a ring. Let u be a unit in R with inverse v. If I is an ideal containing u then $uv = 1$ is also in I since I absorbs products. Again, since I absorbs products, for any $r \in R$, $r1 = r \in R$. Thus $I = R$.

Let $a_1, \ldots, a_n$ be arbitrary elements of r. We want to show

$$I = \{r_1a_1 + r_2a_2 + \cdots + r_na_n : r_i \in R\}$$

is an ideal of R. We can see that I is closed under addition because

$$\begin{aligned}(r_1a_1 + r_2a_2 + \cdots + r_na_n) &+ (s_1a_1 + s_2a_2 + \cdots + s_na_n)\\ &= r_1a_1 + s_1a_1 + r_2a_2 + s_2a_2 + \cdots + r_na_n + s_na_n\\ &= (r_1 + s_1)a_1 + (r_2 + s_2)a_2 + \cdots + (r_n + s_n)a_n\end{aligned}$$

The first step repeatedly uses commutativity and associativity of addition. The last step uses distributivity. The final expression is in a form that shows it is an element of I.

The product of any $t \in R$ with $r_1a_1 + r_2a_2 + \cdots + r_na_n$ is

$$t\big(r_1a_1 + r_2a_2 + \cdots + r_na_n\big) = (tr_1)a_1 + (tr_2)a_2 + \cdots + (tr_n)a_n$$

using distributivity and associativity of multiplication. The result is in a form to show it is in I. □

Definition 4.4.3. We say I **is generated by** $a_1, a_2, \dots, a_s$ if $I = \{r_1a_1 + r_2a_2 + \cdots + r_na_n : r_i \in R\}$. We write $I = \langle a_1, a_2, \dots, a_n \rangle$. The ideal I is **principal** if there exists some $a \in I$ such that $I = \{ra : r \in R\}$. A ring in which all ideals are principal is called a **principal ideal domain**, or a PID for short.

An ideal generated by several elements may also be generated by a single element, and therefore be principal.

Proposition 4.4.4. *Every ideal in $\mathbb{Z}$ is principal. Every ideal in $F[x]$ for F a field is principal.*

Proof. Let I be an ideal of $\mathbb{Z}$. If $I = \{0\}$, there is nothing to prove. Otherwise, let a be the smallest positive integer in I. Let b be any other nonzero element of I. Then, by the properties of ideals, any linear combination of a and b is in I. By the GCD Theorem 1.1.4, $\gcd(a, b) \in I$. But the gcd of a and b is positive and less than or equal to a. Since a is the smallest positive element of I, we must have $\gcd(a, b) = a$. Consequently, an arbitrary element of I is divisible by a, so $I = \langle a \rangle$ is principal.

The same proof applies to $F[x]$ with minor modification. For I a nonzero ideal of $F[x]$, one uses the monic polynomial of lowest degree in I.

□

From the proof we see that the ideals of $\mathbb{Z}$ are in one correspondence with the nonnegative integers. Similarly, the principal ideals, other than the 0-ideal, of $F[x]$ are in one to one correspondence with the monic polynomials.

Exercises 4.4.5.

(a) Let R be a subring of a ring S. Let I be an ideal in S. Show that $I \cap R$ is an ideal in R.

Exercises 4.4.6. Let R and S be rings and consider $R \times S$.

(a) Let I be an ideal in R and J and ideal in S. Show that $I \times J$ is an ideal in $R \times S$.

(b) Show that all ideals in $R \times S$ are of the form $I \times J$.

Quotient Rings

A ring R is an abelian group under addition, so, for any subgroup, we can form the quotient group of R by that subgroup. It is natural, when we take the multiplicative structure of R into account, to want the quotient group to also have a multiplicative structure. The necessary property to make this work is the "absorbs products" requirement in the definition of ideal. We have actually seen this appear in Proposition 4.2.7 which said that the kernel of a homomorphism is closed under multiplication by an arbitrary element of R.

We will write cosets of an ideal I in R in the same form as we did for abelian groups. For any $r \in R$, its coset is $r + I$. The coset $r + I$ may be written in other ways; for any $a \in I$, the cosets defined by r and by $r + a$ are the same. For $s \in R$, and $b \in I$ the cosets defined by s and $s + b$ are also the same. We would like to define multiplication of cosets, but for that to work, the product should be independent of the way we name the coset (as $r + I$ or as $(r + a) + I$). In other words we want the products rs and $(r + a)(s + b)$ to define the same cosets. We have

$$(r + a)(s + b) = rs + as + rb + ab$$

Since $a, b \in I$ and I absorbs products, $as + rb + ab \in I$. Thus $rs + I = (r + a)(s + b) + I$, and we have a well-defined product for cosets of I.

The following proposition summarizes this discussion, and we note that multiplication of cosets is determined by multiplication in R so we get a homomorphism from R to R/I.

Proposition 4.4.7. *Let R be a ring and let I be a proper ideal in R (that is $I \neq R$). Let $R/I = \{r + I : r \in R\}$. Then R/I is a ring, with additive structure defined by R/I as the quotient of the abelian group R by its subgroup I, and multiplicative structure defined by*

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

The additive identity is $0 + I$ and the multiplicative identity is $1 + I$.

The function $R \longrightarrow R/I$ that takes r to $r + I$ is a homomorphism of rings.

Example 4.4.8. In $\mathbb{Z}$, the subgroups $\langle n \rangle$ are also ideals, because multiplication is simply repeated addition, and a subgroup is closed under addition. Thus the quotient of $\mathbb{Z}$ by its subgroup $\langle n \rangle$ is a ring.

Example 4.4.9. Let F be a field. We have seen that any ideal in $F[x]$ is principal, generated by some $m(x) \in F[x]$. Every polynomial is congruent modulo $m(x)$ to its remainder upon division by $m(x)$.

We have already seen two familiar examples of quotient rings in Chapter 1, the integers modulo n and, for F a field, $F[x]$ modulo $m(x)$. The treatment of these in Chapter 1 is from a modular arithmetic perspective and we used brackets to define the equivalence class for elements. We now see them each as a quotients of a ring ($\mathbb{Z}$ or $F[x]$ repectively) and we now write the equivalence classes as cosets rather than using brackets. One convenient aspect of working with these quotient rings is that each element can be *uniquely represented* by the remainder upon division by the modulus. One can use $\{0, 1, \ldots, n - 1\}$ as the elements of $\mathbb{Z}/n$ and omit the brackets or coset notation when the context makes clear that we are working in $\mathbb{Z}/n$ rather than $\mathbb{Z}$. Similarly one can use polynomials of degree less than $\deg(m(x))$ as the elements of $F[x]/m(x)$.

Definition 4.4.10. Let R be a ring and I an ideal in R. A **system of representatives** for R/I is a set $S \subseteq R$ such that each $r \in R$ is congruent modulo I to exactly one element of S.

It can be more challenging to find a system or representatives for R/I in other rings, even in $F[x, y]$ with F a field. In polynomial rings over a field, monomial ideals and principal ideals admit a clear system or representatives, but things get much more complicated when there are several polynomial generators or when the base ring is not a field.

Definition 4.4.11. In a polynomial ring $R[x_1, x_2, \ldots, x_n]$, a **monomial ideal** is an ideal generated by monomials.

For the following two examples it is helpful to think of $F[x, y]$ as a vector space over F with basis $x^i y^j$ for $i, j \geq 0$.

Example 4.4.12. Consider a field F and $F[x, y]/\langle x^\alpha y^\beta \rangle$. Since an ideal is closed under multiplication, every monomial $ax^{\alpha'}y^{\beta'}$ with $\alpha' \geq \alpha$ and $\beta' \geq \beta$ is in $\langle x^\alpha y^\beta \rangle$. Since an ideal is closed under addition, and any sum of monomials with x-degree at least α and y-degree at least β is in $\langle x^\alpha y^\beta \rangle$. Conversely, any multiple of $x^\alpha y^\beta$ must be a polynomial whose terms all have x-degree at least α and y-degree at least β. The multidegrees of the monomials in $\langle x^\alpha y^\beta \rangle$ are those in the region marked Γ shown in Figure 4.1.

Given any polynomial we may subtract off the monomials that are multiples of $x^\alpha y^\beta$ and be left with a polynomial that has terms that have x-degree less than α or y-degree less than β. The multidegrees of these monomials are in the region marked Δ in Figure 4.1. These monomials are a basis for the quotient ring $F[x, y]/\langle x^\alpha y^\beta \rangle$. A set of representatives for $F[x, y]/\langle x^\alpha y^\beta \rangle$ is all polynomials whose terms have multidegree in Δ.

Example 4.4.13. Consider a field F and $F[x, y]/\langle x^3, xy, y^2 \rangle$. A monomial that is in $\langle x^3, xy, y^2 \rangle$ has multidegree that is in the region marked Γ in Figure 4.2. As in the previous example any polynomial whose terms involve just monomials with multidegree in Γ will be in $\langle x^3, xy, y^2 \rangle$. Given any other polynomial we may subtract off elements of $\langle x^3, xy, y^2 \rangle$ and be left with a polynomial all of whose terms are in the set Δ in Figure 4.2. The monomials with multidegree in Δ form a basis for a set of representatives for $F[x, y]/\langle x^3, xy, y^2 \rangle$. The set of representatives is all polynomials whose terms have multidegree in Δ.

Figure 4.1

Figure 4.2

Example 4.4.14. Consider now a principal ideal that is non-monomial, $F[x,y]/\langle y^2+x^3\rangle$. In this quotient ring $y^2=-x^3$. Given an arbitary polynomial we can replace y^2 with $-x^3$ and get a polynomial that is equivalent modulo y^2+x^3 and has terms that are at most degree 1 in y. Thus $\{x^i : i\in\mathbb{N}_0\}\cup\{x^iy : i\in\mathbb{N}_0\}$ is a basis for $F[x,y]/\langle y^2+x^3\rangle$. (It should be clear that no sum of these monomials can be a multiple of y^2+x^3.) Alternatively, we could replace each occurence of x^3 with $-y^2$ and obtain a basis $\{y^j : j\in\mathbb{N}_0\}\cup\{xy^j : j\in\mathbb{N}_0\}\cup\{x^2y^j : j\in\mathbb{N}_0\}$.

We have emphasized the strong relationship between the ring integers of $\mathbb{Z}$ and polynomial rings $F[x]$ for F a field. We have seen a couple of examples in $F[x,y]$. There are analogous examples for $\mathbb{Z}[y]$. We substitute a prime, say 2, for x and consider something analogous to a monomial ideal.

Example 4.4.15. Consider the ideal $\langle 4y^3, 8y^2\rangle$ in $\mathbb{Z}[y]$. The multiples of $4y^3$ will all have degree at least 4 (in y) and will have coefficient a multiple of 4. A set of representatives for $\mathbb{Z}[y]/\langle 4y^3,8y^2\rangle$ is $a_0+a_1y+a_2y^2+a_3y^3+\dots$ in which $a_k\in\{0,1,2,3\}$ for $k\geq 3$ and $a_2\in\{0,1,2,3,4,5,6,7\}$ and $a_0,a_1\in\mathbb{Z}$.

Exercises 4.4.16.

(a) Find the nilpotents, zero divisors, and units in $F[x,y]/\langle x^3y^2\rangle$.

(b) Find the nilpotents, zero divisors, and units in $F[x,y]/\langle x^3,y^2\rangle$.

(c) Find the nilpotents, zero divisors, and units in $F[x,y]/\langle x^3,xy,y^2\rangle$.

(d) Find the nilpotents, zero divisors, and units in $F[x,y]/\langle y^2-x^3\rangle$.

Exercises 4.4.17.

(a) We were careful in the last example to use powers of 2 as the coefficients in the ideal in $\mathbb{Z}[y]$. Find a system of representatives for $\mathbb{Z}[y]/\langle 4y^3,7y^2\rangle$.

(b) Find a system of representatives for $\mathbb{Z}[y]/\langle ay^5,by^4,cy^2,dy\rangle$. [Hint: GCD.]

4.5 Isomorphism Theorems

The First and Third Isomorphism Theorems for rings are quite straightforward extensions of the theorems for groups as are the Factor Theorem and Correspondence Theorem. The first theorem below combines the First Isomorphism Theorem and the Factor Theorem. The Second Isomorphism Theorem, treated at the end of the section, is less central in ring theory.

Theorem 4.5.1 (First Isomorphism and Factor Theorems). *Let $\varphi : R\longrightarrow S$ be a ring homomorphism and let K be the kernel. For any ideal J contained in K the homomorphism φ factors through R/J in the following sense: there is a ring homomorphism $\tilde{\varphi} : R/J\longrightarrow S$ defined by $r+J\longrightarrow\varphi(r)$ such that $\tilde{\varphi}\circ\pi=\varphi$.*

$$\begin{tikzcd}
R \arrow[r, "\varphi"] \arrow[d, "\pi"'] & S \\
R/J \arrow[ur, dashed, "\tilde{\varphi}"'] &
\end{tikzcd}$$

Additionally,

(1) If $J = K$ then $\tilde{\varphi}$ is injective.

(2) If φ is surjective then so is $\tilde{\varphi}$.

(3) If $J = K$ and φ is surjective then $\tilde{\varphi}$ is an isomorphism.

Proof. Let $\varphi : R \longrightarrow S$ be a ring homomorphism with kernel K and J an ideal contained in K. By the Factor Theorem for groups 2.7.11, we know there is well defined group homomorphism $\tilde{\varphi} : R/J \longrightarrow S$. This is because for any $r \in R$ and $j \in J$ we have $\varphi(r + j) = \varphi(r)$, so we can define $\tilde{\varphi}(r + J) = \varphi(r)$ unambiguously. The map is a group homomorphism because

$$\tilde{\varphi}(r_1 + r_2 + J) = \varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2) = \tilde{\varphi}(r_1 + J) + \tilde{\varphi}(r_2 + J)$$

Similarly the map respects multiplication because

$$\tilde{\varphi}(r_1 r_2 + J) = \varphi(r_1 r_2) = \varphi(r_1)\varphi(r_2) = \tilde{\varphi}(r_1 + J) * \tilde{\varphi}(r_2 + J)$$

Finally, $\tilde{\varphi}(1_R + J) = \varphi(1_R) = 1_S$, so $\tilde{\varphi}$ is a homomorphism of rings. By construction, $\varphi = \tilde{\varphi} \circ \pi$.

The kernel of $\tilde{\varphi}$ is $\{r + J : \varphi(r) = 0\}$. This is clearly $K/J = \{k + J : k \in K\}$. When $J = K$, $\tilde{\varphi}$ is injective since $\{k + K : k \in K\} = \{0 + K\}$.

Suppose φ is surjective. For any $s \in S$, there is some $\in R$ such that $\varphi(r) = s$. Then $\tilde{\varphi}(r + J) = \varphi(r) = s$, so $\tilde{\varphi}$ is also surjective. If $J = K$ and φ is surjective we have an isomorphism $\tilde{\varphi} : R/K \longrightarrow S$. □

We can derive the Third Isomorphism Theorem as a corollary. Consider the case $\varphi : R \longrightarrow R/K$ and let J be an ideal of R with $J \subseteq K$. The previous theorem says that $\tilde{\varphi} : R/J \longrightarrow R/K$ is a surjective homomorphism. The kernel is $\{r + J : \tilde{\varphi}(r + J) = 0 + K\}$. But $\tilde{\varphi}(r + J) = \varphi(r)$, so the kernel is $\{r + J : r \in K\} = K/J$. Now applying item (3) of the theorem to $\tilde{\varphi}$ we get an isomorphism between $(R/J)/(K/J)$ and R/K.

We can also prove the Third Isomorphism Theorem directly, which we do below.

Theorem 4.5.2 (Third Isomorphism Theorem)**.** *Let R be a ring with ideals K and J such that $J \subseteq K$. Then K/J is an ideal of R/J and $(R/J)/(K/J) \cong R/K$.*

Proof. I claim there is a well defined function from R/J to R/K defined by $\varphi(r+J) = r+K$. We need only check that if $r_1 + J = r_2 + J$ then $r_1 + K = r_2 + K$. This is clearly true because if $r_1 + J = r_2 + J$ then $r_1 - r_2 \in J$. Since $J \subseteq K$ we have $r_1 = r_2 \in K$ so $r_1 + K = r_2 + K$. We clearly also have

$$\begin{aligned}
\varphi\big((r_1 + J) + (r_2 + J)\big) &= \varphi(r_1 + r_2 + J) \\
&= r_1 + r_2 + K \\
&= (r_1 + K) + (r_2 + K) \\
&= \varphi(r_1 + K) + \varphi(r_2 + K)
\end{aligned}$$

and a similar computation holds for multiplication. Thus φ is a homomorphism and it is surjective.

The kernel is $\{r + J : r + K = 0 + K\}$, but this is $\{r + J : r \in K\} = K/J$. So by the first isomorphism theorem, $(R/J)/(K/J) \cong R/K$. □

From Theorem 4.5.1 any surjective homomorphism $R \longrightarrow S$ gives rise to an isomorphism between R/K and S where K is the kernel of $R \longrightarrow S$. A strengthening of Theorem 4.5.2 is the following.

Theorem 4.5.3 (Correspondence)**.** *Let $R \longrightarrow S$ be a surjective homomorphism of rings with kernel K. There is a one-to-one correspondence, given by φ, between ideals of S and ideals of R containing K.*

$$\begin{aligned}
R &\longrightarrow S \\
I &\longleftrightarrow \varphi(I) \\
\varphi^{-1}(J) &\longleftrightarrow J
\end{aligned}$$

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

- *$K \leq I \leq I'$ if and only if $\varphi(I) \leq \varphi(I')$.*
- *The map φ induces an isomorphism $R/I \cong S/\varphi(I)$.*

Theorem 4.5.4 (Second Isomorphism)**.** *Let S be a subring of R and let J be an ideal in R.*

(1) $S + J$ is a subring of R.

(2) $S \cap J$ is an ideal in S.

(3) $S/(S \cap J) \cong (S + J)/J$.

Proof. The first two items are left as exercises. Consider the homomorphism $\varphi : S \longrightarrow R/J$, which is the composition of the inclusion map : $S \longrightarrow R$ and the quotient map $R \longrightarrow R/J$. The image of φ is $\{s + J : s \in S\}$ and it is the quotient of the subring $S + J$ in R by the ideal J. The kernel of φ is $S \cap J$. By the first isomorphism theorem, $S/(S \cap J) \cong (S + J)/J$.

□

Exercises 4.5.5.

(a) Prove (1) and (2) of the Second Isomorphism Theorem.

Exercises 4.5.6. Let $\varphi : R \longrightarrow S$ be a homomorphism of rings, not necessarily surjective.

(a) Let J be an ideal in S. Show that $\varphi^{-1}(J)$ is an ideal in R.

(b) Give an example to show that for I an ideal in R, $\varphi(I)$ may not be an ideal in S.

4.6 Operations on Ideals and Properties of Ideals

Intersection, Sum and Product of Ideals

Proposition 4.6.1. *Let I and J be ideals. Then $I \cap J$ is an ideal. More generally, if $\mathcal{A}$ is a set of ideals in R then*

$$\bigcap_{I \in \mathcal{A}} I$$

is an ideal in R.

Proof. Let $a, b \in \bigcap_{I \in \mathcal{A}} I$. Then we have $a, b \in I$ for each $I \in \mathcal{A}$. Since each I is closed under addition, $a + b \in I$ for all $I \in \mathcal{A}$. Consequently, $a + b \in \bigcap_{I \in \mathcal{A}} I$.

Similarly, we can show that $\bigcap_{I \in \mathcal{A}} I$ absorbs products. For $r \in R$ and $a \in \bigcap_{I \in \mathcal{A}} I$ we have $a \in I$ for each $I \in \mathcal{A}$. Since each I absorbs products, $ra \in I$ for each $I \in \mathcal{A}$. Thus $ra \in \bigcap_{I \in \mathcal{A}} I$. □

Definition 4.6.2. Let I and J be ideals. The **sum** of I and J is $I + J = \{a + b : a \in I \text{ and } b \in J\}$. The **product** of I and J is $IJ = \langle ab : a \in I \text{ and } b \in J\rangle$. Similarly for ideals $I_1, \ldots, I_n$ in R we can define

$$I_1 + I_1 + \cdots + I_n = \{r_1 a_1 + r_2 a_2 + \cdots + r_n a_n : r_k \in R, a_k \in I_k\}$$
$$I_1 I_2 \ldots I_n = \langle a_1 a_2 \cdots a_n : a_k \in I_k \text{ for all } k \in \{1, \ldots, n\}\rangle$$

There is a subtle, but very important, difference in the two definitions. The product of ideals I and J is defined to be *generated* by the set of elements ab with

$a \in I$ and $b \in J$. The sum of I and J is *the set* of all sums $a + b$ with $a \in I$ and $b \in J$. We must show that this set is in fact an ideal.

Proposition 4.6.3. *Let $I_1, \ldots, I_n$ be ideals in R. Then $I_1 + \cdots + I_n$ is also an ideal.*

Proof. We have to show $I_1 + \cdots + I_n$ is closed under sums and that it absorbs products. Consider two arbitrary elements of $I_1 + I_2 + \cdots + I_n$, which we may write as $a_1 + a_2 + \cdots + a_n$ and $b_1 + b_2 + \cdots + b_n$ with $a_k, b_k \in I_k$ for $k = 1, \ldots, n$. Their sum is, after rearranging using commutativity and associativity, $(a_1 + b_1) + (a_2 + b_2) + \cdots + (a_n + b_n)$. Since each $(a_k + b_k) \in I_k$, the sum is an element of $I_1 + I_2 + \cdots + I_n$.

Let $r \in R$ and $a_1 + a_2 + \cdots + a_n$ with $a_k \in I_k$. Then by distributivity

$$r(a_1 + a_2 + \cdots + a_n) = ra_1 + ra_2 + \cdots + ra_n$$

Since each I_k absorbs products, each $ra_k \in I_k$. Thus

$$ra_1 + ra_2 + \cdots + ra_n \in I_1 + I_2 + \cdots + I_n$$

this shows $I_1 + I_2 + \cdots + I_n$ absorbs products. □

Exercises 4.6.4.

(a) In the integers, show that the sum $\langle a \rangle + \langle b \rangle = \langle \gcd(a, b) \rangle$.

(b) In the integers, show that the intersection $\langle a \rangle \cap \langle b \rangle = \langle \operatorname{lcm}(a, b) \rangle$.

(c) Extend these results to $F[x]$ for F a field.

Proposition 4.6.5. *Let $I_1, \ldots, I_n$ be ideals. Then*

$$I_1 I_2 \cdots I_n \subseteq I_1 \cap I_2 \cap \cdots \cap I_n$$

Proof. The product $I_1 I_2 \cdots I_n$ is generated by elements of the form $a_1 a_2 \cdots a_n$ with each $a_k \in I_k$. Since each I_k absorbs products, $a_1 a_2 \cdots a_n \in I_k$ for all k. Thus $a_1 a_2 \cdots a_n \in I_1 \cap I_2 \cap \cdots \cap I_n$. Since the generators of $I_1 I_2 \cdots I_n$ are all in $I_1 \cap I_2 \cap \cdots \cap I_n$ we have

$$I_1 I_2 \cdots I_n \subseteq I_1 \cap I_2 \cap \cdots \cap I_n$$

□

The sum of an arbitrary set of ideals in R (including an infinite set) is defined in a similar fashion, but requires care because we must restrict to finite sums.

Definition 4.6.6. Let $\mathcal{A}$ be a set of ideals in R. The **sum** of these ideals is

$$\sum_{I \in \mathcal{A}} I = \left\{ \sum_{I \in \mathcal{B}} a_I : a_I \in I, \text{ and } \mathcal{B} \text{ is a finite subset of } \mathcal{A} \right\}$$

Exercises 4.6.7. Let $\varphi : R \longrightarrow S$ be a homomorphism.

(a) For I and ideal in R show that $\varphi^{-1}(\varphi(I)) = I + K$ where $K = \ker \varphi$. In particular, if I contains K, then $\varphi^{-1}\varphi(I) = I$.

Maximal, Prime and Radical Ideals

There are three key properties that ideals may have.

Definition 4.6.8. Let I be a proper ideal of R (that is $I \neq R$). An ideal I is **maximal** if the only ideal properly containing I is R. The ideal I is **prime** when $ab \in I$ implies that either $a \in I$ or $b \in I$. The ideal I is **radical** when $a^n \in I$ for $n \in \mathbb{N}$ implies $a \in I$.

In any integral domain, the zero ideal is prime. This follows directly from the definition of integral domain, $ab = 0$ implies $a = 0$ or $b = 0$.

Let I be a nonzero ideal in the integers, and let d be its positive generator. If d is not a prime number, say $d = ab$ with $a, b < d$, then I is not a prime ideal since $ab \in I$, but $a, b \notin I$. If d is a prime number then $ab \in I$ implies that $d|ab$, and by primality of d, either $d|a$—and therefore $a \in I$—or $d|b$—and then $b \in I$. Consequently, I is prime. We conclude that, for the integers, an ideal is prime if and only if it is generated either by 0 or a prime integer.

Exercises 4.6.9.

(a) Show that the nonzero prime ideals in $\mathbb{Z}$ are also maximal ideals. [Suppose p is a prime number. Try to enlarge $\langle p \rangle$ and show that you get all of $\mathbb{Z}$.]

(b) Let I be a nonzero ideal in $\mathbb{Z}$. We know I is principle; let a be the smallest positive integer in I. Show that I is radical if and only if the prime factorization of a is $a = p_1 p_2 \cdots p_r$ for *distinct* primes p_i.

(c) Extend these results to $F[x]$ for F a field.

Theorem 4.6.10. *All prime ideals are radical. All maximal ideals are prime.*

Proof. Let P be a prime ideal. We will show that if $a^n \in P$ then $a \in P$. This establishes that P is a radical ideal. Suppose $a^n \in P$. Let $m \leq n$ be the smallest power of a that lies in P. If $m > 1$, then we have $a * a^{m-1} = a^m \in P$. By primality, either a or a^{m-1} is in P. This contradicts our assumption on m. Thus $m = 1$ and $a \in P$.

Let M be a maximal ideal. We will show that $ab \in M$ implies either a or b is in M. This establishes that M is prime. Let $ab \in M$. Suppose that $a \notin M$. Since M is a maximal ideal $M + \langle a \rangle = R$. Consequently there is some $m \in M$ and $r \in R$ such that $m + ra = 1$. Multiplying both sides by b we get $mb + rab = b$. Since $ab \in M$, we have $b = mb + rab \in M$. Thus if $ab \in M$ and $a \notin M$ then $b \in M$, as was to be shown. □

The proof that a maximal ideal is prime echoes the proof that an irreducible integer (or polynomial in $F[x]$ for F a field) is prime, Theorem 1.1.9.

Now we show that these properties of ideals are intimately connected with properties of the quotient ring.

Theorem 4.6.11. *Let R be a ring and I and ideal in R.*

- *I is a maximal ideal if and only if R/I is a field.*
- *I is a prime ideal if and only if R/I is an integral domain.*
- *I is a radical ideal if and only if R/I is reduced.*

Proof. We will prove one direction for each claim and leave the other as an exercise.

Suppose I is maximal. Let $r + I$ be an arbitrary element of R/I with $r + I \neq 0 + I$. Since I is maximal, $I + \langle r \rangle = R$, so there is some $a \in I$ and $s \in R$ such that $a + sr = 1$. Then $sr + I = (1 - a) + I = 1 + I$, because $a \in I$. Consequently, $s + I$ is the inverse of $r + I$. Thus an arbitrary nonzero element of R/I has an inverse, and R/I is a field.

Suppose I is a prime ideal. Let $r+I$ and $s+I$ be such that $(r+I)(s+I) = 0+I$. Then $rs + I = 0 + I$ so $rs \in I$. Since I is prime, either $r \in I$ or $s \in I$. Thus, either $r + I = 0 + I$ or $s + I = 0 + I$. This shows R/I has no zero-divisors.

Suppose I is a radical ideal. Suppose that $r + I$ is nilpotent in R/I; that is $(r + I)^n = 0 + I$. Then $r^n + I = 0 + I$, so $r^n \in I$. Since I is radical, we must have $r \in I$, and consequently $r + I = 0 + I$. This shows that R/I has no nonzero nilpotent elements, so R/I is reduced.

□

Exercises 4.6.12. Let $\varphi : R \longrightarrow S$ be a homomorphism of rings and let J be an ideal in S. From Exercise 4.5.6 we know that $\varphi^{-1}(J)$ is an ideal in R.

(a) If J is a radical ideal, show that $\varphi^{-1}(J)$ is a radical ideal in R.

(b) If J is a prime ideal, show that $\varphi^{-1}(J)$ is a prime ideal in R.

(c) Using $R = \mathbb{Z}$ and $S = \mathbb{Q}$ show that $\varphi^{-1}(J)$ may not be maximal when J is maximal.

Exercises 4.6.13.

(a) Show that the intersection of two radical ideals is radical.

(b) Illustrate with an example from $F[x]$ for F a field.

(c) Given an example in $F[x]$ to show that the intersection of two prime ideals may not be prime.

Exercises 4.6.14.

(a) Let $N = \{a \in R : a^n = 0 \text{ for some } n \in \mathbb{N}\}$ be the set of all the nilpotent elements in a ring R. Show that N is an ideal of R. It is called the **nilradical** of R.

(b) Show that R/N is reduced (it has no nonzero nilpotent elements).

(c) Show that N is contained in the intersection of all prime ideals in R. (The reverse containment is also true, but much more difficult to prove.)

(d) Show that if $a \in N$ then $1 - a$ and $1 + a$ are units.

Comaximality and the Chinese Remainder Theorem

Recall that two integers a, b are *coprime* if they have no common factor other than 1. A consequence—the GCD Theorem 1.1.4—is that some linear combination of a and b is equal to 1. Interpreting this in the context of ideals, the ideal $\langle a, b\rangle$ generated by the two coprime integers is all of $\mathbb{Z}$. We can extend this notion of coprime integers (or polynomials) to ideals in a general ring R.

For any two integers m, n we have a homomorphism $\mathbb{Z}/mn \longrightarrow \mathbb{Z}/m$ and $\mathbb{Z}/mn \longrightarrow \mathbb{Z}/n$. This gives a homomorphism into the direct product by Proposition 4.3.4: $\mathbb{Z}/mn \longrightarrow \mathbb{Z}/m \times \mathbb{Z}/n$. The Chinese Remainder Theorem says this is an isomorphism when m, n are coprime.

Theorem 4.6.15 (Chinese Remainder Theorem)**.** *Let m, n be coprime integers. The natural maps from Theorem 4.2.12 give an isomorphism $\mathbb{Z}/mn \longrightarrow \mathbb{Z}/m \times \mathbb{Z}/n$.*

The proof is based on the GCD Theorem, which says that there are integers u and v such that $mu+nv = 1$. More generally we have the property of comaximality for ideals, and a Chinese Remainder Theorem for comaximal ideals whose proof essentially mimics the proof of the Chinese Remainder Theorem for integers.

Definition 4.6.16. Two ideals I and J in a ring R are **comaximal** if $I + J = R$.

Theorem 4.6.17. *Let I and J be proper ideals of R that are comaximal. Then $IJ = I \cap J$ and $R/IJ \cong R/I \times R/J$.*

Proof. Since I and J are comaximal there exist $a \in I$ and $b \in J$ such that $a+b = 1$. We now show that the homomorphism $R \longrightarrow R/I \times R/J$ is surjective. The image of a in R/J is $a + J = (1 - b) + J = 1 + J$ because $b \in J$. The image of a in R/I is $a + I = 0 + I$ because $a \in I$. Similarly, the image of b is $1 + I$ in R/I and it is $0 + J$ in R/J. Thus for an arbitrary element $(r_1 + I, r_2 + J)$ in $R/I \times R/J$, there is a preimage, $r_1 a + r_2 b$.

The kernel of $R \longrightarrow R/I \times R/J$ is $I \cap J$ so the proof is complete once we show $IJ = I \cap J$. We already know that $IJ \subseteq I \cap J$. Let $c \in I \cap J$. Then $ac + bc = c$, but ac and bc are both in IJ so we have expressed an arbitrary element of $I \cap J$ as a sum of two elements in IJ. Thus $IJ = I \cap J$. □

4.7 Fractions

Dealing with fractions is one of the big challenges for primary school students. A key reason for the difficulties is that a fraction can be written in an infinite number of equivalent ways (for example $1/2 = 2/4 = 3/6 \ldots$) and it is necessary to use multiple expressions for a number in order to do arithmetic with fractions. Underlying our use of fractions is an equivalence relation on ordered pairs of integers; which is not something we dare to explain to students. In this section we show that the method used to construct the rational numbers from the integers extends with little modification to an arbitrary ring.

First let's consider some examples to show that there are other rings of interest in between the integers and the rational numbers, that is, rings properly containing $\mathbb{Z}$ but properly contained in $\mathbb{Q}$.

Example 4.7.1. One can verify that the following sets are in fact subrings of $\mathbb{Q}$.

- $R = \{a/2^i : a \in \mathbb{Z}, i \in \mathbb{N}_0\}$.
- $S = \{a/b : a \in \mathbb{Z} \text{ and } b \text{ is an odd integer}\}$.
- $T = \{a/100^i : a \in \mathbb{Z}, i \in \mathbb{N}_0\}$.

Exercises 4.7.2. For the rings R, S, and T:

(a) Identify all the units in each of these rings.

(b) Show that in each of these rings every ideal is principal, generated by some nonnegative integer.

(c) In $\mathbb{Z}$, any two distinct positive integers generate different ideals. Show that is not true in R, S, T. For each of these rings, identify a set of integers that uniquely define all ideals.

(d) Which of these ideals are prime?

The rings in the previous example and exercise are all constructed via the process we now describe.

Definition 4.7.3. Let R be an integral domain. A subset D of $R \setminus \{0\}$ that contains 1 and is closed under multiplication is called a **multiplicatively closed set**.

Let D be a multiplicatively closed set in R. Define a relation on $R \times D$ by

$$(r_1, d_1) \sim (r_2, d_2) \quad \text{when} \quad r_1 d_2 = r_2 d_1$$

Proposition 4.7.4. *The relation above is an equivalence relation. Under this relation for any $r \in R$ and $c, d \in D$, $(r, d) \sim (rc, dc)$*

Proof. The relation is reflexive:$(r, d) \sim (r, d)$ since $rd = rd$.

The relation is symmetric: Suppose $(r_1, d_1) \sim (r_2, d_2)$ so $r_1 d_2 = r_2 d_1$. Then $r_2 d_1 = r_1 d_2$ so $(r_2, d_2) \sim (r_1, d_1)$.

The relation is transitive: Suppose $(r_1, d_1) \sim (r_2, d_2)$ and $(r_2, d_2) \sim (r_3, d_3)$. Then $r_1 d_2 = r_2 d_1$ and $r_2 d_3 = r_3 d_2$. Multiplying the first by d_3 and the second by d_1 we get

$$r_1 d_2 d_3 = r_2 d_1 d_3 = r_2 d_3 d_1 = r_3 d_2 d_1$$

Since R is an integral domain, we can cancel d_2 (see Exercise 4.1.6) to obtain $r_1 d_3 = r_3 d_1$. This shows $(r_1, d_1) \sim (r_3, d_3)$.

The final claim follows from the definition of the relation $rdc = drc$. It may be seen as simplification of fractions. □

A key step in the above proof involved the cancellation for integral domains. The construction of rings of fractions can be generalized to arbitrary rings provided D contains no zero-divisors. It can be generalized to allow D to contain zero-divisors with one small modification to the definition of the equivalence relation.

Theorem 4.7.5. *Let D be a multiplicatively closed set in R. Let $[r, d]$ denote the equivalence class of (r, d). The operations*

- $[r, c] + [sd] := [rd + sc, cd]$*, and*
- $[r, c] \star [s, d] := [rs, cd]$*,*

are well defined. The set $R \times D/\sim$ with these operations is a ring with additive identity $[0, 1]$ and multiplicative identity $[1, 1]$. We denote this ring $D^{-1}R$. The map $R \longrightarrow D^{-1}R$ taking r to $[r, 1]$ is an embedding.

Proof. Let (r, c) and (s, d) be in $S \times D$. Since D is multiplicatively closed, $cd \in D$ so both $(rd + sc, cd)$ and (rs, cd) are in $S \times D$ and their equivalence classes exist.

To show that the operations are well defined, suppose two different representatives for each equivalence class: $(r, c) \sim (r', c')$ and $(s, d) \sim (s', d')$. We want to show that the formula for the equivalence class of the product (and for the sum) is independent of the representatives chosen. We deal with the product first. We want to show that $(rs, cd) \sim (r's', c'd')$, which reduces to $rsc'd' = r's'cd$. We know

$$rc' = r'c \tag{4.3}$$

$$sd' = s'd \tag{4.4}$$

Multiplying the first equation by sd' and the second by rc' we get

$$rc'sd' = r'csd' = r'csd' = r'cs'd$$

Rearranging the factors on the first and last terms gives $rsc'd' = r's'cd$.

For the sum we want to show $(rd+sc, cd) \sim (r'd'+s'c', c'd')$. Multiplying (4.3) bu dd' we get $rc'dd' = r'cdd'$, and multiplying (4.4) by cc' gives $sd'cc' = s'dcc'$. Adding the two equations

$$\begin{aligned} rc'dd' + sd'cc' &= r'cdd' + s'dcc' \\ (rd+sc)c'd' &= (r'd' + s'c')cd \end{aligned}$$

This establishes $(rd+sc, cd) \sim (r'd' + s'c', c'd')$.

Verifying the claims about the additive and multiplicative identity are routine computations. Verification of commutativity and associativity are more involved by fairly straightforward and are left to the reader. We next show that multiplication distributes over addition.

$$\begin{aligned} [r,c]\Big([s,d]+[t,f]\Big) &= [r,c]\Big[sf + td, df]\Big) \\ &= [rsf + rtd, cdf] \\ [r,c]*[s,d] + [r,c]*[t,f]\Big) &= [rs, cd] + [rt, cf] \\ &= [rscf + rtcd, cfcd] \qquad = [rsf + rtd, cdf] \end{aligned}$$

The function $\iota : R \longrightarrow D^{-1}R$ taking r to $[r,1]$ takes the identity to the identity element of R to the identity element of $D^{-1}R$. It respects sums since $[r,1]+[s,1] = [r*1+s*1, 1*1] = [r+s, 1]$ which is the image of $r+s$. The map ι respects products since $[r,1]*[s,1] = [rs, 1*1]$ which is the image of rs. Thus ι is a homomorphism. Suppose $\iota(r) = \iota(s)$. Then $[r,1] = [s,1]$. By the definition of the equivalence relation, $r*1 = s*1$. This shows ι is injective. □

The ring $D^{-1}R$ is often called a **localization** of R.

Exercises 4.7.6.

(a) Let $D = \{30^i : i \in \mathbb{N}_0\}$. Verify that D is multiplicatively closed in $\mathbb{Z}$. Identify all of the prime ideals in $D^{-1}\mathbb{Z}$.

(b) Let $D = \{(x^3 - x)^i : i \in \mathbb{N}_0\}$. Verify that D is multiplicatively closed in $\mathbb{Q}[x]$. Identify all of the prime ideals in $D^{-1}\mathbb{Q}[x]$.

(c) Under what conditions on D does $D^{-1}\mathbb{Z}$ have just one maximal ideal?

(d) Let D be multiplicatively close I be an

Exercises 4.7.7. In this exercise we characterize all localizations of $\mathbb{Z}$.

(a) Let $D = \{d_1, d_2, \ldots, d_t\} \subseteq \mathbb{Z} \setminus \{0\}$. What is the smallest multiplicatively closed subset of $\mathbb{Z}$ containing D and 1? What if D is an infinite subset of $\mathbb{Z} \setminus \{0\}$? We will abuse notation and write $D^{-1}\mathbb{Z}$ for the localization due to the smallest multiplicatively closed subset of $\mathbb{Z} \setminus \{0\}$ containing D.

(b) Suppose $D \subseteq N \subseteq \mathbb{Z} \setminus \{0\}$. Show that there is an injective ring homomorphism $D^{-1}\mathbb{Z} \to N^{-1}\mathbb{Z}$.

(c) Show that any localization of $\mathbb{Z}$ is of the form $P^{-1}\mathbb{Z}$ where P is a subset of the set of primes in $\mathbb{N}$. [You need to identify the set P for a given D and show that $P^{-1}\mathbb{Z} \cong D^{-1}\mathbb{Z}$].

(d) Let P be a subset of the prime integers. Identify all the ideals in $P^{-1}\mathbb{Z}$. Which ideals are prime?

Exercises 4.7.8. Let R be an integral domain and let D be a multiplicative subset of R. We will consider R as a subset of $D^{-1}R$ via the embedding $\varphi : R \longrightarrow D^{-1}R$ which takes r to $r/1$.

(a) Let $D^{-1}I = \{a/s : a \in I, s \in D\}$. Show $D^{-1}I$ is an ideal in $D^{-1}R$.

(b) Show that $D^{-1}I = D^{-1}R$ if and only if $I \cap D \neq \emptyset$.

(c) Let J be an ideal of $D^{-1}R$. Show that $J \cap R$ is an ideal in R.

Parts a - c show we have a function from the set of ideals in $D^{-1}R$ to the set of ideals in R given by $J \mapsto J \cap R$ and a function from the set of ideals in R to the set of ideals in $D^{-1}R$ given by $I \mapsto D^{-1}I$.

(d) Show that $I \mapsto D^{-1}I$ is surjective: That is, show that every ideal in $D^{-1}R$ is $D^{-1}I$ for some ideal I in R. (Hints: If J is an ideal in $D^{-1}R$ then an element of J may be written a/s for $a \in R$ and $s \in D$. Show that $D^{-1}(J \cap R) = J$.)

(e) Show that these two maps of ideals respect intersections. For example, $D^{-1}(I \cap I') = D^{-1}(I) \cap D^{-1}(I')$.

(f) The map $I \mapsto D^{-1}I$ is not injective. Show that it is injective on prime ideals that don't meet D. Conclude that the functions $J \mapsto J \cap R$ and $I \mapsto D^{-1}I$ give a 1-1 correspondence between prime ideals of $D^{-1}R$ and prime ideals of R not meeting D.

Exercises 4.7.9. Let R be an integral domain. A multiplicatively closed set $D \subseteq R$ is *saturated* when

$$xy \in D \Longleftrightarrow x \in D \text{ and } y \in D.$$

There is a theorem saying D is saturated if and only if $R \setminus D$ is a union of prime ideals. Prove one direction of this result as follows.

(a) Let $\mathcal{P}$ be a set of prime ideals and let $D = R \setminus \Big(\cup_{P \in \mathcal{P}} P \Big)$. Show that D is multiplicatively closed and saturated.

Chapter 5

Fields

5.1 First Fields and Automorphisms

Definition 5.1.1. A **field** is a set F with two binary operations, $+$ and $*$, called addition and multiplication, two special elements 0 and 1, and two unary operations, $a \longmapsto -a$, and, for all but the 0 element, $a \longmapsto a^{-1}$ such that

- F is an abelian group under $+$ with identity element 0 and additive inverse $a \longmapsto -a$.
- $F^* = F \setminus 0$ is an abelian group under $*$ with identity element 1 and multiplicative inverse $a \longmapsto a^{-1}$.
- Multiplication distributes over addition: $a * (b + c) = a * b + a * c$.

There are a few fields that should be familiar to you; the following were discussed in the first chapter.

- The rational numbers $\mathbb{Q}$. This is the smallest field that contains the integers.
- The prime fields, $\mathbb{F}_p$ for each prime number p. A fundamental result from modular arithmetic is that each nonzero element in $\mathbb{Z}/p$, the ring of integers modulo p, is invertible. One can compute the inverse of a nonzero element by using the extended Euclidean algorithm. This shows that $\mathbb{Z}/p$ is a field. When studying fields we will write $\mathbb{F}_p$ instead of $\mathbb{Z}/p$.
- The real field, $\mathbb{R}$.
- The field of complex numbers $\mathbb{C}$. The complex numbers also form a vector space of dimension 2 over $\mathbb{R}$ with basis $\{1, i\}$ where $i = \sqrt{-1}$. That is,

every element can be written in a unique way as $a + bi$ for $a, b \in \mathbb{R}$ and the properties of a vector space hold for scalar multiplication by a real number.

- Inside the field of complex numbers is the Gaussian rationals (see 1.2)

$$\mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\}$$

  This is a field, and also a two-dimensional vector space over $\mathbb{Q}$.

Definition 5.1.2. Let K be a field and let F be a subset of K such that F is a field using the operations $*_K$ and $+_K$. We say F is a **subfield** of K and K is an **extension field** of F. We will write $F \leq K$ and also K/F depending on whether the emphasis is on F being a subfield of K or K and extension of F.

Exercises 5.1.3.

(a) Let F and E be subfields of K. Show that $F \cap E$ is a subfield of K.

(b) Let $\mathcal{F}$ be a set of subfields of K, then

$$\bigcap_{F \in \mathcal{F}} F$$

is a subfield of K.

(c) Let $\mathcal{R}$ be a set of subrings of a ring S, then

$$\bigcap_{R \in \mathcal{R}} R$$

is a subring of S.

If $F \leq K$ then K is a vector space over F. We write $[K : F]$ for the dimension. It is also called the **degree** of the extension.

For any field F, there is a ring homomorphism $\mathbb{Z} \longrightarrow F$ taking 1 to 1_F by Theorem 4.2.11. If the kernel is trivial then F contains a subring isomorphic to the integers, and since F is a field it must contain a subfield isomorphic to the rationals, $\mathbb{Q}$. If the kernel is not trivial, then F contains a subring isomorphic to $\mathbb{Z}/m$ for some integer m. Since F is a field, $\mathbb{Z}/m$ cannot have zero-divisors, so m must be prime. Thus we have two cases, a field F either contains a copy of $\mathbb{F}_p$ or $\mathbb{Q}$. This smallest field contained in F is called its **prime field**. We say F has **characteristic** p, when $\mathbb{F}_p \leq F$ or characteristic 0 in the case $\mathbb{Q} \leq F$.

As with groups and with rings, a natural topic to investigate is the functions that respect the structure of fields.

Definition 5.1.4. For fields F and K, a function $\varphi : F \longrightarrow K$ is a homomorphism when

(1) φ is a homomorphim from the group $F, +_F$ to $K, +_K$, and

(2) φ is a homomorphism from the group $F^*, *_F$ to $K^*, *_K$.

Applying Proposition 2.2.2, $\varphi : F \longrightarrow K$ is a homomorphism of fields if it respects addition and multiplication:

$$\begin{aligned} \varphi(a_1 +_F a_2) &= \varphi(a_1) +_K \varphi(a_2), \text{ and} \\ \varphi(a_1 *_F a_2) &= \varphi(a_1) *_K \varphi(a_2) \end{aligned}$$

In these two equations I have emphasized that the addition and multiplication on the left is done in F and the addition and multiplication on the right is in K. Generally, we follow standard practice and do *not* write the subscripts on the operation signs to make the equations more legible. But, don't forget the distinction! We will also usually not write the multiplication sign, unless there is some important reason to use it.

It turns out that a homomorphisms of fields is always injective!

Proposition 5.1.5. *Let $\varphi : F \longrightarrow K$ be a homomorphism of fields. Then $\varphi(a) = \varphi(b)$ implies $a = b$, so φ is injective.*

Proof. Let $\varphi : F \longrightarrow K$ be a homomorphism. Let a be a nonzero element of F. Since $aa^{-1} = 1_F$, applying φ we get $\varphi(a)\varphi(a^{-1}) = 1_K$. Since 0_K does not have a multiplicative inverse, $\varphi(a)$ cannot be 0_K. Thus $a \neq 0_F$ implies $\varphi(a) \neq 0_K$.

Now suppose $\varphi(a) = \varphi(b)$. Then $\varphi(a-b) = 0_K$, and the contrapositive of what we showed in the previous paragraph gives $a - b = 0$, so $a = b$. □

A homomorphism of fields $\varphi : F \longrightarrow K$ is often called an **embedding** of F in K since it places an isomorphic copy of F, namely $\varphi(F)$ inside of K.

The next proposition is completely analogous to results about the composition of homomorphisms of groups, Proposition 2.2.2, and properties of isomorphisms, Proposition 2.2.9.

Proposition 5.1.6.

(1) The composition of two field homomorphisms is a field homomorphism.

(2) The composition of two isomorphisms of fields is an isomorphism of fields.

(3) Let $\varphi : F \longrightarrow K$ be an isomorphism of fields. The inverse function $\varphi^{-1} : K \longrightarrow F$ is also an isomorphism of fields.

Proof. (1) We have already shown that the composition of group homomorphisms is a group homomorphism. Thus the composition of two field homomorphisms $\alpha : F \longrightarrow H$ and $\beta : H \longrightarrow K$ is is both a homomorphism of the additive group $F, +_F$ to $K, +_K$ and a homomorphism of the multiplicate group $F, *_F$ to $K, *_K$. Therefore $\beta \circ \alpha$ is a field homomorphim.

(2) The composition of two bijections is a bijection and the composition of two isomorphisms is an isomorphism.

(3) We can apply the fact that the inverse of an isomorphism of groups is also an isomorphism of groups to prove the result in a similar fashion to the first item.

□

And now the culmination of this section!

Definition 5.1.7. Let K be a field. The **automorphism group** of K is the set of all isomorphisms from K to itself, with the operation of composition. It is written $\mathrm{Aut}(K)$. Let F be a subfield of K. An automorphism σ such that $\sigma(a) = a$ for all $a \in F$ is said to **fix** F. The set of automorphisms K that **fix** F is denoted $\mathrm{Aut}(K/F)$.

The previous proposition shows that the composition of automorphisms is an automorphism and that every isomorphism has an inverse, so $\mathrm{Aut}(K)$ is a group. It is a simple exercise to show that the composition of two automorphisms that fix F also fixes F and that the inverse of an automorphism that fixes F also fixes F.

Corollary 5.1.8. *For K a field,* $\mathrm{Aut}(K)$ *is a group under composition. If F is a subfield of K,* $\mathrm{Aut}(K/F)$*, the automorphisms of K that fix F, form a group under composition.*

What can we say about automorphisms of the fields introduced above? First note that any automorphism has to take 1 to itself. Consider an automorphism φ of $\mathbb{Q}$. We must have $\varphi(1) = 1$. Since φ respects addition,

$$\varphi(\underbrace{1 + \cdots + 1}_{b \text{ terms}}) = \underbrace{\varphi(1) + \cdots + \varphi(1)}_{b \text{ terms}}$$

which shows that $\varphi(b) = b$ for each positive integer b. Since φ also respects additive inverses, $\varphi(-b) = -b$ for positive integers b, so φ is the identity map on the integers. Since φ respects multiplicative inverses, $\varphi(1/b) = 1/\varphi(b) = 1/b$ for any integer b, and since φ respects products $\varphi(a/b) = \varphi(a)\varphi(1/b) = a/b$. Thus we have shown that the only automorphism of $\mathbb{Q}$ is the identity map. A similar (shorter argument) shows that the only automorphism of $\mathbb{F}_p$ is the identity map.

Notice also that there can be no homomorphism from $\mathbb{Q}$ to $\mathbb{F}_p$ since any homomorphism must be injective. There can't be a homomorphism from $\mathbb{F}_p$ to $\mathbb{Q}$ since we would have to map $1_{\mathbb{F}_p}$ to $1_{\mathbb{Q}}$, but $\underbrace{1+\cdots+1}_{p \text{ terms}} = 0$ in $\mathbb{F}_p$ while $\underbrace{1+\cdots+1}_{p \text{ terms}} \neq 0$ in $\mathbb{Q}$.

The reals are vastly more complicated, so let's consider automorphims of $\mathbb{C}$ that fix $\mathbb{R}$; so we consider automorphisms φ such that $\varphi(r) = r$ for $r \in \mathbb{R}$. We know that $i * i = -1$ so $\varphi(i) * \varphi(i) = \varphi(-1) = -1$. We know there are only two square roots of 1 in $\mathbb{C}$, so there are only two possibilities: $\varphi(i)$ is either i itself or $-i$. In the first case φ has to be the identity map, $\varphi(a + bi) = \varphi(a) + \varphi(b)\varphi(i) = a + bi$ since φ fixes the reals. In the second case φ is the conjugation map: $\varphi(a + bi) = a - bi$. It is clear that the composition of the conjugation map with itself is the identity map. Thus, $\text{Aut}\left(\mathbb{C}/\mathbb{R}\right) \cong \mathbb{Z}_2$.

A similar argument applies to the field $\mathbb{Q}(i)$. The field $\mathbb{Q}$ has to be fixed, and the only non-identity automorphims takes $a + bi$ to $a - bi$. Thus we have $\text{Aut}\left(\mathbb{Q}(i)\right) \cong \mathbb{Z}_2$.

This simple example is the model for much of our work in this chapter. For a field K containing another field F, we seek to understand the automorphisms of K that fix F, and to use that knowledge to better understand the field K.

5.2 Constructing Fields

We have three main tools for constructing new fields.

Construction I: In Section 1.1 we showed that the ring of integers modulo a prime forms a field, which we write $\mathbb{F}_p$. Similarly, in Section 1.3 we showed that for F a field and $m(x)$ irreducible, $F[x]/m(x)$ is a field (see Theorem 4.1.9). More generally, for any ring R, Theorem 4.6.11 shows that R/I is a field whenever I is a maximal ideal.

Construction II: The second method is based on the construction of the rational numbers from the integers. For an integral domain R, let $D = R \setminus \{0\}$ and form the ring of fractions $D^{-1}R$ as in Section 4.7. This is a field. For the integral domain $F[x]$ over a field F, the resulting field is written $F(x) = \{a(x)/b(x) : a(x), b(x) \in F[x] \text{ with } b(x) \neq 0\}$.

Construction III: The third method is to take a subfield of a given field. We did this in Section 1.2 when we introduced the subfield of the complex numbers

> $\mathbb{Q}(i)$. Given any field K and a subset $S \subseteq K$ we can take the intersection of all subfields of K containing S. Exercise 5.1.3 show that this is a field (letting $\mathcal{F}$ be the set of all fields containing S). It contains S and it is, by construction a subfield of every field containing S. Thus it makes sense to call it the smallest subfield containing S. We often are interested in the smallest subfield of K containing a specific subfield F and some additional set of elements $S \subseteq K \setminus F$. We write this $F(S)$. If no subfield is specificied, we know that this field must contain one of the prime fields $\mathbb{Q}$, or $\mathbb{F}_p$, so we may write it $\mathbb{F}_p(S)$ or $\mathbb{Q}(S)$ as appropriate.

With the notation of Construction III, given a subset S of K we may also take the intersection of all rings containing S and some subfield F, which we write $F[S]$.

There is a relationship between Construction III and the other constructions, which we can illustrate with two examples inside the complex field. Before introducing the two examples recall Theorem 4.3.9, which we adapt here as follows.

Theorem Let F be a subfield of a field K. For any $s \in K$ there is a unique homomorphism from $F[x]$ to K that takes x to s, namely

$$\begin{aligned} \overline{\varphi} : F[x] &\longrightarrow K \\ \sum_i a_i x^i &\longmapsto \sum_i a_i s^i \end{aligned}$$

Definition 5.2.1. Continuing with the notation as stated above, if the homomorphism is injective we say that s is **transcendental** over F. Otherwise s is **algebraic** over F. The monic generator of the kernel in the theorem is called the **minimal polynomial** of s.

Proposition 5.2.2. *If s is algebraic over F then the minimal polynomial of s is an irreducible polynomial. Consequently, the image of φ as defined in the theorem is a subfield of K. Thus $F[s] = F(s)$. The dimension $[F(s) : F]$ is equal to the degree of the minimal polynomial.*

Proof. Suppose by way of contradiction, that the monic generator of the kernel is $m(x)$ and it factors as $m(x) = f(x)g(x) \in F[x]$. Then $f(s)g(s) = 0$ in K. Since K is a field either $f(s)$ or $g(s)$ is zero. Suppose the former. Then $f(x)$ is in the kernel, and is therefore a multiple of $m(x)$. Since $f(x)$ is also a factor of $m(x)$ we must have that $g(x)$ has degree 0, so it is a constant. This shows that $m(x)$ is irreducible.

Alternatively, we can prove the result by noting that $F[x]$ modulo $m(x)$ is isomorphic to its image in K, which must be an integral domain. Since a reducible polynomial yields a quotient ring with zero divisors by Corollary 4.1.10, the kernel must be generated by an irreducible polynomial. Moreover, the quotient of $F[x]$ by an irreducible polynomial is a field, so the image of φ is actually a subfield of K.

Note that the image of φ consists of polynomials in s. Thus the smallest ring containing F and s is also the smallest field containing F and s: $F[s] = F(s)$. We note also that this is $\left\{\sum_{i=0}^{n-1} a_i s^i\right\}$ where $n = \deg(m(x))$, since the polynomials of degree less than n form a system of representatives for $F[x]/m(x)$. Since each element of $F[s]$ is uniquely expressed as a polynomial in s of degree less than $\deg(m(x))$, the degree of the extension $F(s)/F$ is equal to $\deg(m(x))$. □

Proposition 5.2.3. *If K contains some transcendental element over F then $[K : F]$ is infinite. Conversely, if K is finite dimensional over F then every element of K is algebraic over F.*

Proof. Suppose that $\alpha \in K$ is transcendental over F. Then $F[\alpha]$ is isomorphic to $F[x]$ since it is the image of the injective homomorphism $\varphi : F[x] \longrightarrow K$ that takes x to α and fixes F. In $F[x]$ the powers of x, x^i, are linearly independent, so $F[x]$ is infinite dimensional over F. Since $F[\alpha]$ is isomorphic to $F[x]$, it is also infinite dimensional over F. Since K contains $F[\alpha]$ it is infinite dimensional over F. □

Example 5.2.4. Consider the homomorphism $\varphi : \mathbb{Q}[x] \longrightarrow \mathbb{C}$ that takes x to $i = \sqrt{-1}$. This is not injective, the kernel is $x^2 + 1$ and the image is $\mathbb{Q}[i]$, the ring of polynomials in i with rational coefficients. It is isomorphic to $\mathbb{Q}[x]/(x^2 + 1)$. This is a field because $x^2 + 1$ is irreducible. Thus the field $\mathbb{Q}(i) = \mathbb{Q}[i]$.

Similarly, $\mathbb{Q}(\sqrt{2}) \cong \mathbb{Q}[x]/(x^2 - 2)$ and $\mathbb{Q}(\sqrt[3]{2}) \cong \mathbb{Q}[x]/(x^3 - 2)$. The latter example is subtle though. There are three cube roots of 2 in $\mathbb{C}$, the others are $\sqrt[3]{(2)}\omega$ and $\sqrt[3]{(2)}\omega^2$ where $\omega = (1 - \sqrt{3}i)/2$. Define $\varphi : \mathbb{Q}[x] \longrightarrow \mathbb{C}$ by $\varphi(x) = \sqrt[3]{(2)}\omega$. This gives $\mathbb{Q}(\sqrt[3]{2}\omega) \cong \mathbb{Q}[x]/(x^3 - 2)$. We may do the same for $\sqrt[3]{2}\omega^2$, so there are three embeddings of $\mathbb{Q}[x]/(x^3 - 2)$ in $\mathbb{C}$.

Example 5.2.5. Consider the homomorphism $\varphi : \mathbb{Q}[x] \longrightarrow \mathbb{C}$ that takes x to e where e is the Euler number $e \approx 2.71$. It is not obvious, but e is transcendental, not algebraic []. The homomorphism φ is therefore an isomorphism of $\mathbb{Q}[x]$ with its image $\mathbb{Q}[e]$. The smallest field containing $\mathbb{Q}[e]$ is

$$\mathbb{Q}(e) = \left\{\frac{f(e)}{g(e)} : f(e), g(e) \in \mathbb{Q}[e] \text{ and } g(e) \neq 0\right\}$$

The stipulation that $g(e) \neq 0$ is simply requiring that the coeffiecients of g not all be zero, since no nonzero polynomial in $\mathbb{Q}[x]$ evaluates at e to 0.

The number $\pi \approx 3.14$ is also transcendental [] so there is an isomorphism between $\mathbb{Q}[x]$ and $\mathbb{Q}[\pi]$ and between $\mathbb{Q}(x)$ and $\mathbb{Q}(\pi)$.

The set of algebraic numbers (the complex numbers algebraic over $\mathbb{Q}$) is actually countable, while the complex numbers (and therefore the transcendental numbers) are uncountable.[]

Adjoining a root

Our focus henceforth is on Construction I. Let $m(x)$ be a monic irreducible polynomial of degree $n > 1$ over a field F. Let $m(x) = x^n + m_{n-1}x^{n-1} + \cdots + m_1x + m_0$. In the quotient ring $F[x]/m(x)$ the coset $x + \langle m(x)\rangle$ satisfies

$$\begin{aligned}(x + \langle m(x)\rangle)^n &+ m_{n-1}(x + \langle m(x)\rangle)^{n-1} + m_{n-2}(x + \langle m(x)\rangle)^{n-2} + \cdots \\ &\cdots + m_2(x + \langle m(x)\rangle)^2 + m_1(x + \langle m(x)\rangle) + m_0 \\ &= (x^n + m_{n-1}x^{n-1} + m_{n-2}x^{n-2} + \cdots + m_2x^2 + m_1x + m_0) + \langle m(x)\rangle \\ &= 0 + \langle m(x)\rangle\end{aligned}$$

In other words, we may think of $m(x)$ as having a root in $F[x]/m(x)$, namely the coset $x + \langle m(x)\rangle$.

It is common therefore to give this coset a new symbol, let's call it α, and to speak of the quotient ring as follows: We adjoin a root α of $m(x)$ to obtain the field $F(\alpha)$ in which $x - \alpha$ is now a factor of $m(x)$. In the polynomial ring $F(\alpha)[x]$, we can then factor $m(x)$ by dividing $m(x)$ by $x - \alpha$. A natural question is whether $m(x)/(x - \alpha)$ is now irreducible, or does it factor completely (into linear factors), or something in between?

The upcoming sections explore this question. It is clear though that if $\deg(m(x)) = 2$ then adjoining a root of $m(x)$ will factor $m(x)$ completely, since the quotient $m(x)/(x - \alpha)$ will be another linear factor.

More generally, over $\mathbb{Q}$, every extension by a root of a quadratic is isomorphic to $\mathbb{Q}(\sqrt{D})$ for some square free integer D. The exercise below steps through the proof.

Exercises 5.2.6. Let $m(x) = x^2 + ax + b$ be an irreducible quadratic over $\mathbb{Q}$.

(a) Use the quadratic formula to find two distinct embeddings of $\mathbb{Q}[x]/m(x)$ into $\mathbb{C}$.

(b) Show that these two embeddings have the same image (although the image of $x + \langle m(x)\rangle$ itself is different in the two cases.

(c) Show that there is some square free integer D such that $\mathbb{Q}(\sqrt{D})$ is the same field as the one determined by $\mathbb{Q}[x]/m(x)$.

(d) Conclude that every degree 2 extension of $\mathbb{Q}$ is isomorphic to $\mathbb{Q}(\sqrt{D})$ for some square free integer D. [Suppose K is a degree 2 extension of $\mathbb{Q}$ and let $\alpha \in K \setminus \mathbb{Q}$.]

(e) Conclude also that every degree 2 extension of $\mathbb{Q}$ has one non-trivial automorphism.

Finally, we have this relationship between an automorphim of a field extension and the minimal polynomial of an element in the extension.

Proposition 5.2.7. *Let K be an extension of F. Let $\alpha \in K$ have minimum polynomial $m(x)$ over F. For any $\sigma \in \mathrm{Aut}(F/K)$, $\sigma(\alpha)$ is also a root of $m(x)$.*

Proof. Let $m(x) = x^d + m_{d-1}x^{d-1} + \cdots + m_0$ be the mimimum polynomial for α over F. Each $m_i \in F$ so

$$\begin{aligned} \sigma\Big(m(\alpha)\Big) &= \sigma\left(\sum_{i=1}^{n} m_i\alpha\right) \\ &= \sum_{i=1}^{n} \sigma(m_i\alpha) \qquad \text{since } \sigma \text{ respects sums,} \\ &= \sum_{i=1}^{n} m_i\sigma(\alpha) \qquad \text{since } \sigma \text{ respects products and fixes elements of } F \\ &= m\left(\sigma(\alpha)\right) \end{aligned}$$

Since $m(\alpha) = 0$ we have $m(\sigma(\alpha))$ is also 0. □

5.3 Finite Fields

In this section we characterize finite fields by proving the existance and uniqueness of a field of order p^n for any prime p and $n \in \mathbb{N}$, and by revealing the structure of these fields and of their automorphism groups.

Theorem 5.3.1. *Let K be a field with a finite number of elements.*

(1) K has p^n elements for some prime p and $n \in \mathbb{N}$.

(2) Each element of K is a root of $x^{p^n} - x$, so $x^{p^n} - x$ factors completely, into distinct linear factors, over K.

(3) There is an element $\eta \in K$ whose powers $\eta^1, \eta^2, \ldots, \eta^{p^n-1} = 1$ give all the nonzero elements of K. Consequently, K^* is cyclic of order $p^n - 1$.

(4) K is isomorphic to $\mathbb{F}_p[x]/m(x)$ for some irreducible polynomial $m(x)$ of degree n over $\mathbb{F}_p$. Furthermore $m(x)$ is a factor of $x^{p^n} - x$.

For any prime p and any positive integer n:

(4) There exists a field with p^n elements.

(5) Any two fields with p^n elements are isomorphic.

We use $\mathbb{F}_{p^n}$ to denote the unique field with p^n elements. The automorphism group of $\mathbb{F}_{p^n}$ satisfies:

(6) $\mathrm{Aut}(\mathbb{F}_{p^n})$ *is generated by the Frobenius map, $\varphi(\beta) = \beta^p$ for $\beta \in \mathbb{F}_{p^n}$.*

(7) $\mathrm{Aut}(\mathbb{F}_{p^n}) \cong \mathbb{Z}/n$.

As a first step we prove

Proposition 5.3.2. *A finite field is a vector space over $\mathbb{F}_p$ for some prime p. Consequently, the number of elements of K is a power of p.*

Proof. Suppose that K is a finite field. The smallest field contained in K, its prime field, must be $\mathbb{F}_p$ for some prime number p.

From the definition of a field we can see that K satisfies the properties for a vector space over $\mathbb{F}_p$. For example: if $a \in \mathbb{F}_p$ and $\beta, \gamma \in K$ then $a(\beta + \gamma) = a\beta + a\gamma$ follows from the distributive law, but may be also considered as the property concerning scalar multiplication (by α) of a sum of vectors, $\beta + \gamma$. If the dimension of K over $\mathbb{F}_p$ is n then K has a basis $u_1, \ldots, u_n$ and the elements of K are $a_1u_1 + \ldots, a_nu_n$ for $a_i \in \mathbb{F}_p$. Thus K must have p^n elements. □

Suppose that $q = p^n$ is the number of elements in K. By the definition of a field, the set of nonzero elements of K is a group under multiplication. This group is denoted K^*. Recall that the **order** of an element α in a group G is the smallest positive integer r such that α^r is the identity, or infinity, if no such r exists. The order of an element divides the order of the group (Lagrange's theorem).

Proposition 5.3.3. *Let K be a field with p^n elements. The polynomial $x^{p^n} - x$ factors completely, into distinct linear factors, over K.*

$$x^{p^n-1} - 1 = \prod_{\alpha \in K^*} (x - \alpha) \qquad \textit{and,}$$
$$x^{p^n} - x = \prod_{\alpha \in K} (x - \alpha)$$

Proof. The multiplicative group K^* has $p^n - 1$ elements, so each element $\alpha \in K^*$ has order dividing $p^n - 1$. Thus each $\alpha \in K^*$ is a root of $x^{p^n-1} - 1$ and each $\alpha \in K$ is a root of $x^{p^n} - x$. Since roots correspond to factors of a polynomial, we have p^n linear factors of $x^{p^n} - x$, one for each element of K. Their product has degree p^n so must be equal to $x^{p^n} - x$. □

Recall the following properties from Theorem 2.1.11 and problems immediately following it.

Theorem Let α be an element of order r in an group G.

(1) $\alpha^i = \alpha^j$ iff $i \equiv j \mod r$.

(2) The order of α^i is r/d where $d = \gcd(i, r)$.

(3) Let G be abelian. Let $\beta \in G$ have order s, coprime to $r = \text{ord}(\alpha)$. Then $\text{ord}(\alpha\beta) = rs$.

(4) Let G be abelian. If $\alpha_1, \ldots, \alpha_n$ have orders $r_1, \ldots, r_n$ where the r_i are pairwise coprime, then $\text{ord}(\prod_{i=1}^n \alpha_i) = \prod_{i=1}^n r_i$.

Now we can establish item (3) of the Theorem.

Proposition 5.3.4. *The multiplicative group of a finite field is cyclic.*

Proof. Let K have p^n elements and let the prime factorization of $p^n - 1$ be $\prod_{i=1}^r q_i^{b_i}$. We will show below that for each $i = 1 \ldots, r$ there is an element $\eta_i \in K^*$ of order $q_i^{b_i}$. Let $\eta = \prod_{i=1}^r \eta_i$. Since the $q_i^{b_i}$ are coprime to each other, the theorem above shows that the order of η is $\prod_{i=1}^r q_i^{b_i} = p^n - 1$. Thus η generates the multiplicative group of K.

Now we prove the claim. Suppose that q is prime and q^b appears in the prime factorization of $p^n - 1$, so q^b divides $p^n - 1$ but q^{b+1} does not. Let $t = (p^n - 1)/q^b$ and consider the set $S = \{\alpha^t : \alpha \in K^*\}$. For any $\beta \in S$ the polynomial $x^t - \beta$ has at most t roots so there can be at most t elements of K whose tth power is β. Therefore the cardinality of S is at least $(p^n - 1)/t = q^b$. On the other hand, everything in S is a root of $x^{q^b} - 1$ since

$$(\alpha^t)^{q^b} = \alpha^{p^n - 1} = 1$$

There can be only q^b roots of $x^{q^b} - 1$, so S has at most q^b elements. This shows $|S| = q^b$. Similarly, at most q^{b-1} of the elements in S can be roots of $x^{q^{b-1}} - 1$ so there must be at least $q^b - q^{b-1}$ elements of S whose order in K is q^b. This shows what we wanted: there is some element of K of order q^b.

□

Definition 5.3.5. An element of a finite field whose powers generate the nonzero elements of the field is called **primitive**.

Proposition 5.3.4 says that every finite field has a primitive element. Furthermore, from the lemma, if η is primitive in a field of p^n elements then η^k is also primitive whenever k is coprime to $p^n - 1$. Thus there are $\varphi(p^n - 1)$ primitive elements, where φ is the Euler totient function ($\varphi(n)$ is the number of positive integers less than n and coprime to n).

To prove item (4) of the Theorem we need to use the minimal polynomial of a primitive element.

Proposition 5.3.6. *Let K be a finite field of p^n elements. Let η be any primitive element of K, let $\varphi : \mathbb{F}_p[x] \longrightarrow K$ take x to η and let $m(x)$ generate the kernel (so $m(x)$ is the minimal polynomial of η over $\mathbb{F}_p$). Then K is isomorphic to $\mathbb{F}_p[x]/m(x)$ and* $\deg m(x) = n$. *Furthermore, $m(x)$ divides $x^{p^n} - x$ and $m(x)$ factors completely in K.*

Proof. From the first isomorphism theorem, φ gives rise to an isomorphism from $\mathbb{F}_p[x]/m(x)$ to its image in K. This image must be a field since $m(x)$ must be irreducible. But the image of φ contains η and therefore all of its powers. Thus the image is all of K and we have $K \cong \mathbb{F}_p[x]/m(x)$. The dimension of K is n and the dimension of $\mathbb{F}_p[x]/m(x)$ is $\deg(m(x))$ so the degree of $m(x)$ is n.

By Proposition 5.3.3, $x^{p^n} - x$ factors into linear factors in K and η is one of the roots. This implies that $x^{p^n} - x$ is in the kernel of φ, so $m(x)$ divides $x^{p^n} - x$. Since $x^{p^n} - x$ factors completely in K so to does $m(x)$. □

We can now prove existence and uniqueness for fields of prime power order. We will need the "Freshman's dream":

Proposition 5.3.7. *Let α, β be elements of a field of characteristic p. Then $(\alpha+\beta)^p = \alpha^p + \beta^p$.*

Proof. Expanding $(\alpha+\beta)^p$ using the binomial theorem we get terms like

$$\binom{p}{k}\alpha^k\beta^{p-k}$$

The binomial coefficient really means 1 added to itself $\binom{p}{k}$ times. Since p divides the binomial coefficient when $1 < k < p$ the coefficient is 0 unless $k = 0$ or $k = p$. That gives the result. □

Proposition 5.3.8. *For any prime power there exists a unique field of that order.*

Proof. Uniqueness: Let K and K' be two fields with p^n elements. Let η be a primitive element in K and let $m(x)$ be its minimal polynomial over $\mathbb{F}_p$. The previous proposition showed that η is a root of $x^{p^n} - x$, and $m(x)$ divides $x^{p^n} - x$. By Proposition 5.3.6, $x^{p^n} - x$ factors into distinct linear factors in both K and K' so there must be a root of $m(x)$ in K'. Call this root η'. Then the homomorphism from $\mathbb{F}_p[x]$ to K' that takes x to η' must have image that is a subfield of dimension n in K', and is therefore all of K'. By Proposition 5.3.6, both K and K' are isomorphic to $\mathbb{F}_p[x]/m(x)$ so they are isomorphic to each other.

Existence: By successively factoring $x^{p^n} - x$ and adjoining roots of a nonlinear irreducible factor, we can, after a finite number of steps, arrive at a field in which $x^{p^n} - x$ factors completely. I show below that the roots of $x^{p^n} - x$ form a field. Since the derivative of $x^{p^n} - x$ is -1, $x^{p^n} - x$ does not have multiple roots, so by the roots-factors theorem it has exactly p^n roots. Thus we have a field of p^n elements.

To show the roots of $x^{p^n} - x$ form a field, we need to show that the sum of two roots is a root, that the additive inverse of a root is a root, that the product of two roots is a root and that the multiplicative inverse of a root is a root. These are all trivial except for the case of the sum of two roots, which can be proved using the "Freshman's dream." □

Definition 5.3.9. Let K be a finite field and let $p(x)$ be a polynomial over K. If $p(x)$ is irreducible and the class of x is primitive in $K[x]/p(x)$, then we say $p(x)$ is a **primitive polynomial**.

Example 5.3.10. There is a unique irreducible polynomial of degree 2 over $\mathbb{F}_2$, x^2+x+1. Let's use it to construct the field $\mathbb{F}_4$ as $\mathbb{F}_2[x]/(x^2+x+1)$. Let η be the congruence class $x + \langle x^2+x+1\rangle$, then η is a root of x^2+x+1. We say that we have adjoined η to $\mathbb{F}_2$ to create the field $\mathbb{F}_4$. The elements of $\mathbb{F}_4$ are polynomials in

η of degree less than 2: $\mathbb{F}_4 = \{0, 1, \eta, \eta + 1\}$. Addition is component-wise (relative to the basis $\{1, \eta\}$). Multiplication must take account of $\eta^2 = \eta + 1$. Here is a multiplication table for this field.

$*$	1	η	$\eta+1$
1	1	η	$\eta+1$
η	η	$\eta+1$	1
$\eta+1$	$\eta+1$	1	η

We can also see that η is a primitive element. Since $\mathbb{F}_4^*$ has 3 elements, any element besides 1 generates $\mathbb{F}_4^*$. An alternative to compute in this field is by making a table of powers of η and the corresponding polynomials in η as follows.

exponential form	polynomial form
1	1
η	η
η^2	$\eta+1$

The following example shows that there can be many ways to construct a given field.

Example 5.3.11. Let $p = 2$. We can construct the field $\mathbb{F}_{2^3}$ by adjoining to $\mathbb{F}_2$ a root η of the irreducible polynomial $m(x) = x^3 + x + 1$. Since the degree of $m(x)$ is 3, the elements of the field will be polynomials of degree less than 3 in η.

Here is the "dictionary" between powers of η and corresponding polynomials in η. We use $\eta^3 = \eta + 1$ to compute successive rows in the table.

exponential form	polynomial form
1	1
η	η
η^2	η^2
η^3	$\eta+1$
η^4	$\eta^2+\eta$
η^5	$\eta^2+\eta+1$
η^6	η^2+1

The next line in the table would be

$$\eta^7 = \eta^3 + \eta = \eta + 1 + \eta = 1$$

This is to be expected, since $\mathbb{F}_8^*$ is a cyclic group of order 7. Note that every element of $\mathbb{F}_8^*$ except 1 is primitive, since 7 is prime.

We can use this table to verify that η^2 is another root of $x^3 + x + 1$

$$\begin{aligned}(\eta^2)^3 + \eta^2 + 1 &= \eta^6 + \eta^2 + 1 \\ &= 0\end{aligned}$$

Similarly η^4 is also a root.

$$\begin{aligned}(\eta^4)^3 + \eta^4 + 1 &= \eta^{12} + \eta^4 + 1 \\ &= \eta^5 + \eta^4 + 1 \\ &= (\eta^2 + \eta + 1) + (\eta^2 + \eta) + 1 \\ &= 0\end{aligned}$$

The Finite Field Theorem says that there is a unique field of order 8 so every irreducible polynomial of degree 3 over $\mathbb{F}_2$ must have roots in the field that we constructed. There are 2 monic irreducible polynomials of degree 3 over $\mathbb{F}_2$: $x^3 + x + 1$, which we used to construct this field, and $x^3 + x^2 + 1$.

Similar computations show that η^3, η^5 and η^7 are the roots of $x^3 + x^2 + 1$.

Here is a multiplication table (omitting 0) for the representation of $\mathbb{F}_8$ using $x^3 + x + 1$.

$*$	1	η	$\eta + 1$	η^2	$\eta^2 + 1$	$\eta^2 + \eta$	$\eta^2 + \eta + 1$
1	1	η	$\eta + 1$	η^2	$\eta^2 + 1$	$\eta^2 + \eta$	$\eta^2 + \eta + 1$
η	η	η^2	$\eta^2 + \eta$	$\eta + 1$	1	$\eta^2 + \eta + 1$	$\eta^2 + 1$
$\eta + 1$	$\eta + 1$	$\eta^2 + \eta$	$\eta^2 + 1$	$\eta^2 + \eta + 1$	η^2	1	η
η^2	η^2	$\eta + 1$	$\eta^2 + \eta + 1$	$\eta^2 + \eta$	η	$\eta^2 + 1$	1
$\eta^2 + 1$	$\eta^2 + 1$	1	η^2	η	$\eta^2 + \eta + 1$	$\eta + 1$	$\eta^2 + \eta$
$\eta^2 + \eta$	$\eta^2 + \eta$	$\eta^2 + \eta + 1$	1	$\eta^2 + 1$	$\eta + 1$	η	η^2
$\eta^2 + \eta + 1$	$\eta^2 + \eta + 1$	$\eta^2 + 1$	η	1	$\eta^2 + \eta$	η^2	$\eta + 1$

Now we consider the automorphism group of a finite field. Recall that any automorphism has to take 1 to itself, and must therefore fix the subfield $\mathbb{F}_p$.

Proposition 5.3.12. *The automorphism group of $\mathbb{F}_{p^n}$ is cyclic of order n, generated by the Frobenius map $\varphi : \alpha \longmapsto \alpha^p$.*

Proof. The Frobenius map respects addition, by the Freshman's dream, and it clearly respects multiplication: $\varphi(\alpha\beta) = (\alpha\beta)^p = \alpha^p\beta^p = \varphi(\alpha)\varphi(\beta)$. Thus φ is a homomorphism of fields. Since a homomorphism of fields must be injective, and since an injective function on a finite set is also surjective, we conclude that φ is an automorphism.

Repeatedly composing the Frobenius with itself gives other automorphims and one can inductively establish the formula: $\varphi^t(\alpha) = \alpha^{p^t}$. Since $\mathbb{F}_{p^n}^*$ has order $p^n - 1$ we have for $\alpha \neq 0$, $\varphi^n(\alpha) = \alpha^{p^n} = \alpha^{p^n-1} * a = 1 * \alpha = \alpha$.

Thus φ^n is the identity map. I claim no lower power of φ is the identity map. Suppose that φ^r is the identity automorphism and let η be primitive in $\mathbb{F}_{p^n}$. Then $\eta = \varphi^r(\eta) = \eta^{p^r}$, so $\eta^{p^r-1} = 1$. Since η is primitive it has order $p^n - 1$, so we see $r \geq n$ as claimed.

We need to show that there are no other automorphisms of $\mathbb{F}_{p^n}$. Let η be primitive, and let $m(x) = x^n + m_{n-1}x^{n-1} + \cdots + m_0$ be its mimimum polynomial. The lemma showed that $\varphi^r(\eta) = \eta^{p^r}$ is another root of $m(x)$. Since η is primitive, $\eta, \ldots, \eta^{p^{n-1}}$ are all distinct and thus they form the complete set of roots of $m(x)$.

Let σ be an arbitrary automorphim of $\mathbb{F}_{p^n}$. Then σ must take η to one of these other roots of $m(x)$. The action of σ on η determines σ completely, so if $\sigma(\eta) = \eta^{p^r}$ then $\sigma = \varphi^r$.

In conclusion $\mathrm{Aut}(\mathbb{F}_{p^n})$ is cyclic of order n, and is generated by φ. □

Example 5.3.13. Consider now the degree 3 extension of $\mathbb{F}_3$, the field with 27 elements, $\mathbb{F}_{27}$. In this field there is just one subfield $\mathbb{F}_3$, so there are 24 elements that have a minimal polynomial of degree 3. Each of these minimal polynomials factors completely in $\mathbb{F}_{27}$ by Proposition5.3.6. Thus we have $24/3 = 8$ monic irreducible polynomials of degree 3 over $\mathbb{F}_3$.

Another way to count the number of irreducible monic polynomials of degree 3 over $\mathbb{F}_3$ is to count the number of monic reducible polynomials and subtract that from the total number of monic polynomials. There are 27 monic polynomials of degree 3, since we choose 3 coefficients from $\mathbb{F}_3$. A reducible polynomial is either the product of 3 linear factors or the product of a linear and a quadratic irreducible. There are 3 monic linear polynomials, and 3 monic quadratic irreducibles, so 9 possible products. For a product of linear monic polynomials we choose 3 factors with replacement from the 3 linear polynomials, so there are $\binom{3+2}{3} = 10$ possibilites. Thus the number of irreducibles shoule be $27 - 9 - 10 = 8$. That checks with our computation from the previous paragraph.

The multiplicative group $\mathbb{F}_{27}^*$ is cyclic of order 26. In $\mathbb{Z}/26$ the odd numbers, other than 13, are all generators for the group so there are 12 generators. Consequently in $\mathbb{F}_{27}^*$ there are 12 primitive elements. Each is a root of one of the monic irreducible polynomials of degree 3, so we expect $12/3 = 4$ different primitive monic polynomials of degree 3.

Here are the monic irreducibles of degree 3 over $\mathbb{F}_3$ that are not primitive

$$
\begin{aligned}
&x^3 + 2x^2 + 2x + 2,\\
&x^3 + x^2 + x + 2\\
&x^3 + 2x + 2\\
&x^3 + x^2 + 2,
\end{aligned}
$$

and here are the ones that are primitive.

$$
\begin{aligned}
&x^3 + 2x + 1\\
&x^3 + 2x^2 + x + 1\\
&x^3 + x^2 + 2x + 1\\
&x^3 + 2x^2 + 1
\end{aligned}
$$

5.4 Problems

Exercises 5.4.1. Irreducible polynomials over $\mathbb{F}_2$

(a) (Discussion) Find all irreducible polynomials over $\mathbb{F}_2$ of degree at most 4. You should justify your list.

(b) (HW) Find all all irreducible polynomials over $\mathbb{F}_2$ of degree 5. Use the list from part (a) to explain your result. Notice any patterns in the list of polynomials.

(c) (HW) Determine how many irreducible polynomials of degree 6 there are over $\mathbb{F}_2$ based on part (a). Justify your answer briefly.

Exercises 5.4.2. (Discussion) The field $\mathbb{F}_{2^3}$.

(a) Construct the field $\mathbb{F}_8$ using one of the polynomials from the previous problem that has the appropriate degree (there are two). Make a table showing the powers of the primitive element, call it η, and the corresponding vector form, using the basis $\{1, \eta, \eta^2\}$.

(b) Show that the polynomial that you did not choose also has roots in $\mathbb{F}_8$.

Exercises 5.4.3. (HW)

(a) One of the irreducible polynomials of degree 4 in 11.1(a) has roots which are not primitive. Which one?

(b) Construct the field with 16 elements using one of the primitive irreducible polynomials of degree 4: Make a table showing the powers of the primitive element, call it η, and the corresponding vector form, using the basis $\{1, \eta, \eta^2, \eta^3\}$. Give also the multiplicative order of each element and its minimal polynomial.

(c) Identify the subfield $\mathbb{F}_4$.

(d) Factor over $\mathbb{F}_4$ the irreducible polynomial that you chose to construct $\mathbb{F}_{16}$.

(e) How many elements of $\mathbb{F}_{16}$ are primitive?

Exercises 5.4.4. (HW) The field $\mathbb{F}_{3^2}$.

(a) Write a multiplication table for $\mathbb{F}_3[x]/\langle x^2+x+2\rangle$. [You may omit 0. It may be easier to take the elements in the order $1, x, x+1, x+2$ followed by twice each.]

(b) Find all irreducible polynomials of degree 2 over $\mathbb{F}_3$ and find their roots in the table you constructed.

Exercises 5.4.5. (HW) Factoring a polynomial over different fields.

(a) Factor $x^9 - x$ over $\mathbb{F}_3$.

(b) Factor $x^5 + x^4 + 1$ over $\mathbb{F}_2$ (it is reducible!), $\mathbb{F}_4$ and $\mathbb{F}_8$.

(c) Factor $x^{16} - x$ over $\mathbb{F}_2$, $\mathbb{F}_4$, $\mathbb{F}_8$ and $\mathbb{F}_{16}$.

Exercises 5.4.6. Let $n > m$ be positive integers and $d = \gcd(n, m)$. Show that the intersection of $\mathbb{F}_{p^m}$ and $\mathbb{F}_{p^n}$ is $\mathbb{F}_{p^d}$ as follows.

(a) Recall that the remainder $x^n - 1$ divided by $x^m - 1$ is $x^r - 1$ where r is the remainder when n is divided by m.

(b) Show that the gcd of $x^n - 1$ and $x^m - 1$ is $x^d - 1$.

(c) (HW) Combine the previous results and the theorem that the roots of $x^{p^n} - x$ are the elements of $\mathbb{F}_{p^n}$ to conclude that $\mathbb{F}_{p^d}$ is a subfield of $\mathbb{F}_{p^n}$ iff d divides n. (Strictly speaking $\mathbb{F}_{p^n}$ has a subfield isomorphic to $\mathbb{F}_{p^d}$. See 11.10.)

Exercises 5.4.7. (Optional)The field $\mathbb{F}_{81}$.

(a) (Discussion) The polynomials x^2+x+2 and x^2+2x+2 are both irreducible over $\mathbb{F}_3$. Can you construct $\mathbb{F}_{81}$ by using one of these polynomials and then the other?

(b) (Optional) In a computer algebra system use $m(x) = x^4 + x + 2$ and $r(x) = x^4 + 2x + 2$ to construct two versions of $\mathbb{F}_{81}$. Using a brute force search, find a root of $m(x)$ in the second field and a root of $r(x)$ in the first field. These give isomorphisms between the two fields. Check by hand that each composition is an automorphism of the appropriate version of $\mathbb{F}_{81}$.

(c) (Optional) Factor $x^{80} - 1$ over $\mathbb{F}_3$. For each irreducible factor $a(x)$, find the roots of $a(x)$ in $\mathbb{F}_3[x]/m(x)$.

Exercises 5.4.8. (Optional)The field of 64 elements.

(a) The polynomials $m(x) = x^6 + x + 1$ and $r(x) = x^6 + x^5 + x^4 + x + 1$ are both irreducible over $\mathbb{F}_2$. Using a computer algebra system construct two versions of $\mathbb{F}_{64}$, using $m(x)$ for one and $r(x)$ for the other. Using a brute force search, find a root of $m(x)$ in the second field and a root of $r(x)$ in the first field. These give isomorphisms between the two fields. Check by hand that each composition of the two isomorphisms is an automorphism of the appropriate version of the field.

- **(b)** Factor $x^{63}-1$ over $\mathbb{F}_2$. For each irreducible factor $a(x)$, find the roots of $a(x)$ in $\mathbb{F}_2[x]/m(x)$. Use Sage, but also use your understanding of the theory.
- **(c)** The field $\mathbb{F}_{64}$ can also be constructed as an extension of $\mathbb{F}_4$. Construct $\mathbb{F}_4$, then factor $x^{63}-1$ in $\mathbb{F}_4[x]$. Choose one of the factors of degree 3 to construct $\mathbb{F}_{64}$.
- **(d)** Now create $\mathbb{F}_8$ using an irreducible polynomial of degree 3 over $\mathbb{F}_2$, then factor $x^{63}-1$, then create $\mathbb{F}_{64}$ using an irreducible polynomial of degree 2 in $\mathbb{F}_8[x]$.

Exercises 5.4.9. (Challenge) The number of irreducible polynomials.

- **(a)** Let $a(n)$ denote the number of degree-n irreducible polynomials over $\mathbb{F}_p$ for p prim4. Prove that
$$p^n = \sum_{d|n} d{\cdot}a(d).$$
Hint: use the result about the factorization of $x^{p^d}-x$ factors over $\mathbb{F}_{p^d}$ and about subfields of $\mathbb{F}_{p^n}$.
- **(b)** Prove that
$$\lim_{n\to\infty} \frac{a(n)}{p^n} = 0,$$
meaning that irreducible polynomials are "sparse" in $\mathbb{F}_p[x]$.

Exercises 5.4.10. (Challenge) The algebraic closure of $\mathbb{F}_p$. This problem extends 11.6, which showed that we may consider $\mathbb{F}_{p^d}$ as contained in $\mathbb{F}_{p^n}$ if and only if $d|n$.

- **(a)** Let $\overline{\mathbb{F}}_p = \bigcup_{t\geq 1} \mathbb{F}_{p^t}$. Prove that F is a field.
- **(b)** Prove that $\overline{\mathbb{F}}_p$ is algebraically closed.
- **(c)** Prove that every element of $\overline{\mathbb{F}}_p$ is algebraic over $\mathbb{F}_p$ so there is no algebraically closed field properly contained in $\overline{\mathbb{F}}_p$.
- **(d)** Conclude that $\overline{\mathbb{F}}_p$ is the algebraic closure of $\mathbb{F}_{p_n}$ for any n.

Exercises 5.4.11.

- **(a)** Make a table showing the possible multiplicative orders and the number of elements of each order for $\mathbb{F}_{64}$, $\mathbb{F}_{128}$, and $\mathbb{F}_{256}$. Relate this information to subfields (refer to the previous problem).

Exercises 5.4.12. Irreducible polynomials over $\mathbb{F}_p$. Suppose you have formulas for the number of irreducible monic polynomials of degree m over $\mathbb{F}_p$ for each $m < n$. Using some combinatorial arguments you can then compute the number of monic reducible polynomials of degree n. Subtracting this from the number of monic polynomials of degree n yields the number of monic irreducible polynomials of degree n.

(a) Show that the number of monic irreducible quadratics over $\mathbb{F}_p$ is $(p^2 - p)/2$.

(b) Show that the number of monic irreducible cubics over $\mathbb{F}_p$ is $(p^3 - p)/3$.

(c) You might want to guess at a general formula. A different counting method yields the result more easily than the one above. Try this if you want, noting:

- For $a \in \mathbb{F}_{p^n}$, a is in no proper subfield iff the minimal polynomial for a has degree n.
- Each monic irreducible of degree n has n distinct roots in $\mathbb{F}_{p^n}$.

Exercises 5.4.13. For a given prime p, let $I(d)$ be the set of irreducible polynomials of degree d over $\mathbb{F}_p$.

(a) Show that for $n > 0$,

$$\prod_{d|n} \prod_{f \in I(d)} f = x^{p^n} - x$$

(b) Show that for any $\alpha \in \mathbb{F}_q$,

$$1 + \alpha + \alpha^2 + \alpha^3 + \cdots + \alpha^{q-2} = \begin{cases} 1 \text{ if } \alpha = 0 \\ -1 \text{ if } \alpha = 1 \\ 0 \text{ otherwise} \end{cases}$$

Exercises 5.4.14. Simplifying the task of finding irreducibles.

(a) Let $a \in \mathbb{F}_q^*$. Show that there is an automophism m_a of $\mathbb{F}_q[x]$ that fixes elements of $\mathbb{F}_q$ and takes x to ax.

(b) Argue that $f(x)$ is irreducible if and only if $m_a(f(x))$ is.

(c) Show how this may be used to simplify the search for irreducible polynomials of degree d to those of the form $x^d + x^{d-1} + \cdots$ or those of the form $x^d + 0x^{d-1} + \cdots$

(d) In the second case (the coefficient of x^{d-1} is 0) how can the simplifying technique be extended?

(e) Show that $f(x)$ is primitive if and only if $m_a(f(x))$ is primitive.

(f) Apply these result to small fields to find all irreducible polynomials of degree 2, 3 or 4 over $\mathbb{F}_q$. Then find the primitive polynomials.

Chapter 6

On Beyond $\mathbb{Z}$ and $k[x]$

6.1 The Integers, $k[x]$: Big Questions

This section elaborates on other analogous results between $\mathbb{Z}$ and $k[x]$ for k a field. We include some major conjectures about the integers.

The analogies between $\mathbb{Z}$ and $k[x]$

The fundamental theorems that we have seen for the integers and $k[x]$ are the Quotient-Remainder Theorem, the Greatest Common Divisor Theorem, the Euclidean Algorithm, the Prime-Irreducible Theorem (Euclid's lemma), and Unique Factorization (the Fundamental Theorem of Arithmetic). We have also seen the parallels between modular arithmetic and $k[x]/m(x)$. These parallels extend to comparison of the rational numbers and the field of rational polynomials, $k(x)$ and, beyond that, to number fields and function fields of curves.

Definition 6.1.1. Let $D = \mathbb{Z} \setminus \{0\}$. The **rational numbers** are $\mathbb{Q} = D^{-1}\mathbb{Z}$. A **number field** is a finite extension of the rational numbers. Each number field K is defined (although not uniquely) by an irreducible polynomial $m(x) \in \mathbb{Q}[x]$. $K = \mathbb{Q}[x]/m(x)$.

Definition 6.1.2. Let k be a field and let $D = k[x] \setminus \{0\}$. The field of **rational polynomials** is $k(x) = D^{-1}(k[x])$. An finite field extension of $k(x)$ is a **function field**.

For example, a polynomial in two variables, $f(x, y)$ may be considered an element of $k[x][y]$ let us suppose

$$f(x, y) = f_n(x)y^n + f_{n-1}(x)y^{n-1} + \cdots + f_1(x)y + f_0(x)$$

Dividing by the coefficient of the highest degree term in y we get an element of $k(x)[y]$ that is monic, as a polynomial in y.

$$\frac{f(x, y)}{f_n(x)} = y^n + \frac{f_{n-1}(x)}{f_n(x)}y^{n-1} + \cdots + \frac{f_1(x)}{f_n(x)}y + \frac{f_0(x)}{f_n(x)}$$

Let us assume this polynomial is irreducible. The original polynomial $f(x, y)$ defines a **curve**: the set of ordered pairs $(a, b) \in k^2$ such that $f(a, b) = 0$. We call the field

$$K = k(x)[y]/f(x, y)$$

The function field of $f(x, y)$. Note that computing modulo $f(x, y)$ and modulo $f(x, y)/f_n(x)$ gives the same result since they differ by a unit in $k(x)[y]$.

Example 6.1.3. Let $m(y) = y^3 - (x^3 - x)$. Trust me, this is irreducible (you can use Eisenstein's criteria as we will see later). The function field from $m(y)$ is

$$K = k(x)[y]/(y^3 - (x^3 - x))$$

Each element of K is uniquely represented as a polynomial in y of degree at most 2 with coefficients from $k(x)$.

The analogy between number fields and function fields of curves is particularly strong for the specific case when the base field is finite.

The Riemann hypothesis and Weil conjectures

The Riemann zeta function is a very important tool in analytic number theory that is related to the distribution of prime numbers. More specifically, through various transformations it gives information about the number of primes less than a given integer.

The Riemann zeta function ζ is defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

With some manupulation of infinite series we can write the zeta function as a product

$$\begin{aligned} \zeta(s) &= \prod_{p \text{ prime}} (1 + p^{-s} + p^{-2s} + p^{-3s} + \cdots) \\ &= \prod_{p \text{ prime}} \frac{1}{1 - p^{-s}} \end{aligned}$$

The Riemann zeta function can be extended to the complex plane through analytic continuation, and it has poles at the even negative integers. The **Riemann hypothesis** states that the only other poles are at complex numbers s with $\mathrm{Re}(s) = 1/2$. If it is true the result would have implications for the distribution of primes, but they are not easy to spell out succinctly.

The analytic continuation of ζ, the relationship to the distribution of primes, and other related topics are well beyond what we will study, but the interesting thing is that there are analogous zeta functions for $\mathbb{F}_q[x]$. In fact, number theorists consider not only the zeta function above but the generalized zeta functions for number fields. The Weil conjectures are the geometric analogue. They deal with zeta functions for curves (and projective varieties of higher dimension) over finite fields.

The rather amazing thing is that the Weil conjectures, including the analogue of the Riemann hypothesis, have been proven, while the original Riemann hypothesis has not been proven.

The *abc* conjecture

The *abc* conjecture, formulated in 1985 by Joseph Oesterlé and David Masser, is relatively simple to state. From a very high level one can say that it relates the sum of two numbers with the factorizations of the two numbers and of their sum.

Definition 6.1.4. Let n be an integer with unique factorization $n = up_1^{e_1}p_2^{e_2}\cdots p_t^{e_t}$. The **radical** of n is the square free positive integer $\mathrm{rad}(n) = p_1p_2\cdots p_t$.

Conjecture 6.1.5 (abc)**.** *Let $\epsilon > 0$. There exist only finitely many triples of coprime positive integers a, b, c with $c = a + b$ such that*

$$c > \mathrm{rad}(abc)^{1+\epsilon}$$

One way to gain intuition about this conjecture is to take the extreme case when a and b are large powers of a small prime (or a small number of small primes).

Example 6.1.6. Suppose $a = 2^s$ and $b = 3^t$. Let $c = 2^s + 3^t$ and factor $c = p_1^{e_1}p_2^{e_2}\cdots p_k^{e^k}$ with each $e_i \geq 1$. Note that none of the p_i is equal to 2 or 3.

The conjecture states that except for a finite number of cases

$$2^s + 3^t < \left(2 \cdot 3 \cdot \mathrm{rad}(2^s + 3^t)\right)^{1+\epsilon}$$

Let's take $\epsilon = 1$ just to get a sense for what this inequality says (the conjecture is for arbitrary $\epsilon > 0$, but 1 is adequate for this example.) Simplifying the right hand side and substituting the factorization for $c = 2^s + 3^t$,

$$p_1^{e_1}p_2^{e_2}\cdots p_k^{e_k} < 36(p_1p_2\cdots p_k)^2$$

Consequently,

$$p_1^{e_1-2}p_2^{e_2-2}\cdots p_k^{e_k-2} < 36$$

One easy conclusion to draw is that there can only be a finite number of cases in which $2^s + 3^t = p_1^{e_1}p_2^{e_2}\cdots p_k^{e_k}$ in which $e_i \geq 3$ for all i.

There is an analogous statement for the polynomial ring $k[x]$, but it has been proven, so it is a theorem not a conjecture. In place of an inequality about the size of integers we have an inequality about the degree of polynomials. The **radical** is defined in analogy to the radical of an integer.

$$\operatorname{rad}\Big(u\big(p_1(x)\big)^{e_1}\big(p_2(x)\big)^{e_2}\cdots\big(p_t(x)\big)^{e_t}\Big) = p_1(x)p_2(x)\cdots p_t(x)$$

Theorem 6.1.7 (Mason-Stothers)**.** *Let $a(x)$, $b(x)$, and $c(x)$ be coprime polynomials over a field k such that $a(x)+b(x)=c(x)$ and such that not all of them have vanishing derivative. Then*

$$\max\{\deg(a(x)), \deg(b(x)), \deg(c(x))\} \leq \deg\Big(\operatorname{rad}\big(a(x)b(x)c(x)\big)\Big) - 1.$$

The theorem is not difficult to prove, so we will do it later in the course.

Fermat's last theorem

Fermat's last theorem became an actual theorem in 1995, many years after it was proposed in 1637.

Theorem 6.1.8 (Fermat-Wiles)**.** *For an integer $n \geq 3$. There is no triple of positive integers a, b, c such that $a^n + b^n = c^n$.*

The proof of this result definitely does not fit in the margin of a piece of paper. Wiles used sophisticated algebra, number theory and algebraic geometry to prove it.

There is an analogue for polynomials over a field.

Theorem 6.1.9. *Let k be a field of characteristic 0 and let $n \geq 3$ be an integer. Suppose that $a(x)$, $b(x)$, and $c(x)$ are three coprime polynomials in $k(x)$ satisfying $a(x)^n + b(x)^n = c(x)^n$. Then all three polynomials are constants.*

The result can be extended to the case of finite characteristic, but that requires a subtle additional restriction. The proof of the result for polynomials will be one of your assigned problems.

Why?

We have just listed three conjectures for the integers (well, one, Fermat's last theorem, is now proven) that have analogues for polynomials that are proven. Two of the polynomial theorems are within our ability to prove. What makes the polynomial versions easier than the integer versions?

Exercises 6.1.10. Let R be an integral domain. We would like to formulate a Quotient-Remainder Theorem and a Euclidean Algorithm for $R[x]$.

(a) Show that the Quotient-Remainder theorem holds with no change, provided the divisor is monic.

(b) In $\mathbb{Z}[x]$ let $f(x) = 3x^3 + 5x^2 + x + 6$ and $g(x) = 2x + 5$. Show that there is a integer m, a $Q(x) \in \mathbb{Z}[x]$ and an $R \in \mathbb{Z}$ such that $mf(x) = g(x)Q(x) + R$. What is special about m? What kind of uniqueness statement can you make?

(c) Do the same question as in (b) but with $g(x) = 3x^2 + 2x + 1$. Now the remainder will be a polynomial, $R(x)$, of degree ≤ 1.

(d) With these exercises as motivation, can you formulate a QR theorem for an arbitrary $R[x]$ with R an integral domain?

6.2 Geometry and Polynomials

Let k be a field. One important difference between polynomials and integers is that each polynomial in $k[x]$ can be considered as a function: it determines a function from k to k. Let $f(x) \in k[x]$ with $f(x) = f_n x^n + f_{n-1}x^{n-1} + \cdots + f_1 x + f_0$. The function defined by $f(x)$ is

$$\begin{aligned} f : k &\longrightarrow k \\ a &\longmapsto f(a) = f_n a^n + f_{n-1}a^{n-1} + \cdots + f_1 a + f_0 \end{aligned}$$

We can think about this geometrically by "graphing" the function in k^2. (For $k = \mathbb{R}$ this is the usual notion of graphing, but for $\mathbb{C}$ or finite fields one has to be more schematic.)

We explore and expand on this geometry in this section. We also make use of the universal property of polynomial rings, Theorem 4.3.9, which we reiterate here (see also the beginning of Section **??**).

Theorem [Universal Property of $k[x]$] Let k be contained in some ring R. For any $a \in R$ there is a unique homomorphism from $k[x]$ to R that takes x to a, namely

$$\begin{aligned} \varphi_a : k[x] &\longrightarrow R \\ \sum_i f_i x^i &\longmapsto \sum_i f_i a^i \end{aligned}$$

This map will be called the **evaluation at** a homomorphism.

It is important to observe the distinction between the two functions we have just introduced. The first: given an $f(x) \in k[x]$ we have a function that takes k

to k that evaluates a *fixed* $f(x)$ at *all* elements of k. It is not a homomorphism (except for a very special case). The second: given $a \in k$ there is a homomorphism from $k[x]$ to k. It evaluates *all* $f(x)$ at a *fixed* $a \in k$.

The Roots-Factors Theorem

Let's now think of a fixed $f(x)$ and the graph of $f(x)$ in k^2. That is, in the x-y plane the points $(x, f(x))$ form the graph of $f(x)$. The intersection of this graph with the line $y = 0$ is the set of roots of $f(x)$.

Definition 6.2.1. Let $f(x) \in k[x]$. An element $a \in k$ is a **root** of $f(x)$ when $f(a) = 0$.

Theorem 6.2.2. *Let $f(x) \in k[x]$. Then a is a root of $f(x)$ if and only if $(x-a)$ is a factor of $f(x)$. Consequently, if $f(x)$ has degree n then f has at most n distinct roots.*

Proof. From the Quotient-Remainder theorem $f(x) = (x-a)q(x)+r(x)$ for unique polynomials $q(x)$ and $r(x)$. We also have $\deg(r(x)) < \deg(x-a)$ so $r(x)$ is actually a constant, $r \in k$. Now apply the "evaluate at a" homomorphism. We get

$$f(a) = (a-a)q(a) + r = r$$

In particular, a is a root of $f(x)$ if and only if $f(a) = 0$, which is true if and only if $r = 0$.

If $f(x)$ has t distinct roots, $a_1, \ldots, a_t$, then it has t distinct factors $(x - a_i)$ for $i = 1, \ldots, t$. Then, by unique factorization, $(x-a_1)(x-a_2)\cdots(x-a_t)$ divides $f(x)$. Thus degree $f(x) \geq t$. □

Definition 6.2.3. We say that a is a **root of multiplicity** t for $f(x) \in k[x]$ when $(x-a)^t$ divides $f(x)$ but $(x-a)^{t+1}$ does not.

Let's briefly consider the evaluation at a homomorphism for a fixed $a \in k$.

$$\begin{aligned} \varphi_a : k[x] &\longrightarrow k \\ f(x) &\longmapsto f(a) \end{aligned}$$

The homomorphism is surjective. By the roots-factors theorem, the kernel of this map is the set of polynomials that are multiples of $x-a$. That is, the ideal $\langle x-a \rangle$.

6.3 Derivatives

Another important distinction between $\mathbb{Q}$ and $k(x)$ (for k be a field). Is that polynomials have a derivative. We will show that there is a well-defined derivative function on $k[x]$ without resorting to limits, and which makes sense even over finite fields. We require that the following properties of the proposed derivative (with respect to x) hold.

(1) The derivative d maps $k[x]$ to $k[x]$.

(2) $d(x) = 1$.

(3) Additivity: $d(f + g) = d(f) + d(g)$.

(4) The Product Rule: $d(fg) = fd(g) + gd(f)$.

Exercises 6.3.1. Using just the rules above, show that the following properties would have to hold for the derivative:

(a) $d(0) = 0$ [use linearity].

(b) $d(1) = 0$ [use the product rule].

(c) $d(m) = 0$ for $m = 1 + 1 + \cdots + 1$ (with m terms) [use linearity].

(d) Show that if $k = \mathbb{Q}$ then $d(a) = 0$ for $a \in \mathbb{Q}$ [use the product rule].

For arbitrary fields we turn the result about $\mathbb{Q}$ in the last problem into an additional assumption:

(5) The derivative of a constant is 0: $d(a) = 0$ for $a \in k$.

This was proven from the rules (1)-(4) for derivatives in $\mathbb{Q}[x]$ in the problem above, but, more generally, it has to be assumed. In fact, the derivative can be defined much more broadly (of a ring R relative to a subring (call it k) with property (5) requiring that the derivative of any element of k is 0).

Exercises 6.3.2. Using the five rules above for the proposed derivative, show that the following properties would have to hold.

(a) (HW) $d(ax^m) = amx^{m-1}$ [induction, product rule].

(b) (HW) $d(\sum_{m\in\mathbb{N}_0} a_m x^m) = \sum_{m\in\mathbb{N}_0} a_m m x^{m-1}$.

The previous exercises have shown that the assumptions (1)-(5) determine that a derivative satisfying those rules can have only one possible form.

Definition 6.3.3. Let k be a field, define the **derivative with respect to** x on $k[x]$ over k to be the function d defined by

$$d(\sum_{m\in\mathbb{N}_0} a_m x^m) = \sum_{m\in\mathbb{N}_0} a_m m x^{m-1}$$

Exercises 6.3.4. With this definition:

(a) Show that the properties (1)-(5) do indeed hold.

(b) Show that the derivative map is a homomorphism of vector spaces over k.

(c) (HW) Show that the derivative can be extended to the ring of fractions $k(x)$ and that the quotient rule that you know from calculus must hold (first use the product rule to establish the reciprocal rule $d(1/a(x)) = -a'(x)/(a(x))^2$).

Exercises 6.3.5. The derivative of a composition of polynomials

(a) Find a formula for the derivative of $\big(g(x)\big)^n$.

(b) Use the previous result to derive the formula for the derivative of a composition of polynomials, $f\big(g(x)\big)$, that you learned in calculus. (Write $f(x) = f_n x^n + \cdots + f_0$.)

Exercises 6.3.6. (HW) Let k be a field of characteristic 0 and $f(x) \in k[x]$.

(a) Suppose a is a root of multiplicity m of f. Show that a is a root of multiplicity $m-1$ of $f'(x)$.

(b) Suppose the unique factorization of $f(x)$ is

$$f(x) = u\big(p_1(x)\big)^{e_1}\big(p_2(x)\big)^{e_2}\cdots\big(p_t(x)\big)^{e_t}$$

with the $p_i(x)$ being distinct monic irreducibles. Show that

$$f'(x) = \big(p_1(x)\big)^{e_1-1}\big(p_2(x)\big)^{e_2-1}\cdots\big(p_t(x)\big)^{e_t-1}h(x)$$

for some $h(x)$ that is not divisible by any of the $p_i(x)$.

(c) For $f \in k[x]$ and f monic, show that $\operatorname{rad}(f) = f/\gcd(f, f')$. This shows that the radical of a polynomial is easy to compute, via the Euclidean algorithm and polynomial division.

Strange things happen in finite characteristic.

Exercises 6.3.7. Let p be prime and $q = p^r$ a power of p. We will consider derivatives in $\mathbb{F}_{p^r}[x]$.

(a) Over a field of characteristic 0, the kernel of the derivative map is just the constant polynomials. What is the kernel for $\mathbb{F}_q[x]$?

(b) Give an example of a polynomial over $\mathbb{F}_p$ such that $\gcd(f, f') \neq \operatorname{rad}(f)$.

(c) Consider the polynomial $x^p - t$ in $\mathbb{F}_p(t)[x]$. Show that this polynomial is irreducible. [Assume it factors, take the derivative.]

6.4 Mason-Stothers Theorem and Fermat's Last Theorem for Polynomial Rings

We can now prove the analogues of the *abc*-conjecture and Fermat's Last Theorem for $k[x]$. We assume the characteristic of k is 0.

Mason-Stothers

In previous exercises we showed $\operatorname{rad}(f) = f/\gcd(f, f')$. Let's introduce the unique factorization of f, $f = u\prod_{i=1}^{t} p_i^{e_i}$ with the p_i irreducible, $e_i \geq 1$ and u a constant. Now introduce some notation for the radical of f, denote it f_r, the "excess", f_e, and another term derived from the derivative that I will call f_k.

$$\begin{aligned}
f &= u\prod_{i=1}^{t} p_i^{e_i} \\
f_r &= \prod_{i=1}^{t} p_i \\
f_e &= u\prod_{i=1}^{t} p_i^{e_i-1} = \frac{f}{f_r}
\end{aligned}$$

Taking the derivative,

$$\begin{aligned}
f' &= u\Big(\sum_{i=1}^{t} e_i p_i' p_i^{e_i-1}\big(\prod_{j\neq i} p_j^{e_j}\big)\Big) \\
&= u\Big(\prod_{i=1}^{t} p_i^{e_i-1}\Big)\Big(\sum_{i=1}^{t} e_i p_i' \prod_{j\neq i} p_j\Big) \\
&= f_e f_k
\end{aligned}$$

where,

$$f_k = \sum_{i=1}^{t} e_i p_i' \prod_{j\neq i} p_j = \sum_i e_i p_i' \frac{f_r}{p_i}$$

Note that f_k is comprime to f: for each irreducible factor p_i of f, p_i divides all summands of f_k except for the ith term, which it does not divide.

Now we can prove the Mason-Stothers theorem.

Theorem 6.4.1. *Let k be a field of characteristic 0. Let f, g, h be nonconstant coprime polynomials in $k[x]$ satisfying $f + g = h$. Then*

$$\max\Big(\deg(f), \deg(g), \deg(h)\Big) < \deg\big(\operatorname{rad}(fgh)\big)$$

Proof. Note that if any two of f, g, h have a common factor of positive degree then all do, so it is enough to assume that any two are coprime.

Taking the derivative of both sides of the equation, we have these two equalities:

$$\begin{aligned} f + g &= h \\ f' + g' &= h' \end{aligned}$$

Multiplying the first by g' and the second by g we have

$$\begin{aligned} fg' + gg' &= hg' \\ f'g + g'g &= h'g \end{aligned}$$

Substracting,

$$fg' - f'g = hg' - h'g$$

Now we introduce the notation defined above,

$$\begin{aligned} f_r f_e g_e g_k - f_e f_k g_r g_e &= h_r h_e g_e g_k - h_e h_k g_r g_e \\ f_e g_e\big(f_r g_k - f_k g_r\big) &= h_e g_e\big(h_r g_k - h_k g_r\big) \end{aligned}$$

Observe that f_e, g_e, h_e are coprime and each divides $fg' - f'g$, which equals $hg' - h'g$. Therefore the product $f_e g_e h_e$ divides $fg' - f'g$ and consequently,

$$\deg(f_e) + \deg(g_e) + \deg(h_e) \leq \deg(fg' - f'g)$$

Observe also that

$$\deg(f'g - fg') \leq \max(\deg(f'g), \deg(fg')) = \deg(f) + \deg(g) - 1$$

Now we substitute $\deg(f_e) = \deg(f) - \deg(f_r)$ and similarly for the others.

$$\deg(f) - \deg(f_r) + \deg(g) - \deg(g_r) + \deg(h) - \deg(h_r) \leq \deg(f) + \deg(g) - 1$$

Rearranging,

$$\deg(h) \leq \deg(f_r) + \deg(g_r) + \deg(h_r) - 1$$

Using comprimality of f, g, h,

$$\deg(f_r) + \deg(g_r) + \deg(h_r) = \deg(f_r g_r h_r) = \deg\big(\operatorname{rad}(fgh)\big)$$

So

$$\deg(h) \leq \deg(\operatorname{rad}(fgh)) - 1$$

The argument is easily reworked to show the same bounds for $\deg(f)$ and $\deg(g)$. □

Exercises 6.4.2. Extend the Mason-Stothers theorem to fields of finite characteristic.

(a) Let f, g, h be nonconstant coprime polynomials in $k[x]$ satisfying $f + g = h$. Then either

$$\max\{\deg(f), \deg(g), \deg(h)\} < \deg(\operatorname{rad}(fgh))$$

or each of f, g, and h has derivative 0. [Hint: the key step is either $fg' - f'g = 0$ or it does not equal 0.

The analogue of Fermat's last theorem

Theorem 6.4.3. *Let k be a field of characteristic 0 and let $n \geq 3$ be an integer. Suppose that $a(t)$, $b(t)$, and $c(t)$ are three coprime polynomials in $k(t)$ satisfying $a(t)^n + b(t)^n = c(t)^n$. Then all three polynomials are constants.*

Exercises 6.4.4. Use the Mason-Stothers theorem to prove the analogue of Fermat's last theorem for polynomials.

(a) Show that you can assume, without loss of generality, that $c(x)$ has degree that is greater than or equal to the degrees of $a(x)$, $b(x)$.

(b) Prove the theorem.

(c) (HW) Show that the FLT holds for polynomials that aren't coprime provided $a(t)$ and $b(t)$ aren't associates.

(d) (HW) Show that there do exist polynomials $a(t), b(t), c(t)$ of positive degree such that $a(t)^2 + b(t)^2 = c(t)^2$.

6.5 Unique Factorization Domains

Let R be a unique factorization domain. The goal of this section is to show that $R[x]$ is also a unique factorization domain. We will make use of unique factorization in $K[x]$ where K is the field of quotients of R.

Recall that for $\mathbb{Z}$ and for $k[x]$, with k a field, a key step in proving unique factorization was to show the equivalence between irreducibility and primality. Let us extend the definition of these concepts to an arbitrary integral domain.

Definition 6.5.1. Let R be an integral domain. We say that a **divides** b, written $a|b$, when there is some c such that $ac = b$. An element p in R is **prime** when $p|ab$ implies that either $p|a$ or $p|b$. An element p is **irreducible** when $p = ab$ implies that either a or b is a unit in R.

Observe that p is a prime if and only if $\langle p \rangle$ is a prime ideal.

Some comments on *uniqueness* of factorization

In $\mathbb{Z}$, the unique factorization of an integer is $up_1^{e_1} \cdots p_r^{e_r}$ in which the p_i are distinct *positive* primes, the e_i are positive integers, and u is 1 or -1. In $k[x]$ we stated the unique factorization to be $u(p_1(x))^{e_1} \cdots (p_r(x))^{e_r}$ in which the $p_i(x)$ are distinct *monic* polynomials, the e_i are positive integers, and u is a nonzero constant (an element of k). In each case the uniqueness is up to reordering of the prime power factors, but we did not need to worry about $(2x-2)((1/2x)$ and $(x-1)x$ both being valid factorizations into irreducibles of x^2-x. We chose the monic polynomial once and for all as the one to use in a standard factorization.

Definition 6.5.2. Let R be an integral domain. Two elements of R are **associates** if they differ by a factor that is a unit. That is, if $r_1, r_2 \in R$ we say r_1 and r_2 are associates when there is a unit $u \in R$ such that $r_2 = ur_1$.

It should be clear that the relation of being associates is an equivalence relation. Additionally, if p is irreducible in R then so are all associates of p.

Henceforth, for our unique factorization domain R we will assume that for each equivalence class of irreducibles (under the relation of being associates) we have chosen a unique representative that will be used in the standard factorization. Unique factorization may then be stated as follows. Let Irr be a set of representatives for the irreducibles, containing one element from each equivalence class of irreducible polynomials (under the relation of being associates).

> Each element of R may be written in a unique way as $u\prod_{p\in\mathrm{Irr}} p^{e_p}$. In which u is a unit, each $e_p \in \mathbb{N}_0$, and only a finite number of the e_p are nonzero.

The gcd of two elements $a, b \in R$ is now uniquely defined. Factor each of a, b into irreducibles and for each irreducible p take the minimum power that divides both a and b. Multiply over all irreducibles to get the gcd. Explicitly, if $a = u\prod_{p\in\mathrm{Irr}} p^{d_p}$ and $b = v\prod_{p\in\mathrm{Irr}} p^{e_p}$, then $\gcd(a, b) = \prod_{p\in\mathrm{Irr}} p^{\min(d_p,e_p)}$. Similarly, $\operatorname{lcm}(a, b) = \prod_{p\in\mathrm{Irr}} p^{\max(d_p,e_p)}$.

The field K is a unique factorization domain only in the most trivial sense; since all nonzero elements are all units, there are no irreducibles. On the other hand, K inherits a meaningful unique factorization because it was constructed from R. An element of K can be written a/b with $a, b \in R$. We may express this "in lowest terms" by factoring a and b and dividing each by their gcd. Thus, using the factorizations in the previous paragraph, we let $P = \{p \in \mathrm{Irr} : d_p > e_p\}$ and $Q = \{p \in \mathrm{Irr} : d_p < e_p\}$, and then

$$\begin{aligned}\frac{a}{b} &= \frac{a/\gcd(a,b)}{b/\gcd(a,b)}\\ &= (uv^{-1})\frac{\prod_{p\in P} p^{d_p-e_p}}{\prod_{p\in Q} p^{e_p-d_p}}\end{aligned}$$

This is in lowest terms—the numerator and denominator have no common irreducible factor—and it is the unique expression of $\frac{a}{b}$ in this form.

Proposition 6.5.3. *Any element of K can be uniquely expressed in the form $u\frac{a}{b}$ in which u is a unit, a and b are the product of elements of* Irr*, and a and b are coprime.*

Proving $R[x]$ is a UFD

As we said earlier the key step in proving unique factorization is to show the equivalence between irreducibility and primality.

The first step, an easy one, is that one direction of the equivalence always holds.

Proposition 6.5.4. *In any integral domain every prime element is irreducible.*

Proof. Look at the proof for $\mathbb{Z}$ and see that it works with little alteration. □

In general rings, irreducible elements are not necessarily prime.

Example 6.5.5. Here is a domain in which an irreducible element may not be prime. Consider $\mathbb{Z}[\sqrt{-5}]$. One can check that 2,3, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are all irreducible. One can also verify that the only units in $\mathbb{Z}[\sqrt{-5}]$ are ±1, so no two of these irreducibles are associates. Here are two distinct factorizations into irreducibles that are equal

$$\left(1+\sqrt{-5}\right)\left(1-\sqrt{-5}\right) = 6 = 2 \cdot 3$$

The key element in establishing unique factorization is thus to show that when an element of $R[x]$ is irreducible it is also prime. To get started we will use a coarse factorization for elements of $K[x]$.

Definition 6.5.6. Let $f(x) \in K[x]$ have positive degree and write its coefficients in lowest terms

$$f(x) = u_n \frac{a_n}{b_n} x^n + u_{n-1} \frac{a_{n-1}}{b_{n-1}} x^{n-1} + \cdots + u_1 \frac{a_1}{b_1} x + u_0 \frac{a_0}{b_0}$$

Here each u_i is a unit, a_i and b_i are products of elements of Irr and have no common irreducible factor. Let $A = \gcd(a_1, a_2, \ldots, a_n)$ and let $B = \operatorname{lcm}(b_1, b_2, \ldots, b_n)$. We define the **unit part** of $f(x)$ to be $u_f = u_n$, the **content** of $f(x)$ to be $c_f = A/B$, and the **primitive part** of $f(x)$ to be $f^*(x) = u_f^{-1}(B/A)f(x)$. Clearly, we have $f(x) = u_f c_f f^*(x)$.

We say $f(x) \in K[x]$ is **primitive** when it has positive degree, and $u_f = c_f = 1$. In particular the leading coefficient of f is a product of elements of Irr.

Since the gcd and lcm of a polynomial are uniquely defined—because Irr has a unique representative for each associate class of irreducibles—the factorization of $f(x) \in K[x]$ into the product of its unit part, its content and its primitive part is unique. Note also that a primitive polynomial is actually in $R[x]$ since the lcm of the denominators is 1.

Proposition 6.5.7. *Let $f(x)$ have positive degree, with content c_f and primitive part $f^*(x)$. Then $f^*(x)$ is a primitive polynomial. That is,*

(1) $f^* \in R[x]$,

(2) $c_f = 1$, and

(3) $u_f = 1$.

Proof. (1) Let $f(x)$ be written as follows with each $\frac{a_i}{b_i}$ in lowest terms and define A and B.

$$f(x) = u_n \frac{a_n}{b_n} x^n + u_{n-1} \frac{a_{n-1}}{b_{n-1}} x^{n-1} + \cdots + u_1 \frac{a_1}{b_1} x + u_0 \frac{a_0}{b_0}$$
$$A = \gcd(a_1, a_2, \ldots, a_n)$$
$$B = \operatorname{lcm}(b_1, b_2, \ldots, b_n)$$

Then

$$f^*(x) = \frac{a_n}{A}\frac{B}{b_n}x^n + (u_n^{-1}u_{n-1})\frac{a_{n-1}}{A}\frac{B}{b_{n-1}}x^{n-1} + \cdots + (u_n^{-1}u_1)\frac{a_1}{A}\frac{B}{b_1}x + (u_n^{-1}u_0)\frac{a_0}{A}\frac{B}{b_0}$$

Each $\frac{a_i}{A}$ and each $\frac{B}{B_i}$ is an element of R by the choice of A and B, so $f^*(x) \in R[x]$.

(2) We know that $f^*(x) \in R[x]$; we will show that the gcd of the coefficients of $f^*(x)$ is 1 by showing that for any $p \in \text{Irr}$, p does not divide at least one of the coefficients of $f^*(x)$.

Let $p \in \text{Irr}$ be such that $p|B$, and let p^e be the largest power of p that divides B. Then, for some i, p^e divides b_i, but p^{e+1} does not divide b_i. We thus have $p \nmid (B/b_i)$. Since the coefficients of $f(x)$ were in lowest terms, we also have $p \nmid a_i$. Thus $p \nmid (a_iB/b_i)$. This shows that no irreducible that divides B divides all of the a_iB/b_i. Consequently, the only irreducibles that divide all of the a_iB/b_i actually divide all of the a_i and therefore we have

$$\gcd(a_0B/b_0, \ldots, a_nB/b_n) = \gcd(a_0, \ldots, a_n) = A$$
$$\gcd(a_0B/Ab_0, \ldots, a_nB/Ab_n) = \gcd(a_0/A, \ldots, a_n/A) = 1$$

(3) The leading term of f^* is a product of irreducibles, since A, a_n, B, b_n are. So $u_f = 1$. □

Exercises 6.5.8. For each of the following polynomials, simplify the coefficients, determine the content, the unit part and the primitive part. You should notice a profound similarity between the two examples.

(a) Let $f(x) = -4x^3 + \frac{8}{15}x^2 + \frac{6}{9}x + \frac{16}{20}$ in $\mathbb{Q}[x]$.

(b) Let $f(x) = 2y^2x^3 + \frac{y^3}{y^2+3x+2}x^2 + \frac{y^2+y}{y^2+2y+1}x + \frac{y^4}{y^3+2y^2}$ in $\mathbb{Q}(x)[y]$.

Lemma 6.5.9. *The product of two primitive polynomials is primitive. More generally, let $f(x), g(x) \in K[x]$ both have positive degree. Then $u_{fg} = u_fu_g$, $c_{fg} = c_fc_g$ and $\big(f(x)g(x)\big)^* = f^*(x)g^*(x)$.*

Proof. Suppose that $f(x)$ and $g(x)$ are both in $R[x]$ and that they are both primitive. We will show that for any $p \in \text{Irr}$ there is a coefficient of $f(x)g(x)$ that is not divisible by p. Since p is arbitrary the gcd of the coefficients of $f(x)g(x)$ is 1.

Since $f(x)$ is primitive, we know that p does not divide all coefficients of $f(x)$. Let i be the smallest integer such that f_i is not divisible by p. Similarly, let j be the smallest integer such that g_j is not divisible by p. Then the $(i+j)$th coefficient of $f(x)g(x)$ is

$$f_0g_{i+j} + f_1g_{i+j-1} + \cdots + f_ig_j + \cdots + f_{i+j}g_0$$

We see that p divides every term except for one; it does not divide $f_i g_j$. Thus p does not divide the $(i+j)$th term of $f(x)g(x)$. This was what we needed to show.

Now let $f(x), g(x) \in K[x]$ be arbitrary and $h(x) = f(x)g(x)$. Then $h(x) = u_f c_f f^*(x) u_g c_g g^*(x) = \big(u_f u_g\big)\big(c_f c_g\big)\big(f^*(x)g^*(x)\big)$. The first term is a unit in R, the second term is in K and is a product of irreducibles (or their inverses). The last term is a primitive polynomial in $R(x)$. By the uniqueness of the factorization into unit part, content, and primitive part we are done. □

Corollary 6.5.10. *Let $m(x) \in R[x]$ have positive degree. Then*

(1) If $m(x)$ is irreducible then $c_m = 1$.

(2) Let $m(x)$ be primitive. Then $m(x)$ is irreducible in $R[x]$ if and only if it is irreducible in $K[x]$.

(3) If $m(x)$ is irreducible in $R[x]$ then it is prime in $R[x]$.

Proof. In the factorization $m(x) = u_m c_m m^*(x)$, we have c_m is a product of irreducbiles in R and $m^*(x)$ is primitive and therefore in $R[x]$. If c_m were not 1, then $m(x)$ would be reducible.

Let $m(x) \in R[x]$ be primitive with degree $n > 0$. Any factorization of $m(x)$ in $R[x]$ would involve polynomials of degree less than n and would therefore extend to a factorization in $K[x]$. Thus if $m(x)$ is reducible in $R[x]$ then it is also reducible in $K[x]$.

Conversely, suppose that $m(x)$ is reducible in $K[x]$ with $m(x) = f(x)g(x)$ and both $f(x)$ and $g(x)$ having positive degree. Then, $c_m m^*(x) = c_f c_g f^*(x)g^*(x)$. Bust since $m(x)$ is primitive, $c_f c_g = c_m = 1$. In addition, $m(x) = m^*(x) = f^*(x)g^*(x)$. This shows that $m(x)$ is reducible in $R[x]$.

Finally, suppose $m(x)$ is irreducible in $R[x]$ and suppose $m(x)$ divides $f(x)g(x)$ with both factors in $R[x]$. Consider that $m(x)$ also divides $f(x)g(x)$ in $K[x]$. From the previous item, $m(x)$ is irreducible in $K[x]$ so, it is prime in $K[x]$, so it must divide one of $f(x)$ or $g(x)$. Suppose $f(x) = m(x)h(x)$ is a factorization in $K[x]$. We have $f^*(x) = m^*(x)h^*(x)$. This factorization is in $R[x]$. Since $f(x) = c_f f^*(x)$ with $c_f \in R$, and $m(x) = m^*(x)$, we have $m(x)$ divides $f^*(x)$ and therefore $m(x)$ divides $f(x)$. Thus we have $m(x)$ divides one factor of the product $f(x)g(x)$ *within* $R[x]$. This shows $m(x)$ is prime. □

Theorem 6.5.11. *If R is a unique factorization domain, so is $R[x]$.*

Proof. Existence: The proof is similar to that for $k[x]$. We use induction on the degree, starting with the factorization into irreducibles in R, which gives a base step in degree 0. Assume we have factorization into irreducibles for polynomials of degree less than n. Let $f(x) \in R[x]$ have degree n and let its content be c_f and primitive part $f^*(x)$. Since $f(x) \in R$, so is c_f and it has a factorization into irreducibles. If f^* is irreducible itself we are done. If it can be factored then, since $f(x)$ is primitive, each factor must have degree less than n. Each of the factors can be written as a product of irreducibles by the induction hypothesis; so $f^*(x)$ can also be written as the product of irreducibles.

Uniqueness: The proof is entirely analogous to that in the integers or $k[x]$. It follows from the fact that all irreducibles in $R[x]$ are prime as shown in Corollary 6.5.10.

□

Corollary 6.5.12. *Let k be a field. The polynomial ring in n variables $k[x_1, \ldots, x_n]$ is a unique factorization domain.*

Proof. Induction on n. □

Exercises 6.5.13. Discussion items

(a) Prove that every prime is irreducible in an integral domain.

(b) Explain why the "lowest terms" expression of an element of K is unique.

The following problem shows another useful aspect of considering a polynomial ring $k[x_1, \ldots, x_n]$ as $\Big(k[x_1, \ldots, x_{n-1}]\Big)[x_n]$. As a first step, consider $k[x, y]$ as $\big(k[x]\big)[y]$. That is, we consider an element f of this ring as a polynomial in y with coefficients from $k[x]$. In this context, the leading term of f is the one with highest degree in y and the leading coefficient is an element of $k[x]$. That leading coefficient itself has a highest degree term in x. Now, consider $k[x, y]$ as a bivariate polynomial ring, so an element f is treated as a finite sum of terms of the form $f_{a,b}x^a y^b$ with $f_{a,b} \in k$. Up to this point there was no way to choose a leading term for f. Now we have one: we take the terms of highest degree in y, and among those we choose the term that has highest degree in x.

Exercises 6.5.14. A total ordering of monomials

(a) Following the discussion above, find the leading term of $f = 2x^3y + 4x^2y^2 + 3xy^3 + x^2y^2 + 5x^4y^3 + 7x^6y^2$.

(b) What would the leading term of f be if we considered $k[x, y]$ first as $\big(k[y]\big)[x]$?

(c) In either case (f as a polynomial in y with coefficients in $k[x]$ or as a polynomial in k with coefficients in $k[y]$), show that $\mathrm{LT}(fg) = \mathrm{LT}(f) * \mathrm{LT}(g)$, where LT indicates leading term.

(d) Assuming that the leading term is determined as in (b), finish this extension of the Quotient-Remainder Theorem to $k[x, y]$: Given a and b in $k[x, y]$ with $b \neq 0$, there exist unique q and r satisfying the following two properties....

(e) Prove this extension.

(f) Explain how the QR theorem above s is useful for computing in the quotient ring $k[x,y]/\langle b\rangle$.

(g) Give a recursive definition of the leading term of an element f in $k[x_1,\ldots,x_n]$.

Representatives for the irreducibles in $R[x]$

We can define a set of representatives for irreducibles in $R[x]$ by building on top of Irr_R, the set of representatives for the irreducibles in R. Let $m(x)$ be irreducible in $R[x]$ of degree $n > 1$. We will choose a representative for the set of irreducibles that are associates of $m(x)$. We have $m(x) = u_m c_m m^*(x)$. If $c_m \neq 1$ then $m(x)$ would not be irreducible. So, $c_m = 1$.

Factor m_n, the leading coefficient of $m(x)$, as $m_n = u \prod_{p\in Irr} p_p^e$ (with each e_p nonnegative, and only a finite number nonzero, and u a unit in R). We are allowing the possibility that the product is empty, in which case m_n is a unit. We take $u^{-1}m(x)$ as the representative for the associates of $m(x)$. In other words we require the leading term of an element of $\mathrm{Irr}_{R[x]}$ to be a product of elements of Irr_R, disallowing multiplication by a unit in R.

6.6 Irreducible Polynomials

Throughout this section we assume that R is a unique factorization domain and K is its field of quotients. How can we determine if a primitive polynomial $m(x) \in R[x]$ is irreducible? In this section, we introduce three tests for irreducibility, each is a straightforward generalization of tests (that you may have seen at some point) for irreducibility in $\mathbb{Q}[x]$.

Before proceeding we have the following observation.

Proposition 6.6.1. *Let $\varphi : R[x] \longrightarrow R[x]$ be a ring isomorphism. Then $m(x) \in R[x]$ is irreducible if and only if $\varphi\big(m(x)\big)$ is irreducible.*

It can be helpful to make such a transformation to apply one of the tests below (see 6.6.14).

Leading coefficient and constant term

Proposition 6.6.2. *Let $m(x) = m_n x^n + m_{n-1}x^{n-1} + \cdots + m_1 x + m_0$ be a primitive polynomial of degree n in $R[x]$. Let r, s be coprime elements of R. Then r/s is a root of $m(x)$, as an element of $K[x]$, if and only if $sx - r$ is a factor of $m(x)$ as an element of $R[x]$. This can occur only if $s|m_n$ and $r|m_0$.*

Consequently, the search for roots of $m(x)$ (or for linear factors of $m(x)$) is narrowed to considering r/s (or factors of the form $sx - r$) for s a factor of m_n and r a factor of m_0.

Proof. Clearly if $sx - r$ is a factor of $m(x)$ in $R[x]$ then it is also a factor of $m(x)$ in $K[x]$ and consequently r/s is a root of $m(x)$.

Suppose that r/s is a root of $m(x)$ considered in $K[x]$. Then by the Roots-Factors theorem, $x - r/s$ is a factor of $m(x)$ so $m(x) = (x - r/s)g(x)$ for some $g(x) \in K[x]$. Since $m(x)$ is primitive and the content of $x - r/s$ is $1/s$ (or some associate of it), the content of $g(x)$ must be s (or the appropriate associate). Furthermore $m(x) = (sx - r)(g(x)/s)$ is a factorization of $m(x)$ in $R[x]$. Since the leading term of $m(x)$ is the product of the leading terms of the two factors, $s|m_n$. Similarly, the constant term is the product of the constant terms of the two factors, so $r|m_0$. □

This narrows the search for a linear factor of a primitive polynomial $m(x)$, but it doesn't help determine higher degree factors. For $m(x)$ of degree 2 or 3, $m(x)$ is irreducible if and only if it has no linear factor. For $m(x)$ of degree n, $m(x)$ is irreducible if it has no factor of degree $\lfloor n/2 \rfloor$ or less. The following more general result can help with higher degree.

Proposition 6.6.3. *Let $m(x) = m_n x^n + m_{n-1} x^{n-1} + \cdots + m_1 x + m_0$ be a primitive polynomial of degree n in $R[x]$. Suppose $m(x)$ has a factorization in $R[x]$ as $m(x) = a(x)b(x)$. Then*

(1) $a(x)$ and $b(x)$ are primitive with degrees adding to n;

(2) the product of the leading terms of $a(x)$ and $b(x)$ is m_n;

(3) the product of the constant terms of $a(x)$ and $b(x)$ is m_0.

We know that $\mathbb{Z}$ is a UFD and that $\mathbb{Q}[x]$ is a UFD. Theorem 6.5.11 established that $\mathbb{Z}[x]$ is a UFD. It may be counterintuitive, but to prove that an $m(x) \in \mathbb{Q}[x]$ is irreducible, it is useful to consider its primitive part $m^*(x) \in \mathbb{Z}[x]$. Using the previous proposition is one way. The next exercises are examples. We will treat two other methods in the next subsections.

Exercises 6.6.4. Show that each of the following is irreducible.

(a) $3x^3 + 5x + 10$. [How can we quickly tell that any root must be an integer?]

(b) $xy^3 + (x^2 + 2)y + x^2 + x$. [How can we quickly tell that any root must be a polynomial in x?]

Exercises 6.6.5.

(a) For what integer values of u does $y^3 + uy + 2$ factor in $\mathbb{Q}[y]$?

(b) For what values of $u(x) \in \mathbb{Q}[x]$ does $y^3 + u(x)y + x$ factor in $\mathbb{Q}[x, y]$?

Exercises 6.6.6.

(a) For what integer values of u does $f(y) = y^4 + uy^2 + 4$ factor in $\mathbb{Q}[x]$? Consider separately the case when $f(y)$ has a root, and when it factors as a product of irreducible quadratics.

(b) For what values of $u(x) \in \mathbb{Q}[x]$ does $y^4 + u(x)y^2 + x^2$ factor in $\mathbb{Q}[x, y]$? Consider separately the case when $f(y)$ has a root in $\mathbb{Q}[x]$ and when it factors as a product of irreducible quadratics.

Exercises 6.6.7. Show that $x^4 - 10x + 1$ is irreducible over the rationals as follows.

(a) Show it has no roots.

(b) Try to factor it as a product of quadratics and derive a contradiction.

Exercises 6.6.8. Use the general theorem above to show that one of the polynomials over $\mathbb{Z}$ below is irreducible and to factor the other.

(a)

(b)

Exercises 6.6.9. Consider the polynomials below over $\mathbb{F}_2[x]$. Use Proposition 6.6.2 to factor one of the polynomials and to show that the other is irreducible.

(a)
(b)

Reducing modulo a prime ideal

Given a ring R and ideal J in R, there is a natural homomorphism from $R[x]$ to $R/J[x]$ that simply reduces each coefficent of a polynomial modulo J: the image of $a(x) = \sum a_i x^i$ is $\overline{a}(x) = \sum \overline{a}_i x^i$ where $\overline{a}_i$ is $a_i + J$, the congruence class of a_i in R/J.

Proposition 6.6.10. *Let $m(x) \in R$ be a primitive polynomial of degree $n > 1$. Suppose that J is an ideal J in R such that $m_n \notin J$. Let $\overline{m}(x)$ be the element of $R/J[x]$ obtained by reducing the coefficients of m modulo J. If $\overline{m}(x)$ is irreducible then so is $m(x)$.*

Proof. We will show the logically equivalent statement: if $m(x)$ is reducible then so is $\overline{m}(x)$. Suppose $m(x)$ is reducible and $m(x) = a(x)b(x)$ is a factorization in $R[x]$ with $\deg(a(x))$ and $\deg(b(x))$ both greater than 1. Note that the leading coefficient of $m(x)$ is the product of the leading coefficients of $a(x)$ and $b(x)$. Since m_n is not in J, the leading coefficients of $a(x)$ and $b(x)$ are not in J. Thus $\deg(\overline{a}(x)) = \deg(a(x))$ and similarly $\deg(\overline{b}(x)) = \deg(b(x))$.

Reducing each coefficient modulo J we have in $R/J[x]$, $\overline{m}(x) = \overline{a}(x)\overline{b}(x)$. Since the leading coefficients of $\overline{a}(x)$ and $\overline{b}(x)$ are non zero, we have a nontrivial factorization of $\overline{m}(x)$, as was to be proved. □

Exercises 6.6.11. Test whether the following polynomials are irreducible over the rationals.

(a) $3x^2 - 7x - 5$
(b) $2x^3 - x - 6$
(c) $x^3 - 9x - 9$

Exercises 6.6.12.

(a) Show that $y^4 + 7y^3 + 2y^2 + 1$ is irreducible in $\mathbb{Q}[y]$.
(b) Show that $y^4 + (x^3 + 7)y^3 + x^2y^2 + x^2 + 2x + 3$ is irreducible in $\mathbb{Q}[x, y]$.

Eisenstein's criterion

Proposition 6.6.13 (Eisenstein)**.** *Let $m(x) \in R[x]$ be a primitive polynomial of degree $n > 1$. Suppose that there is a prime ideal P in R such that*

(1) $m_n \notin P$.

(2) $m_i \in P$ for $i = 0, \ldots, n-1$

(3) $m_0 \notin P^2$

Then $m(x)$ is irreducible.

Proof. We assume $m(x)$ satisfies the properties above and that it factors as $m(x) = a(x)b(x)$. We show that one of the factors is an invertible element of R (in particular, of degree 0). Since this is true for an arbitrary factorization, $m(x)$ is irreducible.

Since the leading coefficient of m is not in P, the leading coefficients of $a(x)$ and $b(x)$ are also not in P. The constant term of m is in P but not P^2. Since $m_0 = a_0b_0$ and P is prime, exactly one of a_0 and b_0 is in P and the other is not. Let us assume $a_0 \in P$ and $b_0 \notin P$. Let t be minimal such that $a_t \notin P$ (we know the leading coefficient of $a(x)$ is not in P, so there exists such a t). Then we have

$$m_t = a_0b_t + a_1b_{t-1} + \cdots + a_{t-1}b_1 + a_tb_0$$

Each term a_ib_{t-i} is in P except the last, which is not, so $m_t \notin P$. By our assumptions on $m(x)$, the only coefficient not in P is m_n. Thus $t = n$, $a(x)$ has degree n and $b(x)$ has degree 0. Since $m(x)$ is primitive, $b(x)$ must be a unit. □

Exercises 6.6.14. Use Eisenstein's criterion to show that the following are irreducible polynomials

- **(a)** Let p be a prime. Show that $x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible in $\mathbb{Q}[x]$. [Write it as $(x^p - 1)/(x - 1)$ and substitute $x = y + 1$.]
- **(b)** In $k[x, y]$, for any field k, $y^2 + x^2 - 1$ is irreducible.
- **(c)** In $k[x, y]$, for any field k, $y^n - x^3 - x^2$ is irreducible.

Exercises 6.6.15. Show these are irreducible over any field.

- **(a)** $xy^2 - z$
- **(b)** $xy^2 - z^2$
- **(c)** $x^2 + y^2 + z^2 - 1$ (this took me two steps)
- **(d)** $-4x^3z + 3x^2y^2 - 4y^3 + 6xyz - z^2$, the tangent surface to the twisted cubic. (I couldn't do this.)

Exercises 6.6.16.

- **(a)** Let p be prime. It is clear that $(x^p - 1)$ is not irreducible since it has a root, 1. Show that $\frac{x^p-1}{x-1} = x^{p-1} + x^{p-2} + \cdots + 1$ is irreducible over $\mathbb{Q}$. Use the isomorphism $x \longmapsto (x + 1)$ and Eisenstein's criterion.

Bibliography

[Ash02] Ash, *Abstract algebra: The basic graduate year*, Dover, 2000, revised 2002.

[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.