Abstract Algebra Notes

Preface

These are notes I wrote for my Algebraic Structures classes in 2024 and 2026. My goal is to give a good intuition for the concepts, with a lot of examples, and to make the proofs easy to follow. The chapter on rings and fields will be expanded soon.

If you spot any errors, please send me a note at `heinold@msmary.edu`.

Last updated: April 14, 2026.

Contents

Chapter 1

Introduction

1.1 First steps

The topic of these notes is *abstract algebra*. Depending on your background, this might be your first experience with an abstract topic like this. The focus is on concepts and on proving things, less on computations and applications. There are computations and applications, and we will talk about them, but they are not the main focus.

Most undergraduate programs have a required course in abstract algebra. There are many reasons why. First, it is a foundational class for many other theoretical math classes, particularly those you might study in grad school. It's a good first class for practicing with abstract concepts and proving things, as the concepts are not too complex and only knowledge of basic algebra is needed. This practice with abstract concepts is good for life in general. If you are planning on teaching basic algebra in the future, studying abstract algebra will give you a deeper understanding of what you have to teach. Finally, the material is actually very interesting.

If you are someone who absolutely needs to see a subject applied to other things, there are applications of abstract algebra that we'll see later in these notes. Abstract algebra is important in modern cryptography. It is used in checksums for network protocols. It shows up in crystallography, chemistry, and particle physics. It even explains properties of Rubik's Cubes and other puzzles.

What abstract algebra is about

The arithmetic we are all familiar with is about numbers and operations on them, things like $2+3$ or $4.09 \times \sqrt{2}$. These operations have certain properties. For instance, addition is commutative, which means that the order we add things in doesn't matter. For instance, $2+3$ and $3+2$ both give the same result. Addition is also associative. This means that $2+3+4$ can be done as $(2+3)+4$, which says to do $2+3$ and then add 4 to that, or it can be done as $2+(3+4)$, which says to do $3+4$ first and then add 2.

Throughout the 18th and 19th centuries, people started noticing that operations on other types of things followed properties that were very similar to the properties of arithmetic on real numbers. In some cases, these were just different number systems, but in others the objects were quite different, such as the symmetries of a polygon. In the 19th and early 20th centuries, people decided these things were all instances of more general or *abstract* concepts, which they gave names like *group*, *ring*, and *field*. These abstract objects are what we study in abstract algebra. Things we can show are true of an abstract group then apply to all specific instances of groups, like real numbers, symmetries of a polygon, and many other things.

1.2 The idea of a group

Group properties

Before we formally define a group, let's try to get some intuition for it. It took a long time of people playing around with examples before an actual definition was created. Groups start with two things – a set of objects and an operation on those objects. Two examples we will start with are the set of all integers along with the addition operation and the set of non-zero real numbers with the multiplication operation.

Closure Looking at the set of integers and the addition operation, one thing to notice is that whenever you add two integers, you get another integer and not some other type of object. For instance, $2+2=4$, $3+8=11$, and $-2+-7=-9$ are all operations on integers that result in another integer. This property is referred to as *closure*. There are many examples of sets and operations that don't have the closure property. For instance, if we looked at just the set of positive integers and the subtraction operation, we don't have the closure property. For instance, $2-3=-1$ and -1 is not a positive integer.

Identities Looking at the set of all integers with the addition operation, notice that the number 0 is special. If you add it to another integer, there is no change. For instance, $2+0=2$ and $4+0=4$. The term used for this is that 0 is called an *identity*. If we switch to the set of real numbers with the multiplication operation, then the identity is the number 1. It's the only real number where if you multiply it by any real number, the result doesn't change. For instance, $2\times 1=2$ and $6.43\times 1=6.43$.

Inverses Working again with the set of all integers with the addition operation, every integer has an additive *inverse*. The additive inverse of an integer n is a number that you can add to it to get 0 (the additive identity). For instance, the additive inverse of 7 is -7 since $7+-7=0$, and the additive inverse of 12 is -12 because $12+-12=0$. In general, $-n$ is the additive inverse of n.

Looking at real numbers with the multiplication operation, every nonzero number has a multiplicative inverse. The multiplicative inverse of a real number x is a number you can multiply by it to get 1 (the multiplicative identity). For instance, the multiplicative inverse of 7 is $\frac{1}{7}$ since $7\times\frac{1}{7}=1$, and the multiplicative inverse of $\frac{2}{3}$ is $\frac{3}{2}$ since $\frac{2}{3}\times\frac{3}{2}=1$. In general, $\frac{1}{x}$ is the multiplicative inverse of x, provided $x\neq 0$.

Associativity and commutativity As you might remember from basic algebra, the *associative property* for addition says that $(a+b)+c=a+(b+c)$. That is, if we want to add three numbers, it doesn't matter how we group them. For instance, if we want to do $2+3+4$, we can first do $2+3=5$ and then do $5+4$ to get 9. However, we could also do $3+4=7$ first and then do $2+7$ to get 9.

Multiplication is also associative. For instance, $2\times 3\times 4$ can be done via $2\times 3=6$ and then $6\times 4=24$, or we could do $3\times 4=12$ and then $2\times 12=24$. Associativity is a nice property for an operation to have. Without it, things are trickier to work with. The subtraction operation is not associative. Look at $2-3-4$: we have $(2-3)-4=-1-4=-5$, but $2-(3-4)=2-(-1)=3$.

An operation has the *commutative* property if the order of its operands doesn't matter. For instance, addition and multiplication are both commutative. Namely, $a+b=b+a$ and $a\times b=b\times a$. However, subtraction is not commutative since it is not always true that $a-b=b-a$. For instance, $2-3=-1$, while $3-2=1$. Order matters for subtraction.

1.3 Different kinds of arithmetic

Modular arithmetic A different type of arithmetic we all have real-world experience with is clock arithmetic. The thing that makes clock arithmetic different is the wrap-around effect. For instance, 3 hours after 11 o'clock

is 2 o'clock in the 12-hour system, and 5 hours after 22 o'clock is 3 o'clock in the 24-system. This generalizes to something called *modular arithmetic*.

The *modulus* or *mod* operation refers to the remainder left when dividing two integers. For instance, $17 \bmod 5 = 2$ because when you divide 17 by 5, you get 3 with 2 left over. The left-over 2 is the modulus. A quick way to compute $a \bmod m$ is to find the closest multiple of m less than a and subtract from a. For instance, $58 \bmod 7$ is 2 since 56 is the closest multiple of 7 less than 58, and $58 - 56 = 2$. Note that if $a < m$, then $a \bmod m$ is just a.

For any integer m greater than 0, we can take the set of integers from 0 to $m-1$ along with a modified addition operation where $a + b$ is done by adding like normal and then modding by m. For instance, if $m = 24$, then we're looking at the integers 0 to 23. Adding $2 + 3$ still gives 5, but adding $20 + 7$ gives 3, since $20 + 7 = 27$ and $27 \bmod 24$ is 3.

This system of arithmetic, called *modular arithmetic*, shares many properties with ordinary addition of integers, but there are some differences. For instance, the associative and commutative properties still hold, and 0 is still the additive identity. However, inverses are different. For ordinary integers, the inverse of a is $-a$. Specifically, the inverse is the thing you have to add to a so that the result is the identity, 0. In modular arithmetic mod 24, when you add two things and they come out to 24, then we mod by 24, we get 0. So inverses are what you have to add to get to 24. For instance, the inverse of 17 is 7 since $17 + 7 = 24$ (and $24 \bmod 24$ is 0).

Multiplication in modular arithmetic Multiplying numbers in modular arithmetic is interesting. We multiply like normal and then take the mod. For instance, if the modulus is 7, then to multiply 2 and 5, we do $2 \cdot 5 = 10$ and then $10 \bmod 7 = 3$. So the product of 2 and 5 is 3 in this system of arithmetic. Below is a multiplication table mod 7 for the integers 1 to 6.

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

Multiplication is still commutative and associative in this system, and the multiplicative identity is 1, just like with regular multiplication. However, we see that inverses are quite different here. For instance, the inverse of 2 is 4 since $2 \cdot 4$ results in 1 in this system and $4 \cdot 2$ also results in 1.

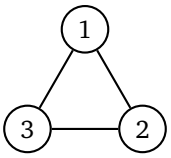
If we change the mod to 6, interesting things happen. Below is the multiplication table for it.

	1	2	3	4	5
1	1	2	3	4	5
2	2	4	0	2	4
3	3	0	3	0	3
4	4	2	0	4	2
5	5	4	3	2	1

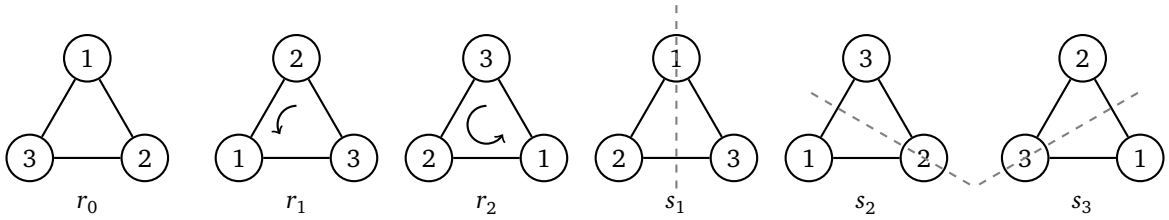
Doing $2 \cdot 3$ gives 6 and when we mod by 6, we get 0. So it's possible in this system to multiply two values and get 0. That's not something that can happen in ordinary arithmetic. Also, some numbers don't have a multiplicative inverse mod 6. In fact, only 1 and 5 have inverses, and each is its own inverse.

There are a lot of interesting properties to modular arithmetic, and we will explore some of them later. Many more properties are studied in number theory. Modular arithmetic also plays a critical role in modern cryptography. As a final note for now, note that modular arithmetic is usually defined differently than we have done here, using equivalence classes. We will cover that later.

Symmetries Below is an equilateral triangle with its three vertices labeled.



We are interested in the *symmetries* of the triangle. Imagine the triangle is sitting on the ground and it has left an indentation on the ground. Symmetries are operations where we pick up the triangle and put it back so that it fits exactly in the indentation. Except for putting it back exactly as we found it, all of these symmetries will move vertices around. There are precisely six ways we can do this, shown below.

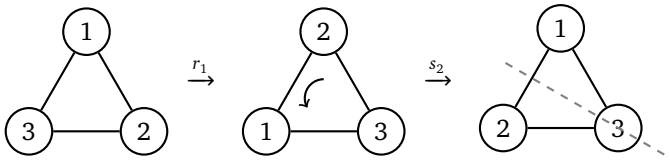


The first operation is where we put the triangle back exactly as we found it. The next two are rotations by 120° and 240°. Note that the first operation can also be seen as a rotation by 0°. The other three operations are reflections about various axes indicated by dashed lines. These are where we pick up the triangle and flip it over. The three reflections indicate the three ways to flip it. We'll use r_0, r_1, and r_2 to represent the rotations and s_1, s_2, and s_3 for the reflections.

An interesting sort of arithmetic arises if you compose these symmetries, that is if you do one followed by another. For instance, if you do r_1 twice in a row, you get r_2, which makes sense since two 120° rotations is a 240° rotation. We could write this as $r_1 \circ r_1 = r_2$. If you do r_1 three times in a row, you get back to r_0 since three 120° rotations is a 360° rotation, the same as a 0° rotation. Note that the rotation r_0 is an identity, since it has no effect on the shape, and composing it with any other symmetry just gives that symmetry.

Doing any reflection two times in a row takes us back to the identity since each reflection undoes itself. Basically, if you flip something over twice in a row, you end up back where you started.

We can also compose reflections and rotations together. For instance, we can do $s_2 \circ r_1$. The $\circ$ operation denotes function composition, since people usually think about symmetries as functions that act on the vertices of the triangle. So $s_2 \circ r_1$ means $s_2(r_1(T))$ where T is the triangle. This means that the right operation, r_1, is done first and then the left operation, s_2, is done after that. We see below that the result of doing these two operations is the same as s_1, so $s_2 \circ r_1 = s_1$.



With a little work, we can work out a table for what we get by composing all the different possible operations together. The result is below.

	r_0	r_1	r_2	s_1	s_2	s_3
r_0	r_0	r_1	r_2	s_1	s_2	s_3
r_1	r_1	r_2	r_0	s_2	s_3	s_1
r_2	r_2	r_0	r_1	s_3	s_1	s_2
s_1	s_1	s_3	s_2	r_0	r_2	r_1
s_2	s_2	s_1	s_3	r_1	r_0	r_2
s_3	s_3	s_2	s_1	r_2	r_1	r_0

Notice that r_0 acts as an identity. Notice also that every element has an inverse, where if we do that symmetry followed by its inverse, the result is as if we didn't move the triangle at all. Namely, the inverse of an element x is something to compose it with that results in the identity r_0. We have the r_0 is its own inverse, the reflections are also each their own inverses, and the inverse of r_1 is r_2 (and vice-versa). Notice also that this type of arithmetic is not commutative. The order we do the symmetries in matters. For instance, $s_2 \circ r_1 = s_1$, but $r_1 \circ s_2 = s_3$.

Chapter 2

Introduction to Groups

2.1 Definition of a group

In the preceding section, we saw a few different types of arithmetic. There are many others that we will look at throughout the course of these notes. But let's get to the definition of a *group*. A group is, roughly speaking, a set and an operation on items in that the set that follows a few useful properties that make it possible to do arithmetic and algebra on the items. Here is the formal definition.

Definition 2.1. *A* group *is a set G along with a binary operation $*$ satisfying four properties:*

1. Closure*: If $a, b \in G$, then $a * b \in G$.*
2. Associativity*: If $a, b, c \in G$, then $(a * b) * c = a * (b * c)$.*
3. Identity*: There exists a particular element $e \in G$, called an* identity*, with the property that for every $a \in G$, we have $a * e = a$ and $e * a = a$.*
4. Inverses*: For every $a \in G$, there exists an* inverse *of a, denoted a^{-1}, with the property that $a * a^{-1} = e$ and $a^{-1} * a = e$.*

If the operation is also commutative, that is if $a * b = b * a$ for every $a, b \in G$, then the group is called abelian.

It's usually tedious to specify the operation using the $*$ symbol, so most of the time we will just write ab instead of $a * b$. We will often refer to the group operation as a "product" or as "multiplication", even though it might be something quite different from ordinary multiplication. A lot of times, we will just refer to the group by G, but if it's necessary to specify the operation, we will use something like $(G, *)$.

Mathematicians like to keep definitions as concise as possible. These four properties are a pretty small set of properties that allow us to do useful algebra. Let's look briefly at those four properties. The closure property tells us that when we do the operation, the result has to stay in G. If doing ab gives us some element c that wasn't in G, then we wouldn't be able to work with c effectively. Associativity tells us that we can group things however we like and the result won't change. The identity is a value that leaves things unchanged when operating with it. The inverse of an element a is an element that undoes the result of a, bringing us back to the identity.

The last three rules are important since it's hard to get much done without them. For instance, suppose we want to solve the equation $ax = b$ for x. We can do this by multiplying both sides by a^{-1}. This turns the equation into $a^{-1}ax = a^{-1}b$. Use the associative property to write the left side as $\left(a^{-1}a\right)x$. We then use the inverse rule to rewrite $a^{-1}a$ as the identity e, giving us $ex = a^{-1}b$. Finally, we use the identity rule to rewrite ex as x, and thus the equation is solved as $x = a^{-1}b$. This simple process uses properties 2, 3, and 4. If we didn't have all of those properties, we couldn't solve simple equations, which would severely limit what we could do algebraically.[2]

[2] There are algebraic structures such as semigroups, monoids, and magmas, that only have some of the four properties, but not others. We won't cover them in these notes, but it's nice to know they exist in case you want to look them up somewhere.

Examples

Let's look at several examples of groups as well as several examples of things that are not groups. To show something is a group, we have to verify all four properties hold. To show something is not a group, we just have to show that one of the properties does not hold. This type of thinking applies to all definitions, by the way, not just groups.

Example 2.1. The integers with the addition operation. The symbol $\mathbb{Z}$ is usually used to denote the integers, and $(\mathbb{Z}, +)$ denotes the group of integers under addition. Usually when addition is used as the group operation, people write $a + b$ instead of ab.

To show this is a group, we need to verify all four properties of the definition hold. However, this gets a little tricky to do at this stage since we would have to properly define what the integers actually are and what addition actually is. This can be done through a set of axioms called the Peano axioms, but going through that would take a while, so we will just throw up our hands and take it as obvious from real-world experience that the sum of two integers is an integer and that addition is associative.

The identity is $e = 0$, and this satisfies the equations in property 3, namely $0 + a = a$ and $a + 0 = a$ for any integer a. The inverse of any integer a is $-a$ since $a + -a = 0$ and $-a + a = 0$. Again, we won't rigorously prove these facts, but we could if we really wanted to. Note also that $(\mathbb{Z}, +)$ is an abelian group because addition is commutative, namely $a + b = b + a$ for all integers a and b.

Example 2.2. The nonzero real numbers with the multiplication operation. People usually denote this group by $\mathbb{R}^*$. Like in the previous example, We'll take it as given that the real numbers are closed under multiplication and that multiplication is associative. It is possible to prove these facts rigorously, but it would take a fair bit of work to give a proper definition of the real numbers. The identity is 1 and the inverse of the real number x is $\frac{1}{x}$. All four properties of a group are satisfied. It is also an abelian group because multiplication is commutative.

Example 2.3. The positive integers with the multiplication operation (not a group). The positive integers with the multiplication operation are not a group because not every element has an inverse. In fact nothing except 1 has an inverse since if $n > 1$, there is no integer we can multiply n by to get 1.

Example 2.4. All real numbers with the multiplication operation (not a group). Here, almost every real number has an inverse, but there is one number, 0, that does not have an inverse. This small failure is enough to make all of $\mathbb{R}$ not a group under multiplication.

Example 2.5. Several examples. Here is a list of a few groups:

- $\mathbb{R}$ — All real numbers under addition
- $\mathbb{R}^*$ — All nonzero real numbers under multiplication
- $\mathbb{Q}$ — All rational numbers (fractions of integers like $2/3$, $9/4$, or $3/1$) under addition
- $\mathbb{Q}^*$ — All nonzero rational numbers under multiplication
- $\mathbb{C}$ — All complex numbers (numbers of the form $a + bi$ where $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$) under addition
- $\mathbb{C}^*$ — All nonzero complex numbers under multiplication

Example 2.6. The even integers with addition. The even integers, $(0, 2, -2, 4, -4,$ etc.$)$ are sometimes denoted as $2\mathbb{Z}$. They form a group under addition. To show this, we have to show the four properties are satisfied. To do this, it helps to have a proper definition of what it means to be even. In general, an integer is considered even if it is a multiple of 2, namely that it's of the form $2n$ for some $n \in \mathbb{Z}$.

First, the evens are closed under addition. To see this, note that if x and y are both evens, then $x = 2m$ and $y = 2n$ for some integers m, n by the definition of even numbers. Then $x + y = 2m + 2n = 2(m + n)$. This shows

$x + y$ is even because it is a multiple of 2, namely $2(m + n)$. Besides that, we know that addition is associative, 0 still works as the identity here, and every even integer $x = 2m$ has an inverse that is also in the set of evens (the inverse of $2m$ is $-2m$, which we can write as $2(-m)$, a multiple of 2). So the evens under addition satisfy all the properties of a group.

Example 2.7. Perfect squares and their negatives under addition (not a group). We are looking at the set $\{\ldots, -16, -9, -4, -1, 0, 1, 4, 9, 16, \ldots\}$. This satisfies associativity, identity, and inverses, but it is not closed. For instance, 1 and 4 are perfect squares, but $1 + 4 = 5$, which is not a perfect square.

Example 2.8. The set $\{1, -1, i, -i\}$ under multiplication. Here i is the famous imaginary number $\sqrt{-1}$. It satisfies $i^2 = -1$. There are 16 possible ways to multiply two items from the set, and it's not hard to check them all to see that the result always gives something in the set, so we do have closure. See the table below. We have associativity because multiplication of complex numbers is associative. The identity is 1. Each item does have an inverse in the group. In particular, 1 is its own inverse, as is -1 (since $-1 \times -1 = 1$), and we have that i and $-i$ are inverses of each other since $i \times -i = -i^2 = -(-1) = 1$. Here is a multiplication table (called a *Cayley table*) showing the result of the group operation for all 16 possible combinations.

	1	-1	i	$-i$
1	1	-1	i	$-i$
-1	-1	1	$-i$	i
i	i	$-i$	-1	1
$-i$	$-i$	i	1	-1

Example 2.9. Integers modulo n with addition. Earlier, we looked at arithmetic mod n, where we add two numbers and then mod the result by n. We will use $\mathbb{Z}_n$ to denote the set $\{0, 1, 2, \ldots, n-1\}$ with addition done mod n. You might see this group written as $\mathbb{Z}/n\mathbb{Z}$ in other sources. This does satisfy all the group properties. Since we mod by n, we will always get a result in the range from 0 to $n-1$, so we have closure. Even with the mod operation, this form of addition is still associative, though we won't prove it. The identity is 0, and the inverse of any element a is $n - a$.

As an example, $\mathbb{Z}_7$ uses integers 0, 1, 2, 3, 4, 5, and 6. Here is a Cayley table for it.

	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	2	3	4	5	6	0
2	2	3	4	5	6	0	1
3	3	4	5	6	0	1	2
4	4	5	6	0	1	2	3
5	5	6	0	1	2	3	4
6	6	0	1	2	3	4	5

Example 2.10. (Some) integers modulo n with multiplication. If we take the integers $1, 2, \ldots, n-1$ and use multiplication mod n, sometimes we get a group and sometimes we don't. The closure and associative properties still hold, and 1 is the identity. The problem comes with inverses. Depending on the value of n, sometimes everything will have an inverse and sometimes not.

For instance, when $n = 5$, we do get a group. The inverse of 1 is itself, the inverse of 2 is 3 (since $2 \times 3 = 6$ and $6 \bmod 5 = 1$), and the inverse of 4 is itself (since $4 \times 4 = 16$, and $16 \bmod 5 = 1$).

However, when $n = 6$, we don't get a group. This is because 2, 3, and 4 do not have inverses. There is nothing to multiply 2 by that will give a result of 1. In particular, the multiples of 2 modulo 6 are 2, 4, and 0.

The key difference is that 5 is prime and 6 isn't. It turns out that any value that shares common factors with n cannot have an inverse. The reverse is also true in that if n and a share no factors in common, then a does have a multiplicative inverse mod n. These are not hard to prove, but we won't do so here.

The upshot is that whenever n is prime, the integers $\{1, 2, \ldots, n-1\}$ form a group under multiplication mod n. If n is not prime, all the integers from 1 to $n-1$ that share no common factors with n form a group, called the

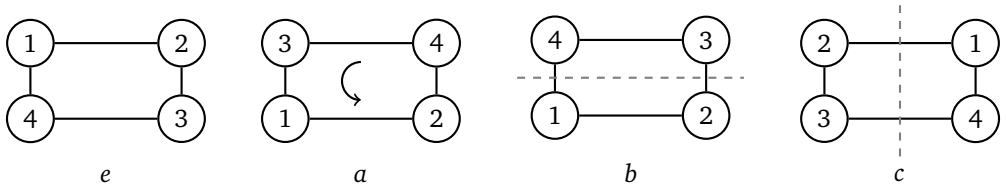
group of units mod n. It is denoted by U_n. For example, below is the Cayley table of the group U_{10} of integers relatively prime to 10 under multiplication.

	1	3	9	7
1	1	3	9	7
3	3	9	7	1
9	9	7	1	3
7	7	1	3	9

Example 2.11. Dihedral groups. Earlier, we looked at the symmetries of a triangle. There are six of them, and we looked at the Cayley table. Those six symmetries form a group where the operation is function composition. The Cayley table is shown again below. Notice that we have closure since every composition of symmetries results in a symmetry. Function composition is associative, though we won't prove it. The table can be used to check that r_0 is the identity and that every element does have an inverse. This group is called D_3, the *dihedral group* of symmetries of an equilateral triangle (3-sided polygon). In a similar way, the symmetries of larger regular polygons also form groups. These groups are not abelian, since, as we saw earlier, the order we do the symmetries in does matter. We will investigate dihedral groups in more detail later on in these notes.

	r_0	r_1	r_2	s_1	s_2	s_3
r_0	r_0	r_1	r_2	s_1	s_2	s_3
r_1	r_1	r_2	r_0	s_2	s_3	s_1
r_2	r_2	r_0	r_1	s_3	s_1	s_2
s_1	s_1	s_3	s_2	r_0	r_2	r_1
s_2	s_2	s_1	s_3	r_1	r_0	r_2
s_3	s_3	s_2	s_1	r_2	r_1	r_0

Example 2.12. Klein four-group. This is a group, usually denoted V, is named for Felix Klein, who helped develop a lot of what became known as group theory. It is the group of symmetries of a non-square rectangle. There are four elements. One is the identity e, which leaves the rectangle in place. The next is a, which is a rotation by 180°. The other two are b and c, which are reflections about horizontal and vertical lines, respectively, through the center of the rectangle. See below.



We can work out a Cayley table for composing these symmetries, shown below.

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

This is a group since the four properties are satisfied. In particular, the table shows the operation is closed. The operation of composing two symmetries is associative since function composition is associative. The identity is e, and each element has an inverse, namely itself. That is, doing a 180° rotation twice in a row or doing a reflection twice in a row leads us back to the original orientation of the sides. We can see this via the diagonal line of e's in the table.

Example 2.13. Another operation defined by a Cayley table (not a group). What if we try creating our own Cayley table? Will the result necessarily be a group? Here is one try below.

	e	a	b	c
e	e	a	b	c
a	a	e	c	c
b	b	c	e	c
c	c	c	c	e

Like with the Klein four-group, e is the identity and each element is its own inverse. However, associativity fails. In particular, look at the expression aab. If we group the a's together, we have $(aa)b = eb = b$, but if we group the middle a with the right b, we have $a(ab) = ac = c$. It actually turns out to be pretty tricky to create a group this way. As we'll see later, for groups with four elements, there are only two possible Cayley tables that turn out to give groups. One is the Klein four-group, and the other is $\mathbb{Z}_4$. Its table looks like below if we set $e = 0$, $a = 1$, $b = 2$, and $c = 3$.

	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

Example 2.14. Direct products. If $(G, *)$ and $(H, \diamond)$ are both groups, we can build a group from the Cartesian product $G \times H$, called the *direct product*. The elements of $G \times H$ are elements of the form (g, h), with $g \in G$ and $h \in H$. The group operation on (g_1, h_1), (g_2, h_2) is given by $(g_1 * g_2, h_1 \diamond h_2)$. That is, we do the G operation on the first part of the pair and the H operation on the second part of the pair.

Closure follows from the fact that G and H are both closed and that $G \times H$ contains all possible pairs. Associativity follows directly from the fact that G and H have that property. The identity is (e_G, e_H), where e_G and e_H are the identities in their respective groups. The inverse of (g, h) is (g^{-1}, h^{-1}). Here are a few examples.

1. Take $\mathbb{Z}_5 \times V$, where $V = \{e, a, b, c\}$ is the Klein four-group. A few typical elements are $(0, a)$, $(2, b)$, and $(4, e)$. The group operation on $(3, a)$ and $(4, b)$ is $(3, a)(4, b) = (3 + 4, ab) = (2, c)$.
2. The direct product $\mathbb{Z}_2 \times \mathbb{Z}_3$ is a group with 6 elements $(0, 0)$, $(0, 1)$, $(0, 2)$, $(1, 0)$, $(1, 1)$, and $(1, 2)$. The group operation is $+$ for both groups, and we can write $(0, 2) + (1, 2) = (0 + 1, 2 + 2) = (1, 1)$.
3. The direct product $\mathbb{U}_{10} \times \mathbb{Z}_3$ is a group with 12 elements, $(1, 0)$, $(1, 1)$, $(1, 2)$, $(3, 0)$, $(3, 1)$, $(3, 2)$, $(7, 0)$, $(7, 1)$, $(7, 2)$, $(9, 0)$, $(9, 1)$, and $(9, 2)$. The group operation is multiplication mod 10 for the first half and addition mod 3 for the second. A typical operation would be $(7, 1)(9, 2) = (7 \cdot 9, 1 + 2) = (3, 0)$.
4. For an infinite example, consider $\mathbb{Z} \times \mathbb{R}^*$. There are infinitely many items in this direct product. Some examples include $(2, 3.7)$, $(-2, \pi)$, and $(47, \sqrt{2})$. Here $\mathbb{Z}$ uses the addition operation and $\mathbb{R}^*$ uses multiplication. An example operation would be $(2, 3.9)(10, 2.21) = (2 + 10, 3.9 \cdot 2.21) = (12, 8.619)$.

Example 2.15. Matrices We aren't assuming familiarity with linear algebra in these notes, so we won't use many matrix examples. However, for those readers that have seen some linear algebra, matrix arithmetic gives examples of groups. For example, the set of all $n \times n$ matrices with real entries form a group. A very common group in higher mathematics is the *general linear group* $\mathrm{GL}(n, \mathbb{R})$ of nxn invertible matrices with real entries. Another important matrix group is the *special linear group*, $\mathrm{SL}(n, \mathbb{R})$ of $n \times n$ (invertible) matrices with determinant 1 and real entries. Both of these are examples of infinite non-abelian groups.

Notational notes Whenever we refer to $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, addition will be the assumed operation. We will use $\mathbb{R}^+$ to refer to the positive real numbers under multiplication and $\mathbb{R}^*$ to refer to the nonzero real numbers under multiplication. The group $\mathbb{Z}_n$ will always use addition. When we want modular multiplication, we will use the group U_n of integers relatively prime to n.

2.2 Basic properties of groups

In this section we'll start to understand what groups are and how to work with them.

One of the parts of the definition of a group is that there is an identity element e with the property that $ea = a$ and $ae = a$ for every element a in the group. That is, doing the operation with e essentially has no effect. A natural question is if there could be more than one element in the group with this property. In other words could a group have two identities? The answer is no.

Proposition 2.1. *The identity in a group is unique.*

Proof. To prove this, we will assume that e_1 and e_2 are both identities and show it must be true that $e_1 = e_2$. Since e_1 is an identity, $e_1 a = a$ for any a in the group. If we plug in $a = e_2$, this tells us $e_1 e_2 = e_2$. Next, since e_2 is an identity, $ae_2 = a$ for any a in the group. Plugging in $a = e_1$ gives $e_1 e_2 = e_1$. So we have $e_1 e_2 = e_2$ and $e_1 e_2 = e_1$. Therefore, $e_1 = e_2$. □

To show something is unique, a proof like this is often used. We assume that two things exist and show they must be equal to each other. These proofs are also sometimes done by contradiction, where we assume two different things exist and arrive at a contradiction. As another example of uniqueness, let's show that inverses are unique.

Proposition 2.2. *Each element in a group has a unique inverse.*

Proof. The definition of a group guarantees that each element a has an inverse. To show the inverse is unique, suppose that x and y are both inverses of a. We need to show that $x = y$. Since x and y are both inverses of a, we have $xa = e$ and $ay = e$. Using these, we can write $x = xe = x(ay) = (xa)y = ey = y$. □

The proof above relies on a bit of a dirty trick: rewriting x as xe. Dirty tricks like this are useful in algebra proofs, and it's good to build up a toolbox of these tricks. The reason this trick is helpful is that it introduces the identity e into the picture. Then we make use of the fact that y is an inverse of a to get $ay = e$, which we then plug into our formula. Associativity then lets us regroup $x(ay)$ into $(xa)y$, and after that we take advantage of the fact that x is an inverse of a to replace xa with e. At the last step, we can say $ey = y$ since e is the identity. Reading the expression $x = xe = x(ay) = (xa)y = ey = y$ from left to right allows us to see that x is equal to y.

This proof is a little slick. When you're first learning to prove things, you probably won't end up with something this efficient. It takes some work, false starts and failed attempts, before you get something that works. Once you have something that works, look for places to trim down extraneous steps to get something efficient like this. It's good to make proofs efficient to make them easier to follow, but just like with writing, if you make things too efficient, it can be challenging for readers to follow.

It's worth noting at this point that the two facts we have proved apply to *every* group. The fact that there is only one identity in a group means that there is only one identity matrix, there is only one identity symmetry, there is only one identity element in modular arithmetic, etc. Likewise, each matrix has exactly one inverse, each symmetry has exactly one inverse, etc. We prove this facts abstractly, for groups in general, and then the result holds for any objects that satisfy the group properties. That is the power of abstract algebra.

Next, we have a proposition that is helpful when working with larger expressions:

Proposition 2.3. *The associative law works for expressions of any size. That is, no matter how we insert parenthesis into an expression, the result will be the same.*

The proof of this relies on (strong) induction, but it is a little tedious, so we will skip the details. Here is a little idea of how it works. Suppose we have the expression $abcd$ and we want to show that $(ab)(cd) = a(b(cd))$. Let $x = cd$. Then $(ab)(cd) = (ab)x$. By the associative law from the definition of a group, $(ab)x = a(bx)$. Plugging $x = ab$ back into this gives $a(bx) = a(b(cd))$, which is what we want. Similar arguments like this allow us to show that associativity works for all the ways of parenthesizing $abcd$. We can then move to $abcde$ and use the fact that we can parenthesize expressions of size 3 or 4 however we like to show that it also works for expressions of 5 things. Strong induction can be used to make this formal for expressions of whatever length.

As we know, algebra relies a lot on solving equations. Here is a proposition about the cancelation property we use in solving equations:

Proposition 2.4. *Cancelation works in a group. In particular, if we have $ax = ay$, then we must have $x = y$. Likewise, if $xa = ya$, then $x = y$.*

Proof. Starting with $ax = ay$, multiply both sides of the equation by a^{-1} on the left. This gives $a^{-1}ax = a^{-1}ay$. By associativity, we can group this as $(a^{-1}a)x = (a^{-1}a)y$. By the definition of inverses, $a^{-1}a = e$, so we have $ex = ey$. By the definition of identities, we have $ex = x$ and $ey = y$, so $x = y$, as desired. A similar proof applies if we start with $xa = ya$, except that we need to multiply on the right instead of on the left to start. □

Note how the proof uses associativity, identity, and inverses, which are three key properties of a group. When people define things in math, they like to keep the definitions as small as possible. Associativity, identity, and inverses are sort of the minimum you need to have in order to get useful stuff done, as we see in the proof above.

It's worth comparing this to how things would work in high school algebra. Given $ax = ay$, we would divide both sides by a. This would cancel the a's on both sides to leave us with $1 \cdot x = 1 \cdot y$. Then, since 1 times anything is itself, we get $x = y$. When working with groups, instead of division, we have multiplication by the inverse. This is a little like multiplying by $1/a$ in high school algebra. Also, notice that the identity e of a group behaves just like the number 1 does in high school algebra.

Note also that there are two parts to the proof, $ax = ay$ and $xa = ya$. Unless the group is commutative, order matters, so we have to worry about whether a is on the left or the right in the proof.

Here is a useful fact with a similar proof to the one above.

Proposition 2.5. *The equation $ax = b$ has the unique solution $x = a^{-1}b$.*

Proof. If we multiply both sides of the equation by a^{-1} on the left, then we get $a^{-1}ax = a^{-1}b$. We can use associativity to group the left side into $(a^{-1}a)x$. Then by the definition of inverses, this becomes ex, and by the definition of the identity, $ex = x$. Thus, $x = a^{-1}b$. This is the only solution to the equation because a^{-1} itself is unique. □

This technique of multiplying both sides by an inverse to essentially cancel out a term shows up all over the place in proofs. It gets tedious to constantly mention associativity, identity, and inverse properties like we do in the proof above, so in the future, we will not. This technique of multiplying by an inverse is just like in ordinary algebra when you divide both sides of an equation by something to cancel things out. Instead of dividing, here we multiply by an inverse. If the group you're working with is not necessarily commutative, then be careful which side you multiply on. For instance, multiplying ax on the left by a^{-1} will cancel out the a to give x, but it won't work to multiply on the right. We would just get axa^{-1}, which we can't reduce unless the group is commutative.

The previous proposition is also useful when trying to find the inverse of an element. According to the definition of inverses, to show that x is the inverse of a, we would need to show $ax = e$ and $xa = e$. However, the previous proposition implies that we can just show one or the other. For instance, if we know $ax = e$, then the proposition says we can solve this to get $x = a^{-1}$.

Example 2.16. As a quick example of multiplying by inverses, suppose we want to solve $abcd^{-1}f = gh$ for c. We can start by left-multiplying by a^{-1} to get $bcd^{-1}f = a^{-1}gh$. Then left-multiply by b^{-1} to get $cd^{-1}f = b^{-1}a^{-1}gh$. Next, right-multiply by f^{-1} to get $cd^{-1} = b^{-1}a^{-1}ghf^{-1}$. Finally, right-multiply by d to get $c = b^{-1}a^{-1}ghf^{-1}d$.

Here is a hopefully obvious property of inverses, that the inverse of an inverse is the original element:

Proposition 2.6. *For any element a of a group, we have $\left(a^{-1}\right)^{-1} = a$.*

Proof. The equation is just saying that the inverse of a^{-1} is a. The inverse of a^{-1} is a value x that satisfies $a^{-1}x = e$. Multiplying both sides by a on the left gives $x = a$. □

Here is another important proposition that uses multiplication by inverses in its proof:

Proposition 2.7. *For any elements a and b of a group, we have $(ab)^{-1} = b^{-1}a^{-1}$.*

Proof. Let x be the inverse of ab. That means $(ab)x = e$. Left-multiply by a^{-1} to get $bx = a^{-1}$. Then left-multiply by b^{-1} to get $x = b^{-1}a^{-1}$. □

Notice how the order flips. It will only be true that $(ab)^{-1} = a^{-1}b^{-1}$ if the group is abelian. The fact that the order reverses is something that often trips up beginning students since it runs counter to how things work in high school algebra. The rule generalizes to more than two terms. For instance, $(abc)^{-1} = c^{-1}b^{-1}a^{-1}$. Below is the general rule, which we prove by induction.

Proposition 2.8. *For any elements $a_1, a_2, \ldots, a_n$ in a group, we have $(a_1a_2\ldots a_n)^{-1} = a_n^{-1}a_{n-1}^{-1}\ldots a_1^{-1}$.*

Proof. The base case is $n = 2$ terms, which we just proved. Now assume that the result is true for n terms, namely that $(a_1a_2\ldots a_n)^{-1} = a_n^{-1}a_{n-1}^{-1}\ldots a_1^{-1}$. Now consider $n+1$ terms, $(a_1a_2\ldots a_na_{n+1})^{-1}$. We can write it as $(xa_{n+1})^{-1}$ where $x = a_1a_2\ldots a_n$. Using the $n = 2$, case we have $(xa_{n+1})^{-1} = a_{n+1}^{-1}x^{-1}$ and then our induction hypothesis tells us what x^{-1} equals. Plugging that in gives $(a_1a_2\ldots a_na_{n+1})^{-1} = a_{n+1}^{-1}a_n^{-1}a_{n-1}^{-1}\ldots a_1^{-1}$, as desired. □

People first learning induction can have trouble with the concept of what it accomplishes. The reasoning almost seems circular since it seems like we are assuming the thing we are trying to prove. Induction is actually about assuming something works for n and showing if that is true then it also works for $n+1$. As a specific example, suppose we are going from $n = 3$ case to the $n = 4$ case in the statement above. That is, assuming $(a_1a_2a_3)^{-1} = a_3^{-1}a_2^{-1}a_1^{-1}$, we want to show that $(a_1a_2a_3a_4)^{-1} = a_4^{-1}a_3^{-1}a_2^{-1}a_1^{-1}$. To do this, we write $(a_1a_2a_3a_4)^{-1} = ((a_1a_2a_3)a_4)^{-1}$ by grouping the first three terms. Letting $x = a_1a_2a_3$, the expression becomes $(xa_4)^{-1}$. We use the rule for two terms to make this $a_4^{-1}x^{-1}$. And since $x = a_1a_2a_3$ and we know the rule works for 3 terms, we get $x^{-1} = a_3^{-1}a_2^{-1}a_1^{-1}$. We plug this in and get $(a_1a_2a_3a_4)^{-1} = a_4^{-1}a_3^{-1}a_2^{-1}a_1^{-1}$. Now that we know the rule works for $n = 4$ terms, we can use the same argument to extend it to $n = 5$ terms. Then the same argument can extend the formula to 6 terms, 7 terms, etc. The induction proof just does this once for a general n, showing how to get from it working for n terms to it working for $n+1$ terms.

Proofs by induction are useful in situations like this, where we have a rule that works for two terms and we want to extend it to an arbitrary number of turns. They are useful in many other places, too. The key idea is usually to find the way to turn the $n+1$ case into the n case. A common technique is to group the first n terms, like above. Some other useful things for these breakdowns are $x^{n+1} = x \cdot x^n$, $(n+1)! = n \cdot n!$ and $x_1 + x_2 + \ldots x_{n+1} = (x_1 + x_2 + \ldots x_n) + x_{n+1}$.

Powers

In ordinary arithmetic, powers are defined as repeated multiplication. For instance, x^2 is just x times x, x^3 is x times itself three times, etc. Negative powers are defined in terms of reciprocals. For instance, x^{-2} is $1/x^2$. Finally, x^0 is 1. A similar thing works for groups, just that we use inverses instead of reciprocals, and 1 is replaced with the identity. Here is the formal definition.

Definition 2.2. *Let a be an element of a group and let n be a positive integer. We define a^n as $aa\cdots a$, which is a multiplied by itself n times. We define a^{-n} as $a^{-1}a^{-1}\cdots a^{-1}$, which is a^{-1} multiplied by itself n times. Finally, a^0 is defined to be the group's identity, e.*

The rules of exponents from ordinary algebra work with groups.

Proposition 2.9. *If a is an element in a group and m and n are integers, then $a^na^m = a^{m+n}$ and $(a^n)^m = a^{mn}$.*

The proofs are not hard, but we will skip them as they are not particularly illuminating. They just rely on associativity and the definitions, but it gets a little tedious to go through all the possible cases for m and n being positive, negative, or 0.

We need to be careful with $(ab)^n$. For instance, $(ab)^2$ can be rewritten as $abab$, but not necessarily as a^2b^2 or b^2a^2 unless the group is abelian.

Order

The following concept shows up extensively.

Definition 2.3. *The* order *of an element a in a group is the smallest positive integer n such that $a^n = e$. If no such integer exists, then the order is considered infinite.*

As an example, consider the group U_7, which has elements 1, 2, 3, 4, 5, 6. Here are the powers of all the elements, where we stop showing powers once we get a 1.

	a	a^2	a^3	a^4	a^5	a^6
1	1					
2	2	4	1			
3	3	2	6	4	5	1
4	4	2	1			
5	5	4	6	2	3	1
6	6	1				

We see that 1, the identity, has order 1. The identity of a group always has order 1 and it's always the only order 1 element. We see that 2 and 4 have order 3 and that 3 and 5 have order 6.

As another example, look at the additive group Z_{10}. The operation here is $+$, and a power in a group is just repeatedly applying the operation. So 2^3 actually means $2+2+2$ here. This takes a little getting used to. To find the orders of the elements, we need to find how many times we need to add them until first get to a multiple of 10. For 1, the order is 10, since $1+1+1+1+1+1+1+1+1+1 = 10$. For 2, the order is 5 since $2+2+2+2+2 = 10$. For 3, the order is 10 again, since $3+3+3+3+3+3+3+3+3+3 = 30$, which is a multiple of 10. The orders of all the elements are given in the table below.

element	0	1	2	3	4	5	6	7	8	9
order	1	10	5	10	5	2	5	10	5	10

If the group is finite, then every element has a finite order. That is, if you multiply an element by itself enough times, eventually you end up with the identity. This might be a little surprising. We state this as a proposition and prove it below.

Proposition 2.10. *If a group has finitely many elements, then every element has finite order.*

Proof. Let a be an element of the group and look at a, a^2, a^3, a^4, etc. By closure, each of those is an element of the group. Because the group is finite, eventually there will be repeats among the powers of a. That is, we must have $a^i = a^j$ for some unequal i and j. Assume i is the smaller of the two powers. Multiplying both sides of this equation by a^{-i} gives $a^{-i}a^i = a^{-i}a^j$ and thus $a^{j-i} = e$. This shows that some power of a must be the identity. □

2.3 Proving things

Because proofs are such an important part of most abstract algebra courses, it's worth spending a little time thinking about how to prove things.

There is no general algorithm to follow for proving things. Each problem is its own thing. For the proofs we will be considering, the primary tools to use are definitions, basic algebra rules, and theorems. Sometimes, all that is needed is to tie some definitions together using a little bit of algebra and the occasional theorem.

However, sometimes cleverness is needed. A lot of proofs require some big or clever idea to work. This can take a long time to find. With enough practice, you get quicker at finding these clever ideas, since variations of things

that worked in other contexts might work for your specific problem. Often your first several tries at a proof will not lead anywhere. Keep trying, and if you find yourself stuck, take a break and come back to the problem later. When you walk away from a problem, your mind might be working on the problem in the background, and when you come back to it, you might find you can solve it quickly. Fresh eyes are also helpful because when you stare at something too long, you can easily miss something or get stuck in a rut. Coming at it fresh allows you to move past these things. It's similar to writing something and reading it back over – if you read it right away, you might miss typos, but when you come back to it, you will often catch things your mind skipped over earlier.

It's also helpful to talk things out with other people. Sometimes, the very act of trying to put something into words is enough to help you clarify the problem you're having, enough that you can get past a sticking point.

Though there is no general rule for what proof technique to use for any situation, there are a few things that are helpful.

1. Many proofs rely just on carefully verifying a definition. Start those by carefully going through the definition. Here are some examples.
   (a) If you need to show something is a group, you have to verify the closure, associative, identity, and inverse properties hold.
   (b) To show a set is closed under an operation $*$, always start by picking two arbitrary elements a and b in the set and show that $a * b$ is also in the set.
   (c) Any time you need to show a group G is abelian, start by letting a and b be two arbitrary elements in G. The goal is then to show that $ab = ba$.
   (d) Any time you need to show the order of g is n, there are two things to be done: (1) show that $g^n = e$ and (2) show that $g^k \neq e$ for any $0 < k < n$.
   (e) If you want to show that x is the inverse of y, show that $xy = e$. (Note that the definition also has $yx = e$, but a remark following Proposition 2.5 implies showing that is redundant.)
2. Induction proofs usually show up in specific scenarios, like when you want to show something is true for any number of times. Examples we have seen are showing the associative rule works for any number of elements and showing the rule $(ab)^{-1} = b^{-1}a^{-1}$ works for products of any number of items.
3. Proof by contradiction is a handy tool that applies in many situations. It is particularly appropriate when the thing to be proved involves a negative. For instance, when showing the order of an element is n, part of the proof is in showing that *no* positive power less than n works. This is a negative sort of thing, so contradiction fits naturally.

   Contradiction is also useful if you want to show two things are *different*. You can assume they are the same, and show that leads to a contradiction somewhere. By assuming things are the same, you get an equation to work with.

   In places where proofs by contradiction work, proving the contrapositive also often works.
4. The topic of these notes is algebra, and many times a direct proof will just involve doing a bunch of algebraic manipulations. Sometimes dirty tricks are helpful, so it's good to build up a toolbox of such tricks.
5. As a general rule, if a statement includes a certain hypothesis, that hypothesis should be used in the proof somewhere. For instance, consider this statement: *If G is abelian, then* $(ab)^2 = a^2b^2$. The proof of this should use the fact that G is abelian. If it doesn't, there are two possibilities: (1) you made a mistake somewhere, (2) the hypothesis wasn't actually needed. Of these, (1) is much more likely. When you're working on your own research, (2) can happen, but when you doing textbook exercises, chances are the person who made the exercises put the hypothesis there for a reason.

To help get some more practice proving things, let's look at a few typical examples.

Example 2.17. *Prove that a group is abelian if and only if $(ab)^{-1} = a^{-1}b^{-1}$ for every a and b in the group.*

Solution: This is an "if and only if" statement. That means it is two statements in one. One of those is that we have to prove that if a group is abelian, then $(ab)^{-1} = a^{-1}b^{-1}$ for every a and b in the group. This is sometimes called the forward implication. The other statement to prove is that if the formula is true, then the group is abelian. This is sometimes called the reverse implication.

Let's start with the forward implication. Assume the group is abelian. By Proposition 2.7, for every a and b in the group, we have $(ab)^{-1} = b^{-1}a^{-1}$. Because the group is abelian, we can swap around the order of the right side into $a^{-1}b^{-1}$. Therefore, $(ab)^{-1} = a^{-1}b^{-1}$, as desired.

Now, for the reverse implication, assume the formula holds for all pairs of elements in the group. Let a and b be any two group elements. To show the group is abelian, we need to show that $ab = ba$. Look at $(a^{-1}b^{-1})^{-1}$. Plugging this into the formula we are assuming is true gives $(a^{-1}b^{-1})^{-1} = (a^{-1})^{-1}(b^{-1})^{-1}$. The right side simplifies into ab. Next, try using $(a^{-1}b^{-1})^{-1}$ with Proposition 2.7. This gives $(a^{-1}b^{-1})^{-1} = (b^{-1})^{-1}(a^{-1})^{-1}$. The right side here simplifies into ba. Therefore, we have $ab = ba$ since both are equal to $(a^{-1}b^{-1})^{-1}$.

Let's talk a little about the strategy behind the proof. First, when doing any proof, it's unspeakably important to be familiar with definitions. Here, the main definition to work with is *abelian*. A group is abelian if $ab = ba$ for every a and b in the group. In proving the forward implication, we need to use that definition in our algebraic manipulations, and in the reverse implication, we need to show that the definition is satisfied.

The forward implication is fairly straightforward as we just use the given formula and then use the abelian property to flip around the order. If and only proofs often work like this, where one direction is easy and the other is trickier. The backward implication takes some thought. When you're working on something like this, you'll find it takes several tries before you get something that works. When you see a proof in a text, it often looks so simple and logical, but what's hiding is all the false starts and dead ends that went into finding that simple proof, as well as all the work that went into trimming it down into something short and efficient.

Example 2.18. *Prove that for every element a in a group, the order of a is the same as the order of a^{-1}.*

Solution: As mentioned above, definitions are critically important. The definition of the order of an element a is that it's the smallest positive integer n such that $a^n = e$. The two key parts of this definition are the formula $a^n = e$ and the fact that n is the *smallest* positive value that works. To show something has a specific order, we need to show that the formula is satisfied and that we've found the smallest possible value.

We will start by assuming that a has order n and use that to show that a^{-1} also has order n. First, note that using rules of exponents, $(a^{-1})^n = a^{-n} = (a^n)^{-1}$. Since a has order n, we know that $a^n = e$, and hence $(a^n)^{-1} = e^{-1} = e$. Thus, $(a^{-1})^n = e$. Now we have to show that there is no smaller value that works. We will do this by contradiction. We will assume that there is some positive $k < n$ with $(a^{-1})^k = e$. The same computation we just did shows that $(a^{-1})^k = a^{-k} = (a^k)^{-1}$. So we have $(a^k)^{-1} = e$. Multiplying both sides by a^k tells us that $a^k = e$. But this is a contradiction because the smallest positive power of a that equals the identity is a^n, and $k < n$. Thus n must be the smallest power of a^{-1} that gives the identity.

Proofs by contradiction work by assuming the thing you want to prove is not true and reasoning from there until you get to something that is not true. Since you get yourself into an impossible situation, that means your initial assumption has to be wrong, meaning the thing you want to prove actually is true. Not all mathematicians are fans of proofs by contradiction. They can usually be rewritten into a direct proof if needed. I often find the contradiction approach to be simpler and more natural.

Just as a quick real world example of a proof by contradiction, suppose you are looking for your the keys to your home and you want to prove to yourself that they are not in your car. You might think to yourself, if I had left them in my car, then I would not have been able to get into my home yesterday because the home was locked and no one was there. Here the contradiction part is you assume they are in the car, and then the contradiction is the impossible situation that you didn't get back into your home, when clearly you are there. This is just to show that proof by contradiction is a technique that we use all the time without even realizing it.

Example 2.19. *Prove or disprove: In an infinite group, all elements besides the identity must have infinite order.*

Solution: A lot of times when a textbook exercise says "prove or disprove", they really mean disprove (though

not always). Here we will disprove the statement. To prove something, we need a logical argument why it is always true. To disprove something, we just need a single counterexample.

Our counterexample is the group of nonzero real numbers under multiplication. Most elements, such as 5, have infinite order, since no positive power of 5 can equal the identity 1. However, -1 has the property that $(-1)^2 = 1$. Therefore it has order 2.

2.4 Subgroups

You might be familiar with the concept of a subset of a set. A set is a collection of things, like $\{1,2,3,4,5\}$. A subset is a part of that set, like $\{1,2\}$ or $\{2,3,5\}$. This "sub" idea appears all over math, and group theory is no exception. Here is an important definition.

Definition 2.4. *A subgroup of group $(G,*)$ is a nonempty subset H of the elements of G such that $(H,*)$ is also a group.*

In other words, if we only take some of the elements of the group and use the same operation as the bigger group, and the result still satisfies all the group axioms, then we have a subgroup. As we'll see, we can't just take any elements. Only very carefully choosing the elements will produce a subgroup.

For example, if we take the group $\mathbb{Z}_6$ under (modular) addition, the subset $\{0,2,4\}$ forms a subgroup. The "sub" part comes from the fact that 0, 2, and 4 are elements of the parent group $\mathbb{Z}_6$. The "group" part comes from the fact that $\{0,2,4\}$ under modular addition satisfies all four properties of a group. In particular, it is closed since modulo 6, we have $0+2=2+0=2$, $0+4=4+0=4$, $2+2=4$, $2+4=4+2=0$, and $4+4=2$, showing that all combinations stay within $\{0,2,4\}$. The operation is the same as in the parent group, so it is still associative. The identity 0 is in the subgroup. And each item in the subgroup has an inverse in the subgroup, namely that 0 is its own inverse and 2 and 4 are inverses of each other.

On the other hand, $\{1,3,5\}$ is not a subgroup of $\mathbb{Z}_6$. There are a couple of reasons why. First, it is not closed, since $1+3=4$ and 4 is not in the subgroup. Also, it does not contain the identity. It does at least satisfy the associative and inverse properties, but that doesn't matter since it fails the closure and identity properties.

To show something is a subgroup, we typically show that the subgroup is closed, contains the identity, and contains the inverses of every element of the subgroup. It's not necessary to check associativity since that comes from the fact that the operation is associative in the parent group.

An important note about showing something is a subgroup If we pick a few random elements from a group, most of the time they won't form a subgroup. Usually there needs to be some special structure in order for a subset to have a chance of being a subgroup. That is, there is some property that all the elements of the group satisfy. For instance, the set of even integers turns out to be a subgroup of $\mathbb{Z}$. The property defining this subset is evenness.

As another example, given any group, the set of all elements of order 2 in that group, along with the identity, forms a subgroup of any group. The main property that all of those elements (except the identity) share is that they have order 2.

Here is more on each of the properties we need to show, keeping in mind that there is usually some property defining the subset we are trying to show is a subgroup.

1. **Closure**: We have to take two things, a and b, in the subset and show ab is in the subset. That is, assuming a and b have the property defining the subgroup, we have to show ab also has that property.
2. **Identity**: We need to take the identity from the parent group and show that it has the property that defines the subset.
3. **Inverses**: We have to pick an element a in the subset, i.e. having the subset's defining property, and show that a^{-1} also has the property.

Some examples

Example 2.20. Proper and trivial subgroups. The entire group itself is always technically a subgroup of itself, but we often want to exclude it from consideration. The term *proper subgroup* refers to any subgroup that is not the entire group. Also, the set just containing the identity is always a subgroup. This subgroup is called the *trivial subgroup*.

Example 2.21. Find all the subgroups of $\mathbb{Z}_6$. The trivial subgroup $\{0\}$ and $\mathbb{Z}_6$ are two subgroups. Another one is $\{0,2,4\}$, which we saw earlier. There is one more: $\{0,3\}$. It's closed because adding 0 and 0, 0 and 3, or 3 and 3 always results in 0 or 3 (modulo 6). We have the identity 0 in the group, and 0 and 3 are each their own inverses.

There are no other subgroups. To show why, first suppose 1 is in the subgroup. Then by closure, $1+1=2$, $1+1+1=3$, $1+1+1+1=4$, and $1+1+1+1+1=5$ are all in the subgroup. This means the only subgroup containing 1 is the entire group $\mathbb{Z}_6$. Similarly, if the group contains 5, then note that $5+5$, $5+5+5$, $5+5+5+5$, and $5+5+5+5+5$ are 4, 3, 2, and 1 respectively, so the only subgroup containing 5 is also the entire group. Also, any subgroup containing 2 must also contain $2+2=4$, and any subgroup containing 4 must also contain $4+4$, which is 2, so $\{0,2\}$ and $\{0,4\}$ are not subgroups. The only other possibility then is $\{0,2,3,4\}$, but this fails closure since $2+3=5$ is not in the set.

Example 2.22. Subgroups of the dihedral group D_3. Recall that D_3 is the group of symmetries of an equilateral triangle. Here is the Cayley table for reference:

	r_0	r_1	r_2	s_1	s_2	s_3
r_0	r_0	r_1	r_2	s_1	s_2	s_3
r_1	r_1	r_2	r_0	s_2	s_3	s_1
r_2	r_2	r_0	r_1	s_3	s_1	s_2
s_1	s_1	s_3	s_2	r_0	r_2	r_1
s_2	s_2	s_1	s_3	r_1	r_0	r_2
s_3	s_3	s_2	s_1	r_2	r_1	r_0

Notice in the upper left corner the rotations form sort of their own portion of the table where when we compose two rotations, we get another rotation. This shows closure. Also, the identity is r_0. The inverse of r_0 is itself, while r_1 and r_2 are inverses of each other. Thus $\{r_0, r_1, r_2\}$ form a subgroup.

Since each reflection is its own inverse, the following are also subgroups: $\{r_0, s_1\}$, $\{r_0, s_2\}$, and $\{r_0, s_3\}$. Besides these, there is the trivial group $\{r_0\}$ and D_3 itself. There are no other subgroups. It's a little tedious to show this, but try some others to convince yourself. No matter what subset you pick, you find that the closure property fails.

Example 2.23. A non-example The subset $\{-1,0,1\}$ is not a subgroup of $\mathbb{Z}$. Even though it satisfies the identity and inverse properties, it does not satisfy closure. In particular, 1 is in the subset, but $1+1=2$ is not in the subset.

Example 2.24. Even integers under addition. Let's show that the even integers form a subgroup of $\mathbb{Z}$ under addition. Recall that a number is even if it is of the form $2k$ for some integer k. We'll demonstrate both approaches here. For Approach #1, we need to show closure, identity, and inverses. The defining property of this subset is that everything in it is even.

Closure: Pick a and b in the subset. That means they are even, so we can write $a=2j$ and $b=2k$ for some integers j and k. Then $a+b=2j+2k=2(j+k)$. This is in the form of 2 times an integer, so it is even. In summary, we pick two things with the subset's defining property, being even, and show the operation on them also has that defining property.

Identity: The identity in the parent group $\mathbb{Z}$ is 0. It is in the subset because 0 it is even. In particular, 0 can be written as $2(0)$, showing it is a multiple of 2, i.e., even.

Inverses: Let a be in the subset. That is, a is even and we can write $a=2k$ for some integer k. The inverse of a is $-a=-2k=2(-k)$. We see that $-a$ is a multiple of 2, i.e. even, so it is in the subset.

Very similar proofs can be used to show that the multiples of 3 form a subgroup, as do the multiples of 4, 5, or indeed any integer.

Example 2.25. $\mathbb{Q}$ is a subgroup of $\mathbb{R}$ under addition. Recall that the rational numbers $\mathbb{Q}$ are all fractions of the form $\frac{a}{b}$ with $a, b \in \mathbb{Z}$. They are closed under addition since $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$, which is still a fraction with an integer numerator and denominator. The identity, 0, is a rational number (it can be written as $\frac{0}{1}$), and each rational number $\frac{a}{b}$ has an inverse $\frac{-a}{b}$ in $\mathbb{Q}$. Thus $\mathbb{Q}$ forms a subgroup of $\mathbb{R}$ under addition.

Example 2.26. A direct product subset. Let $A \times B$ be a direct product of groups A and B. Recall this is the group of all pairs of the form (a, b) with $a \in A$ and $b \in B$, where the group operation for A is used for the first half of the pair and the group operation for B is used for the second half. Recall also that group operation on (a_1, b_1) with (a_2, b_2) is $(a_1 b_1, a_2 b_2)$.

Let's show that $\{(a, e_B) : a \in A\}$ is a subgroup. These are all the pairs where the second item is the identity of B and the first item can be anything from A. Just to emphasize, the key property defining this subset is that the first half of the pair is something from A and the second half must always be the identity. For example, if $A = \mathbb{Z}_4$ and $B = \mathbb{Z}_6$, then $\{(a, e_B) : a \in A\}$ has the elements $(0, 0)$, $(1, 0)$, $(2, 0)$, and $(3, 0)$.

Closure: We have closure since if (a_1, e_B) and (a_2, e_B) are two items in the subset, then their product $(a_1, e_B)(a_2, e_B) = (a_1 a_2, e_B e_B) = (a_1 a_2, e_B)$ is an element of the subset since $a_1 a_2 \in A$ by closure of A and the second item is still e_B. That is, $(a_1 a_2, e_B)$ satisfies the defining rule of the subset that it's a pair with the first thing from A and the second thing is the identity of B.

Identity: The identity (e_A, e_B) of the parent group is in the subset since its first part, e_A is an element of A and its second part is the identity of B.

Inverses: Given (a, e_B) in the subset, we need to show its inverse is in the subset. Its inverse is (a^{-1}, e_B^{-1}), which we can rewrite as (a^{-1}, e_B). This is in the subset because a^{-1} is in A (since A is a group) and its second part is the identity of B.

Example 2.27. Something general Let G be a group, let g be a particular element of G, and let H be a subgroup of G. Let's prove that $gHg^{-1} = \{ghg^{-1} : h \in H\}$ is a subgroup of G. This is sometimes called the *conjugate subgroup* of H. If the group is abelian, then it's the same as H, but it won't necessarily be H if the group is not abelian. As an example of what gHg^{-1} actually is, suppose $G = D_3$, $g = r_1$, and $H = \{r_0, s_1\}$. We get gHg^{-1} by multiplying all the elements of H by g on the left and g^{-1} on the right. In our example, $g = r_1$ and $g^{-1} = r_2$ and we get $gHg^{-1} = \{r_1 r_0 r_2, r_1 s_1 r_2\} = \{r_0, s_3\}$.

To prove the conjugate subgroup really is a subgroup, remember that the key property that its elements share is that they are of the form ghg^{-1} for some $h \in H$.

Identity: To show the identity is in gHg^{-1}, we need to show e is of the form ghg^{-1} for some $h \in H$. It is, if we take $h = e$ since $e = ge^{-1}$ and $e \in H$ because H is a subgroup itself.

Closure: Take two elements satisfying the property, say gh_1g^{-1} and gh_2g^{-1}, with $h_1, h_2 \in H$. Multiply them to get

$$(gh_1g^{-1})(gh_2g^{-1}) = gh_1(g^{-1}g)h_2g^{-1} = g(h_1h_2)g^{-1}.$$

Since H is a subgroup, $h_1h_2 \in H$, so we have written the product as an element of gHg^{-1}.

Inverse: Take $h \in H$ and look at the inverse of ghg^{-1}. Using Proposition 2.7, we have

$$(ghg^{-1})^{-1} = (g^{-1})^{-1}h^{-1}g^{-1} = gh^{-1}g^{-1}.$$

This is an element of gHg^{-1} since $h^{-1} \in H$.

Example 2.28. Another general example. An important concept in groups is something called the *center* of the group. Given a group G, the center of G is $Z(G) = \{z : za = az \text{ for all } a \in G\}$. Roughly speaking, it is the set of items of G that commute with every item of the group. Recall that if a group is not abelian, then the

operation is not commutative. That is, ab is not always equal to ba. But sometimes it might be, and there might be particular items z for which $za = az$ is always true. These items form the center of the group.

To show that the center is a subgroup, we will show the three subgroup properties hold. The defining property of the center is that every element z in it must satisfy $za = az$ for every $a \in G$. This property must work into all three parts of our proof.

Closure: Suppose $z_1, z_2 \in Z(G)$ and $a \in G$. We need to show that $z_1 z_2 \in Z(G)$. That is, we need to show that $(z_1 z_2)a = a(z_1 z_2)$ for every $a \in G$. We have

$$(z_1 z_2)a = z_1(z_2 a) = z_1(az_2) = (z_1 a)z_2 = (az_1)z_2 = a(z_1 z_2).$$

The first equality above uses associativity, then we use the fact that z_2 is in the center to flip $z_2 a$ into az_2, then we use associativity again, then the fact that z_1 is in the center, and finally associativity.

Identity: The identity e from G must be in the center. This is because for any a in the group, ea and ae both equal a. That is, $ea = ae$ for every a, showing the identity e is in the center.

Inverses: Assume $z \in Z(G)$. We need to show that $z^{-1} \in Z(G)$. Let $a \in G$. Since z is in the center, we know that $az = za$. Multiply both sides of this on the right by z^{-1} to get $a = zaz^{-1}$. Then multiply both sides of this on the left by z^{-1} to get $z^{-1}a = az^{-1}$, which is what we need to show that z^{-1} is in the center.

Below is a useful proposition that gives another way to show something is a subgroup.

Proposition 2.11. *A nonempty subset H of a group G, using the same operation as G, is a subgroup of G if and only if $ab^{-1} \in H$ for every $a, b \in H$.*

Proof. The forward implication is the shorter one to prove. For that, we assume H is a subgroup, and we need to explain why ab^{-1} is in H for every a and b in H. Because H is a subgroup, b^{-1} is in H. And because H is a subgroup, it has the closure property. Thus, ab^{-1} must also be in H.

The backward implication takes a little more work. We will assume that $ab^{-1} \in H$ for every $a, b \in H$. We need to show how this means that H must satisfy the closure, identity, and inverse properties (keeping in mind we already know the operation is associative). First, for the identity property, pick any element $a \in H$. Take $b = a$. Then since $ab^{-1} \in H$ by our assumption, and $ab^{-1} = aa^{-1} = e$, this means $e \in H$. Next for the inverse property, let $c \in H$. We want to show that $c^{-1} \in H$. Take $a = e$ (which we just showed is in H) and $b = c$. Then since $ab^{-1} \in H$ by our assumption and $ab^{-1} = ec^{-1} = c^{-1}$, we have $c^{-1} \in H$. Finally, for closure, let c and d be any elements of H. We want to show that $cd \in H$. Take $a = c$ and $b = d^{-1}$ (which we just showed is in H). Then since $ab^{-1} \in H$ by our assumption and $ab^{-1} = c(d^{-1})^{-1} = cd$, we must have $cd \in H$, as desired. □

The idea of the proposition and proof is ab^{-1} being in H whenever a and b are is a way to encode the closure, identity, and inverse properties of a subgroup in a single property. This approach can sometimes be more efficient for showing something is a subgroup, as opposed to verifying the three properties separately.

Example 2.29. As an example, let's use this criterion to prove that the even integers are a subgroup of $\mathbb{Z}$. To do this, let a and b be even integers. Then $a = 2j$ and $b = 2k$ for some integers j and k. Since the operation is addition, the inverse of b is $-b$, and we want to show that $a + -b$ is an even integer. Plugging in, $a + -b = 2j - 2k = 2(j - k)$, and this is of the form of 2 times an integer, so it is even.

Example 2.30. Let G be an abelian group. Let's show that the set of all elements of order 2 in G along with the identity form a subgroup. For this, we pick elements a and b that are either the identity or have order 2 and show that ab^{-1} is either the identity or has order 2.

Let a and b be in the subset. Let's break this into cases. It's possible that $a = e$ or $b = e$. If $b = e$, then $ab^{-1} = a$, and a was defined to be in the subset, so there's nothing left to show here. If $a = e$, then $ab^{-1} = b^{-1}$. We showed in Example 2.18 that an element and its inverse have the same order. Thus, b^{-1} must either be the identity or have order 2. The only other case to consider is if both a and b have order 2. Let's look at $(ab^{-1})^2$. This equals $ab^{-1}ab^{-1}$, which we can rewrite as $a^2(b^{-1})^2$ since the group is abelian. Since a and b^{-1} have order

2 (using Example 2.18 again for b^{-1}), this product then becomes ee, which is e. So by the definition of order, ab^{-1} must have either order 2 or order 1 (i.e., be the identity).

Being abelian is important here. There are nonabelian groups where this fails.

As a final note in this section, we'll mention that for finite groups, we actually only need to check closure. Once we have closure, we can use that to show we also have the identity and inverses. The proof is a really good exercise, so we will leave it to the reader.

Proposition 2.12. *If G is a finite group and H is a nonempty subset of G that is closed under the group operation of G, then H is a subgroup of G.*

2.5 Background on functions

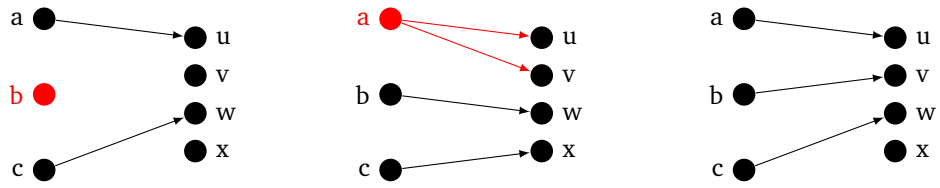
Some background on functions will be helpful for the upcoming material. You're probably used to functions as formulas, like $f(x) = x^2$. But in higher math, functions are thought of as more than just formulas. In general, a function takes inputs in one set and associates them with outputs in another set. If the input set is A and the output set is B, we write a function f as $f : A \to B$. The set of inputs is called the *domain*, and the set of possible outputs is called the *codomain*. Not everything in the codomain will be necessarily an output of some input. The codomain just specifies potential outputs. The *image* of f, denoted by $f(A)$, is the set of all outputs of items in A. Specifically, it is $\{f(a) : a \in A\}$.

For example, $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is a function whose domain and codomain are both $\mathbb{R}$. That is, the inputs and outputs are both real numbers, with the output being the square of the input. The image, $f(\mathbb{R})$, is the interval $[0, \infty)$, all nonnegative real numbers. Note that we could also define the function as $f : \mathbb{R} \to [0, \infty)$. In that case, the codomain and image would be the same thing. In many cases, when creating a function, we won't always have a nice picture of what its image is, so we won't be able to make the codomain equal to the image.

As another example, $f : \mathbb{Z} \to \{0, 1, 2\}$ given by $f(n) = n \bmod 3$ is a function whose domain is all integers, $\mathbb{Z}$, and whose codomain and image are both the set $\{0, 1, 2\}$, which are the possible remainders when dividing by 3.

One more example would be $f : W \to \mathbb{Z}$, where W is the set of all strings of capital letters and $f(w)$ is defined to be the length of w. Here, thinking in terms of inputs and outputs, the inputs are strings, like AAB or ABCDE, and the outputs are integers, like 3 and 5, for those inputs.

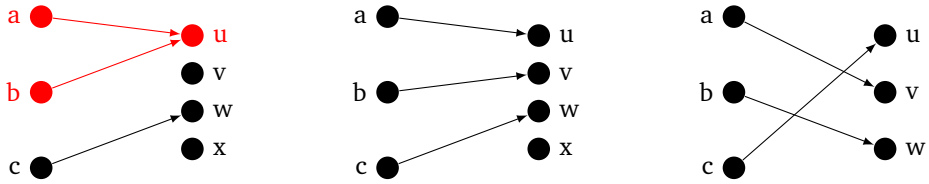
Well-defined functions When defining a function, it is important that every possible input have exactly one output. That is, every input must have an output, and it's not allowed to have more than one output. A function satisfying this is said to be *well-defined*. In the figure below, the domain is $\{a, b, c\}$ and the range is $\{u, v, w, x\}$. The function on the left is not well defined because b doesn't have an output. The function in the middle is not well defined because a has two outputs. The function on the right is well defined.



For another example, let W be the set of all strings of capital letters, and let f be defined such that $f(w)$ is the location of the first A in w. This is not well-defined since if w does not contain an A, then we don't have a value for $f(w)$. As one more example, suppose W' is the set of all strings of capital letters that do contain an A, and we define $f(w)$ as the location of the A in w. This is not well-defined because there could be multiple A's in w and we haven't said which one to pick. To be well defined, we must have only one possible output for each input.

One-to-one functions For some functions, an output can come from multiple inputs. For instance, for a function that returns the length of a string of capital letters, the output 3 corresponds to inputs AAA, AAB, AAC, and many others. A function for which this does not happen, where each output is never hit more than once, is called *one-to-one*. The formal definition of this is that $f : A \to B$ is one-to-one provided whenever $f(a_1) = f(a_2)$, we have $a_1 = a_2$. This formal definition is useful in proving functions are one-to-one.

The example on left is not one-to-one because the output u comes from two inputs, a and b. The functions in the middle and right are both one-to-one.

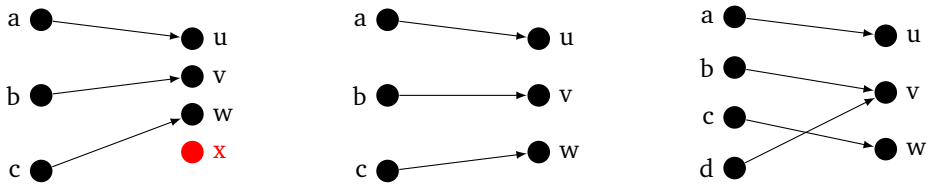


As another example, consider the function $f : W \to W$ which takes strings of capital letters and returns the reverse of that string. It is one-to-one. For instance, looking at a typical output string, such as ABC, there is only one input string we can have that gives that output, namely CBA.

On the other hand, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not one-to-one. The output 4, for instance, is hit by two inputs, 2 and -2.

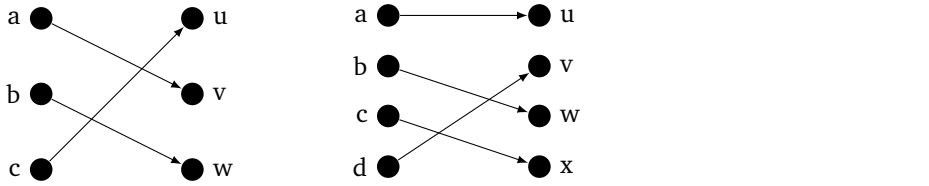
Onto functions A function is called *onto* if the image and codomain are the same. That is, everything in the codomain is actually hit by some input. Formally, $f : A \to B$ is onto if for every $b \in B$, we have $b = f(a)$ for some $a \in A$.

The function below on the left is not onto because the output x is not hit. There is no input that maps to it. The middle and right functions are both onto because every output is hit by some input.



The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not onto because the negative numbers in the codomain are not hit. On the other hand, consider $f : W \to \{0, 1, 2, \dots\}$ defined by $f(w)$ being the number of A's in w. This is onto since every output value is hit by some input. For instance, the inputs B, BA, BAA, BAAA, BAAAA, hit the outputs 0 to 4. For any n, the string of a B followed by n A's will hit the output n.

Bijective functions A function is called a *bijection* if it is both one-to-one and onto. In particular, if $f : A \to B$ is a bijection, then every output in B is associated with exactly one input in A. Both examples below are bijections. Notice that each output is hit, showing the function is onto, and each output is hit by only one input, showing the function is one-to-one.



Notice that both the domain and codomain are the same size in both examples. This is true in general; namely, if the domain and codomain of a bijection are finite sets, then they must be the exact same size.

The function $f : W \to W$ that reverses strings is a bijection. For instance, an output string like CBA comes from

one, and only one, input string, its reverse, ABC in this case. Another example is $f : \mathbb{Z} \to 2\mathbb{Z}$ given by $f(n) = 2n$. Each even integer k is the output of the integer $k/2$, so the function is onto, and there is no other integer besides $k/2$ that satisfies $f(k/2) = k$.

Formally showing things are one-to-one and onto To show something is one-to-one, for small examples, it would be okay to just list out all the possibilities and show nothing repeats. For instance, if $f : \mathbb{Z}_3 \to \mathbb{Z}_5$ is given $f(n) = 2n \bmod 5$, we could just list out all the values of the function, namely $f(0) = 0$, $f(1) = 2$, and $f(2) = 4$. No output is repeated, so the function is one-to-one. We can also list out all the possibilities to show that a function is onto, just making sure that every output shows up at least once.

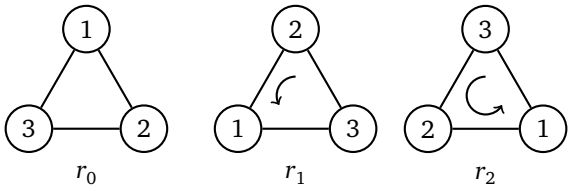
More often, though, we need to use the definition directly. That is, we assume $f(a) = f(b)$ and show that means that $a = b$. For instance, consider $f : \mathbb{R} \to 2\mathbb{R}$ given by $f(x) = 2x$. To show this is one-to-one, we start with $f(a) = f(b)$, which, using the formula, means $2a = 2b$. Simplifying, we get $a = b$, which completes the proof.

To use the definition to show a function onto, we pick an arbitrary element in the codomain and find an element in the domain that maps to it. For example, let's use $f(x) = 2x$ again. We pick an arbitrary element y in the codomain $\mathbb{R}$. We need to find x such that $f(x) = y$. Since $f(x) = 2x$, this means we want $2x = y$. Solve for x to get $x = y/2$. Thus, $f(y/2) = y$, which completes the proof.

then a must equal b. For this function, $f(a) = 2a$ and $f(b) = 2b$. So $f(a) = f(b)$ means $2a = 2b$. Canceling out the twos gives $a = b$, as desired.

2.6 Isomorphisms

Recall that D_3 is the group of symmetries of a triangle. Below are the three rotations r_0, r_1, and r_2, which are rotations by 0°, 120°, and 240°, respectively. These form a subgroup of D_3.



Below on the left is the Cayley table for this subgroup, and on the right is the Cayley table for Z_3.

	r_0	r_1	r_2
r_0	r_0	r_1	r_2
r_1	r_1	r_2	r_0
r_2	r_2	r_0	r_1

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

Notice how similar they are. In D_3, the elements r_0, r_1, and r_2 play a role that is exactly analogous to 0, 1, 2, respectively, in $\mathbb{Z}_3$. For instance, $r_1 r_2 = r_0$ in this subgroup is exactly analogous to $1 + 2 = 0$ in $\mathbb{Z}_3$. This is the idea of an *isomorphism*. An isomorphism is where two groups have exactly the same structure, just possibly with different names for their elements. This notion of isomorphism shows up all over in mathematics. In group theory, it shows up when two groups have elements that can be put into a one-to-one correspondence with each other such that the group operation in each group behaves the exact same way, even if it is a different operation (like function composition versus addition in this example).

Below, we define this concept precisely.

Definition 2.5. *A* homomorphism *between groups $(G, *)$, and $(H, \diamond)$ is a function $\phi : G \to H$ satisfying $\phi(g_1 * g_2) = \phi(g_1) \diamond \phi(g_2)$ for all $g_1, g_2 \in G$. A homomorphism that is also a bijection is called an* isomorphism. *If there exists an isomorphism between groups G and H, we say G and H are isomorphic and write $G \cong H$.*

In the definition, we explicitly show the operations, but usually we will just abbreviate things as $\phi(g_1g_2) = \phi(g_1)\phi(g_2)$. The one exception is if the group operation is addition. For instance, if we are looking at a homomorphism between $\mathbb{Z}$ and itself, the equation would be $\phi(g_1 + g_2) = \phi(g_1) + \phi(g_2)$. Sometimes, the operations are mixed. For instance, a homomorphism from U_{10} to $\mathbb{Z}_4$ would have to satisfy $\phi(g_1g_2) = \phi(g_1) + \phi(g_2)$, since the operation in the domain, U_{10}, is multiplication, and the operation in the codomain, $\mathbb{Z}_4$ is addition.

We will focus for now on isomorphisms, and later we will look at homomorphisms in general. Looking at the definition in more detail, an isomorphism is a function between two groups that identifies each element of a group with one, and only one, element of another group. Suppose the isomorphism identifies g_1, g_2, and g_3 in G with h_1, h_2, and h_3, respectively in H. The homomorphism equation says if $g_1 * g_2 = g_3$ in G, then we must have $h_1 \diamond h_2 = h_3$ in H. Putting this all together, the definition says that isomorphic groups have the same number of elements and the group operations in the two groups behave in the same way.

Note that if we have an isomorphism ϕ from G to H then the inverse function ϕ^{-1} is an isomorphism from H to G. The proof of this is left as an exercise for the reader. So to show two groups are isomorphic, we just need an isomorphism from either one to the other.

The main idea of this section is that if two groups are isomorphic, then they are really the same group, just possibly presented differently. Below are many examples.

Example 2.31. The set $\{0, 2, 4\}$ is a subgroup of $\mathbb{Z}_6$. Working modulo 6, we have $2 + 2 = 4$, $2 + 4 = 0$, and $4 + 4 = 2$. Notice how similar this is to arithmetic in $\mathbb{Z}_3$, where we have $1 + 1 = 2$, $1 + 2 = 0$, and $2 + 2 = 1$, working modulo 3. The elements 0, 2, and 4 in the $\mathbb{Z}_6$ subgroup are analogous to the elements 0, 1, and 2 of $\mathbb{Z}_3$. The addition operation on those elements of $\mathbb{Z}_6$ behaves just the way it does on the elements of $\mathbb{Z}_3$. The Cayley tables below help make this clear.

	0	2	4
0	0	2	4
2	2	4	0
4	4	0	2

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

This is the idea of isomorphism, where two groups behave exactly the same way, just with possibly different names for the elements. The isomorphism here from $\mathbb{Z}_3$ to the subgroup is given by $\phi(0) = 0$, $\phi(1) = 2$, and $\phi(2) = 4$, or more simply by $\phi(n) = 2n$. This identifies 0 to 0, 1 to 2, and 2 to 4. Notice how $1 + 2$ is 0 in $\mathbb{Z}_3$ and $2 + 4$ is 0 in the subgroup. The homomorphism property $\phi(x + y) = \phi(x) + \phi(y)$ tells us since 1 and 2 add to 0 in $\mathbb{Z}_3$, their images $\phi(1)$ and $\phi(2)$ must add to the element corresponding to 0 in the other group.

Example 2.32. Recall that U_{10} is the group $\{1, 3, 7, 9\}$ of elements with multiplicative inverses mod 10. This group is isomorphic to $\mathbb{Z}_4$. Let's look at the Cayley tables of the two groups.

	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

	1	3	9	7
1	1	3	9	7
3	3	9	7	1
9	9	7	1	3
7	7	1	3	9

We have ordered the U_{10} table to have 9 come before 7. Notice how the identities 0 in the left table and 1 in the right table show up in the same four locations in the table. Notice also how the off-diagonal is all 3s in the left table and all 7s in the right table. Further, notice that everywhere there is a 1 in the left table, there is a 3 in the right table, and everywhere there is a 2 in the left table, there is a 9 in the right table.

Putting this together, we can identify $0 \leftrightarrow 1$, $1 \leftrightarrow 3$, $2 \leftrightarrow 9$, and $3 \leftrightarrow 7$. We can write this as an isomorphism $\phi : \mathbb{Z}_4 \to U_{10}$ given by $\phi(0) = 1$, $\phi(1) = 3$, $\phi(2) = 9$, $\phi(3) = 7$. With a little work, one could show that this is the same as the equation $\phi(n) = 3^n \bmod 10$.

This function is definitely a bijection as each output in U_{10} comes from exactly one input in $\mathbb{Z}_4$. To show the homomorphism property, we would have to show that $\phi(m + n) = \phi(m)\phi(n)$ (keeping in mind that the operation is addition in the domain $\mathbb{Z}_4$ and multiplication in the codomain U_{10}. This follows directly from rules

of exponents since

$$\phi(m+n) = 3^{m+n} \bmod 10 = 3^m 3^n \bmod 10 = (3^m \bmod 10)(3^n \bmod 10) = \phi(m)\phi(n).$$

Note that the mod doesn't affect the exponentiation, though we won't prove it.

As an example of what the homomorphism rule is saying here, consider $2+3$ in $\mathbb{Z}_4$. It comes out to 1. Now look at what 2, 3, and 1 map to in U_{10} under ϕ, namely 9, 7, and 3. And in U_{10}, the product of 9 and 7 is 3. The homomorphism property says this sort of thing will always happen.

Example 2.33. The Klein four-group and $\mathbb{Z}_2 \times \mathbb{Z}_2$ are isomorphic. Below are the Cayley tables for each.

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	(0,0)	(0,1)	(1,0)	(1,1)
(0,1)	(0,1)	(0,0)	(1,1)	(1,0)
(1,0)	(1,0)	(1,1)	(0,0)	(0,1)
(1,1)	(1,1)	(1,0)	(0,1)	(0,0)

Notice that the main diagonal consists only of the identity e in the Klein four-group table, and the diagonal consists only of the identity $(0,0)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2$ table. This suggests $e \longleftrightarrow (0,0)$. Next notice that the off-diagonal has all c's for the Klein four-group and all $(1,1)$'s for $\mathbb{Z}_2 \times \mathbb{Z}_2$. This suggests $c \longleftrightarrow (1,1)$. For there, it's not hard to come up with the following isomorphism:

$\phi(e) = (0,0)$
$\phi(a) = (0,1)$
$\phi(b) = (1,0)$
$\phi(c) = (1,1)$.

It's a little tedious to check all the cases, but it is possible to show that this does satisfy the homomorphism property. Just as an example, in the Klein four-group, we have $ab = c$. Replacing a, b, and c, with what they map to gives $(0,1)+(1,0) = (1,1)$, which is true. That is, $\phi(ab) = \phi(a) + \phi(b)$.

Example 2.34. The Klein four-group V is *not* isomorphic to $\mathbb{Z}_4$. Here are their Cayley tables side-by-side.

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

The e's down the main diagonal of the Klein four-group table tell us that every element in that group has order 2. We don't see the identity 0 going down the diagonal in the $\mathbb{Z}_4$ table. In particular, not everything in $\mathbb{Z}_4$ has order 2. As we'll see in a little bit, the strategy to showing two groups are not isomorphic is to find some property that one has that the other doesn't. The property here is that the Klein four-group has four elements of order 2, while $\mathbb{Z}_4$ doesn't.

We could also directly use the homomorphism property to show the groups are not isomorphic. To that end, assume ϕ is an isomorphism mapping $\mathbb{Z}_4$ to V. First, we must have $\phi(0) = e$. This is because $\phi(0) = \phi(0+0) = \phi(0)\phi(0)$. Multiplying both sides by the inverse of $\phi(0)$ gives that $\phi(0) = e$. Now look at $\phi(1+1)$. According to the homomorphism rule, it must equal $\phi(1)\phi(1)$. Whatever $\phi(1)$ maps to, we know that $\phi(1)\phi(1)$ must equal e, since all the elements of V have order 2. However, $\phi(1+1) = \phi(2)$, and $\phi(2)$ can't equal e as we already know $\phi(0) = e$ and ϕ is supposed to be a bijection.

As we'll see later, every group of size 4 is isomorphic to either $\mathbb{Z}_4$ or $\mathbb{Z}_2 \times \mathbb{Z}_2$. That is, there are only two possible structures groups with four elements can follow.

Example 2.35. A 2×2 matrix is a table of values of the form $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Addition is defined by

$$\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} + \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix} = \begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{bmatrix}.$$

For instance,

$$\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} + \begin{bmatrix} 4 & 7 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 2 & 6 \end{bmatrix}.$$

Consider the set of all 2×2 matrices where the upper left entry is an integer and the rest of the entries are 0. Examples include $\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}$. It's straightforward to show that this set forms a group under addition, though we'll leave out the details.

This group is isomorphic to $\mathbb{Z}$ under addition. The isomorphism from $\mathbb{Z}$ to this group is given by

$$\phi(n) = \begin{bmatrix} n & 0 \\ 0 & 0 \end{bmatrix}.$$

It's not hard to show that this is bijective and satisfies the homomorphism property, but we won't do so. The main idea here is that this matrix group is $\mathbb{Z}$ in disguise, basically just $\mathbb{Z}$ with a few extraneous pieces surrounding it. Remember, the idea of groups being isomorphic is that they are essentially the same group, just presented differently.

Example 2.36. As a related example, it's not hard to show that all 2×2 matrices whose bottom two entries are 0 (and whose top two entries are not) is isomorphic to $\mathbb{Z} \times \mathbb{Z}$. The isomorphism here is

$$\phi((n,m)) = \begin{bmatrix} n & m \\ 0 & 0 \end{bmatrix}.$$

Basically, these matrices behave like ordered pairs of integers. The two extraneous 0s at the bottom of the matrix don't anything of interest.

Example 2.37. Let $2\mathbb{Z}$ denote even integers under addition. This group is isomorphic to $\mathbb{Z}$ under addition via the isomorphism $\phi : \mathbb{Z} \to 2\mathbb{Z}$ given by $\phi(n) = 2n$. This function is one-to-one since if $\phi(n) = \phi(m)$, then $2n = 2m$, and this simplifies to $m = n$. The function is onto since if k is an even integer, then $k = 2j$ for some j, which tells us $k = \phi(j)$. It satisfies the homomorphism property since
$\phi(n+m) = 2(m+n) = 2m + 2n = \phi(m) + \phi(n)$.

Example 2.38. Here's one that's maybe a little surprising: the real numbers $\mathbb{R}$ under addition are isomorphic to the positive real numbers $\mathbb{R}^+$ under multiplication. The isomorphism is $\phi : \mathbb{R} \to \mathbb{R}^+$ given by $\phi(x) = e^x$. A look at the graph of the exponential function e^x should be enough to convince you that e^x is one-to-one and onto, though a formal proof would require a little more work, so we will skip it. For the homomorphism property, we have $\phi(x+y) = e^{x+y} = e^x e^y = \phi(x)\phi(y)$.

That isomorphism itself helps us see why the two groups are isomorphic: any two real numbers a and b can be uniquely written in the form $a = e^x$ and $b = e^y$ for some x and y since the exponential function is a bijection. Multiplying a and b corresponds to adding the exponents, namely $ab = e^x e^y = e^{x+y}$. [1]

Example 2.39. Every group is isomorphic to itself. The function $\phi(x) = x$ is a very simple isomorphism showing that this is the case. However, it's not the only isomorphism. For instance, for the group $\mathbb{Z}_4$, the function $\phi(n) = 4 - n$ is an isomorphism. It keeps 0 and 2 fixed and swaps 1 and 3. It is bijective, and it is not hard to check that it satisfies the homomorphism property. Isomorphisms from a group to itself are called *automorphisms*. Describing all the automorphisms of a group is an interesting area of study in group theory. One interesting fact is the set of all automorphisms of a group itself has the structure of a group.

[1] This isomorphism allows us to transform multiplication problems into addition problems. For instance, if we want to multiply 8.2 and 10.3, we first find x and y such that $e^x = 8.2$ and $e^y = 10.3$. These values are $\ln(8.2) \approx 2.104$ and $\ln(10.3) \approx 2.332$. Then add these to get $\ln(8.2) + \ln(10.3) \approx 2.406$. This value is e^{x+y}, that is, $e^{\ln(8.2)+\ln(10.3)}$, Using our approximate values, we have $e^{2.406} \approx 84.44$, which is a little off from the exact answer 84.46. Using more decimal places would get a closer answer. This process is actually how multiplication used to be done before calculators. Instead of e and the natural log, 10 and $\log_{10}$ were used, and people would either look values up in tables or use a mechanical device, called a slide rule, to do the logarithms and powers.

Showing groups are or are not isomorphic

To show two groups are isomorphic, find an isomorphism between them. This is easier said than done. It takes work to examine the structure of the groups and find which things should map to which other things. Coming up with an efficient description of the function can be tricky.

Generators One thing that sometimes helps is to look for a *generator*. A generator of a group is an element g such that every element of the group is some power of g. Not all groups contain a generator like this, but those that do are called *cyclic groups*. For example, $\mathbb{Z}_n$ is a cyclic group with generator 1. This is because 1, $1+1$, $1+1+1$, etc. generate all the items of the group. As another example, U_7 is generated by 3 since modulo 7 we have $3^1 = 3$, $3^2 = 2$, $3^3 = 6$, $3^4 = 4$, $3^5 = 5$, and $3^6 = 1$, and that accounts for all the elements of the group. We will look at generators and cyclic groups in more detail later.

Note that for a finite group, the order of a generator must be the same as the size of the group. No element can have an order larger than the order of the group as there are just not enough elements for the powers to all be distinct. The order can't be smaller than the size of the group because if $a^k = e$, then $a^{k+1} = a$, $a^{k+2} = a^2$, etc., and the powers would end up repeating before they have a chance to cover all the elements.

We can use generators to find isomorphisms in some cases. Here are two examples:

Example 2.40. To find an isomorphism from $\mathbb{Z}_6$ to U_7, note first that both groups do have the same number of elements, so they have a chance of being isomorphic. Use the generator 3 of U_7. In particular, the isomorphism is $\phi(n) = 3^n \bmod 7$. Since 3 is a generator, the powers of 3 cover all the elements of U_7, and 3 has order 6, so the function is onto and one-to-one. The homomorphism property is satisfied because $\phi(m+n) = 3^{m+n} = 3^m 3^n = \phi(m)\phi(n)$.

Example 2.41. To find an isomorphism from $\mathbb{Z}_4 \times \mathbb{Z}_5$ to $\mathbb{Z}_{20}$, note first that both groups do have 20 elements. Next, let's look for a generator of $\mathbb{Z}_4 \times \mathbb{Z}_5$. Since 1 is always a generator of $\mathbb{Z}_n$, it's worth trying $(1,1)$ to see if it works here. Repeatedly adding $(1,1)$ to itself gives $(1,1)$, $(2,2)$, $(3,3)$, $(0,4)$, $(1,0)$, $(2,1)$, $(3,2)$, $(0,3)$, $(1,4)$, $(2,0)$, $(3,1)$, $(0,2)$, $(1,3)$, $(2,4)$, $(3,0)$, $(0,1)$, $(1,2)$, $(2,3)$, $(3,4)$, $(0,0)$. So we see that we do get every element. The isomorphism is then $\phi(n) = (1,1)^n$. Since exponentiation in this group is just repeated addition, the formula reduces to $\phi(n) = (n \bmod 4, n \bmod 5)$.

Showing things are not isomorphic Showing things are not isomorphic is often easier than showing they are since we just have to find any one property that is true in one group that is not true in the other. Below is a list of some things that often work. There are many others besides these. Some of these facts are consequences of things we'll prove in the next section. Proofs of others are left as exercises for the reader.

To show two groups are *not* isomorphic, we can

- show they don't have the same number of elements
- show one is abelian and the other isn't
- show they don't have the same numbers of elements of a given order (like if G has 3 elements of order 2 and H has 4)
- show they don't have the same subgroups (like if one has a subgroup with 10 items and the other doesn't)
- show one is a cyclic group and the other isn't

These are just some of the easiest and most common things to check. There are many others. If you can't find a property that is different in the two groups but still think they are not isomorphic, it might be necessary use the definition more directly, like to show that there cannot exist a bijection between the two sets or to show that the homomorphism property would lead to some contradiction. Here are a few examples using the items from the list above.

Example 2.42. The group $\mathbb{Z}_6$ is not isomorphic to U_6 since $\mathbb{Z}_6$ has 6 elements, while $U_6 = \{1, 5\}$ only has 2.

Example 2.43. The dihedral group D_3 is not isomorphic to $\mathbb{Z}_6$. Though both groups have six elements, $\mathbb{Z}_6$ is abelian, while D_3 is not.

Example 2.44. The group of units U_{20} is not isomorphic to $\mathbb{Z}_8$. Note that U_{20} has elements 1, 3, 7, 9, 11, 13, 17, and 19. It's not hard to work out their orders are 1, 4, 4, 4, 2, 4, 4, and 2, respectively. On the other hand, $\mathbb{Z}_8$ has an element of order 8, namely 1. Since U_{20} doesn't have an element of order 8, the groups are not isomorphic.

Note that this approach is equivalent to saying that U_{20} doesn't have a generator, while $\mathbb{Z}_8$ does (namely 1). In other words, U_{20} is not cyclic, while $\mathbb{Z}_8$ is.

Here is another approach: There are two subgroups of size 4 in U_{20}, namely $\{1, 3, 7, 9\}$ and $\{1, 9, 13, 17\}$. However, there is only one subgroup of size 4 in $\mathbb{Z}_8$, which is $\{0, 2, 4, 6\}$.

Example 2.45. The groups $\mathbb{Z}$ and $\mathbb{Q}$, both under addition, are not isomorphic. Note that 1 is a generator for $\mathbb{Z}$ since every item in $\mathbb{Z}$ can be gotten by repeatedly adding 1 or its inverse -1. However, there is no generator for $\mathbb{Q}$. To see why not, suppose $g = \frac{a}{b}$ is a generator. Then every rational number would be gotten by adding g or $-g$ to itself enough times. However, the only rational numbers we get by doing this will have a denominator that is something that b is divisible by. So, any prime p that is not a factor of b cannot be gotten. Thus, since $\mathbb{Z}$ is cyclic and $\mathbb{Q}$ is not, the two groups are not isomorphic.

2.7 Homomorphisms

An isomorphism between two groups tells us the two groups are really one and the same. A homomorphism between two groups can tell us something about the structure of one group as it relates to the other.

To show something is a homomorphism, we have to verify the equation $\phi(g_1 * g_2) = \phi(g_1) \diamond \phi(g_2)$, where $*$ and $\diamond$ stand for the group operations for the domain and codomain, respectively. To show something it is a not a homomorphism, we can find particular g_1 and g_2 for which the formula fails. Let's look at some examples and non-examples of homomorphisms.

Example 2.46. $\phi : \mathbb{Z} \to \mathbb{Z}_2$ given by $\phi(x) = x \bmod 2$. This is a homomorphism because $\phi(x + y) = (x + y) \bmod 2 = ((x \bmod 2) + (y \bmod 2)) \bmod 2 = \phi(x) + \phi(y)$. One could formally prove that it's okay to move the mods around like this using the division algorithm, which is covered later, but we will leave out those details.

This homomorphism partitions $\mathbb{Z}$ into two pieces, those that map to 0 and those that map to 1. The ones that map to 0 are what we call *even numbers* and those that map to 1 are called *odd numbers*. The arithmetic rules in $\mathbb{Z}_2$ are $0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, and $1 + 1 = 0$. These are the same as the rules for adding evens and odds in $\mathbb{Z}$. This demonstrates that a homomorphism from one group to another allows us to use information about the one group to learn things about the other.

In general, $\phi : \mathbb{Z} \to \mathbb{Z}_n$ given by $\phi(x) = x \bmod n$ tells us that the integers can be partitioned into n sets based on their remainder modulo n, and the arithmetic rules in $\mathbb{Z}_n$ tell us how integers behave in relation to being divisible by n.

Generalizing a little further, $\phi : \mathbb{Z} \to \mathbb{Z}_n$ given by $\phi(x) = kx \bmod n$ turns out to be a homomorphism for any integer k.

Example 2.47. For real numbers, the function $\text{sgn}(x) : \mathbb{R}^* \to \{-1, 1\}$ given by $\text{sgn}(x) = \frac{x}{|x|}$ is a homomorphism. This function is called the sign function of x since it returns the sign of x, namely as either -1 or 1, depending on whether x is negative or positive. It is a homomorphism because $\text{sgn}(xy) = \frac{xy}{|xy|} = \frac{xy}{|x||y|} = \frac{x}{|x|} \cdot \frac{y}{|y|} = \text{sgn}(x)\text{sgn}(y)$.

This homomorphism shows that the rules of arithmetic in $\{-1, 1\}$ under multiplication correspond to the rules

in $\mathbb{R}$ that a positive times a positive is positive, a positive times a negative is a negative, etc.

Example 2.48. Consider $\phi : \mathbb{R} \to \mathbb{Z}$ given by $\phi(x) = \lfloor x \rfloor$. Here, $\lfloor x \rfloor$ denotes the floor function, which returns the greatest integer less than or equal to x. In particular, for positive x, the floor is gotten by dropping the decimal part of x. This is *not* a homomorphism. To show that, we will show that it's not always true that $\phi(x+y) = \phi(x) + \phi(y)$. That is, we want to show that $\lfloor x+y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor$ is not always true.

It's definitely true for some values of x and y. For instance, $\lfloor 2.3 + 2.6 \rfloor = 4$, and $\lfloor 2.3 \rfloor + \lfloor 2.6 \rfloor = 4$ as well. However, $\lfloor 2.3 + 2.8 \rfloor = 5$, while $\lfloor 2.3 \rfloor + \lfloor 2.8 \rfloor = 4$. Since the homomorphism equation fails in at least one case, that is enough to say that ϕ is not a homomorphism.

Example 2.49. If you're familiar with linear algebra, the determinant function is a homomorphism from the group of $n \times n$ invertible matrices under multiplication to $\mathbb{R}^*$. This is because $\det(AB) = \det(A)\det(B)$. This homomorphism gives a link between matrix multiplication and multiplication of real numbers in $\mathbb{R}$.

Example 2.50. Given any groups G and H, consider the function given by $\phi(g) = e_H$ for all g. It is a nice exercise to prove that this is a homomorphism. This homomorphism doesn't tell us much information about either group, but it is worth noting that there is always at least one homomorphism between any pair of groups, namely this trivial one.

Example 2.51. Let's look at homomorphisms from $\mathbb{Z}_m$ to $\mathbb{Z}_n$ given by $\phi(x) = x \bmod n$. Sometimes these will be homomorphisms, and sometimes they won't. The homomorphism equation here is $\phi(x+y) = \phi(x) + \phi(y)$. But we have to be careful with $\phi(x+y)$ since $x+y$ must be done mod m, and then $\phi(x+y)$ must be done mod n. If the moduli are not compatible, then things can go wrong. Let's look at some examples.

1. Look at $\phi : \mathbb{Z}_3 \to \mathbb{Z}_2$ given by $\phi(x) = x \bmod 2$. Here is a table showing what gets mapped to what:

x	0	1	2
$\phi(x)$	0	1	0

   Notice the asymmetry here, where there are more 0s than 1s in the output. This indicates something is wrong. Taking $x = y = 2$ in the homomorphism equation gives
   $\phi(x+y) = \phi((2+2) \bmod 3) = ((2+2) \bmod 3) \bmod 2 = 1$, but $\phi(x) + \phi(y) = (2 \bmod 2) + (2 \bmod 2) = 0$.
   Thus ϕ is not a homomorphism.

2. Let's try $\phi : \mathbb{Z}_4 \to \mathbb{Z}_8$ given by $\phi(x) = x \bmod 8$. The key thing is we're going from a smaller to a bigger group. Here is a table:

x	0	1	2	3
$\phi(x)$	0	1	2	3

   In particular, the codomain is 0 through 7, and 4 through 7 are not hit by anything. This leads to a problem if we take $x = y = 3$ in the homomorphism equation. We get
   $\phi(x+y) = \phi((3+3) \bmod 4) = ((3+3) \bmod 4) \bmod 8 = 2$. But, we also get
   $\phi(x) + \phi(y) = (3 \bmod 8) + (3 \bmod 8) = 6$.

3. Now, let's try $\phi : \mathbb{Z}_{12} \to \mathbb{Z}_4$

x	0	1	2	3	4	5	6	7	8	9	10	11
$\phi(x)$	0	1	2	3	0	1	2	3	0	1	2	3

   Notice how everything repeats neatly here and we get equal numbers of 0s, 1s, 2s, and 3s. This function is a homomorphism. We have $\phi(x+y) = \phi((x+y) \bmod 12) = ((x+y) \bmod 12) \bmod 4$. Things work out well here since modding by 12 and then by 4 doesn't cause any inconsistencies. And we have
   $\phi(x) + \phi(y) = (x \bmod 4) + (y \bmod 4)$.

   One can prove in general that $\phi : \mathbb{Z}_m \to \mathbb{Z}_n$ given by $\phi(x) = x \bmod n$ is a homomorphism if and only if m is divisible by n. One can further prove that $\phi(x) = kx \bmod n$ for a nonzero k is a homomorphism if and only if km is divisible by n.

Facts about homomorphisms

Here we cover a few useful facts about homomorphisms. Since isomorphisms are just special homomorphisms, everything here applies to them as well.

First, recall that the homomorphism equation says $\phi(xy) = \phi(x)\phi(y)$. What if we have three terms, like $\phi(xyz)$? In that case, we can group xyz into $(xy)z$ to turn it into a product of two items and use the homomorphism rule. That is,

$$\phi(xyz) = \phi((xy)z) = \phi(xy)\phi(z) = \phi(x)\phi(y)\phi(z).$$

This idea can be extended using a proof by induction to give the following:

Proposition 2.13. *For any homomorphism $\phi : G \to H$, we have $\phi(g_1 g_2 \cdots g_n) = \phi(g_1)\phi(g_2)\cdots\phi(g_n)$.*

Next, the following properties are used all the time when working with homomorphisms. Note that the first and second parts are actually special cases of the third, but as those special cases are the most common ones, it's worth singling them out.

Proposition 2.14. *Let $\phi : G \to H$ be a homomorphism. Then*

1. $\phi(e_G) = e_H$
2. $\phi(g^{-1}) = \phi(g)^{-1}$
3. $\phi(g^n) = \phi(g)^n$ *for all* $n \in \mathbb{N}$.

Proof. For the first part, we have $\phi(e_G) = \phi(e_G e_G) = \phi(e_G)\phi(e_G)$. Multiplying through by $\phi(e_G)^{-1}$ gives $e_H = \phi(e_G)$.

For the second part, start with $e_H = \phi(e_G) = \phi(g g^{-1}) = \phi(g)\phi(g^{-1})$. Multiplying both sides of this by $\phi(g)^{-1}$ gives $\phi(g)^{-1} = \phi(g^{-1})$, as desired.

For the third part, by the previous proposition, if $n > 0$, we have $\phi(g^n) = \phi(gg\cdots g) = \phi(g)\phi(g)\cdots\phi(g) = \phi(g)^n$. If $n = 0$, then $\phi(g^n) = \phi(e_G) = e_H$ by the first part. If $n < 0$, say $n = -k$ for some $k > 0$, then $\phi(g^n) = \phi((g^{-1})^k) = \phi(g^{-1})^k = (\phi(g)^{-1})^k = \phi(g)^n$. □

The first part of the proposition says that homomorphisms always send the identity to the identity. So a quick way to check that something isn't a homomorphism is if the identity of G gets mapped to something that's not the identity of H. The second part tells us that if g is mapped to h by a homomorphism, then g^{-1} must get mapped to h^{-1}. A consequence of the third part is that if g is a generator of a (cyclic) group, then just knowing what ϕ does to g tells us about $\phi(g')$ for any other g' in the group. That is because $g' = g^k$ for some k, and then $\phi(g') = \phi(g^k) = \phi(g)^k$.

The following propositions are occasionally useful. Their proofs are all good for practicing working with homomorphisms. We will leave the proofs out in the hope that you will use them as good exercises.

Proposition 2.15. *Let $\phi : G \to H$ be a homomorphism. If A is a subgroup of G, then its image $\phi(A) = \{\phi(a) : a \in A\}$ is a subgroup as well.*

Proposition 2.16. *If B is a subgroup of H, then the preimage $\phi^{-1}(B) = \{g \in G : \phi(g) \in B\}$ is a subgroup of G.*

Proposition 2.17. *If $\phi : G \to H$ be a homomorphism, then the* kernel *of ϕ, defined by $K = \{g \in G : \phi(g) = e_H\}$, is a subgroup of G.*

Proposition 2.18. *The homomorphism $\phi : G \to H$ is one-to-one if and only if the only element a satisfying $\phi(a) = e_H$ is e_G.*

Proposition 2.19. *If $\phi : G \to H$ is a homomorphism and $g \in G$ with finite order, then the order of $\phi(g)$ is a divisor of the order of g. If ϕ is also one-to-one, then the orders of g and $\phi(g)$ are equal.*

Chapter 3

Important Groups and Families of Groups

3.1 A few facts from number theory

In this section, we will give a few facts from basic number theory that are useful in proving things about groups. We won't prove them. The proofs are all relatively simple and can be found in any textbook on number theory.

Proposition 3.1. *(Division Algorithm) For any integers a and b, with $b \neq 0$, there exist unique integers q and r such that $a = bq + r$ with $r \in \{0, 1, \ldots |b| - 1\}$.*

The division algorithm is about dividing two numbers, a and b. In particular, when doing a/b, we get a quotient q and remainder r. For instance, when dividing $a = 17$ by $b = 5$, we get a quotient $q = 3$ and a remainder $r = 2$. The division algorithm says we can write this in equation form as $17 = 5(3) + 2$, with the quotient 3 and remainder 2 being unique. The name "division algorithm" comes from its relation to the long division algorithm from elementary school. This proposition is helpful because it gives an equation to use when working with mods (which are remainders).

With $n = 2$, the division algorithm says all integers can be written in the form $2q$ (evens) or $2q + 1$ (odds). With $n = 3$, it says all integers can be written in one of the following three forms: $3q$, $3q + 1$, $3q + 2$. That is, every integer is either a multiple of 3 or it leaves a remainder of 1 when dividing by 3, or it leaves a remainder of 2. With $n = 4$, every integer is of one of these forms: $4q$, $4q + 1$, $4q + 2$, or $4q + 3$. The division algorithm says that similar ideas work for all integers.

Next, here is the formal definition of a familiar concept.

Definition 3.1. *The* greatest common divisor*, or* gcd*, of two integers a and b is the largest positive integer that both are divisible by. It is denoted by* $\gcd(a, b)$*. If* $\gcd(a, b) = 1$*, then a and b are said to be* relatively prime*.*

There is an algorithm, called the *Euclidean algorithm* that can be used to find the gcd. That algorithm is essentially just a repeated application of the division algorithm. One consequence of that algorithm is the following useful proposition that gives us an equation to use when working with the gcd.

Proposition 3.2. *Given any nonzero integers a and b, let $d = \gcd(a, b)$. Then there exist integers x and y such that $ax + by = c$ if and only if c is divisible by d.*

The proof, which we won't cover, involves writing out the steps of the Euclidean algorithm, and then working backwards. For a hint of how all this works, to find the gcd of 186 and 72, the Euclidean algorithm is the following process:

$$\begin{aligned}
186 &= 72(2) + 42 \\
72 &= 42(1) + 30 \\
42 &= 30(1) + 12 \\
30 &= 12(2) + 6 \\
12 &= 6(2) + 0
\end{aligned}$$

We stop when we get a remainder of 0, and the gcd is the remainder 6 in the line above the remainder 0 line. Then we do this sneaky process where we start with the equation from the second-to-last line, solve it for the gcd, and then continually plug in the equations from the lines above, each one solved for the remainder. In our example, we do the following:

$$\begin{aligned}
6 &= 30(1) - 12(2) \\
30(1) - [42 - 30(1)](2) &= 30(3) - 42(2) \\
[72 - 42(1)](3) - 42(2) &= 72(3) - 42(5) \\
72(3) - [186 - 72(2)](5) &= 72(13) - 186(5)
\end{aligned}$$

At the end, we get $72(13) + 186(-5) = 6$, showing $\gcd(72, 186) = 6$ can be written as a linear combination of 72 and 186.

The most important special case of the preceding proposition is when $\gcd(a, b) = 1$. In that case, we can say $\gcd(a, b) = 1$ if and only if there exist integers x and y such that $ax + by = 1$. This is useful because it gives us an equation to work with anytime we have relatively prime integers. Below is one more occasionally useful definition:

Definition 3.2. *The* least common multiple*, or LCM, of integers a and b, denoted* $\text{lcm}(a, b)$*, is the smallest positive integer that is divisible by both a and b.*

For instance, the LCM of 6 and 10 is 30 since the multiples of 6 are 6, 12, 18, 24, 30, ... and the multiples of 10 are 10, 20, 30, ..., with 30 being the smallest positive multiple they have in common. If a and b are relatively prime, then their LCM will be ab, which is the largest value possible. This idea generalizes to the following useful formula:

Proposition 3.3. *Let $a, b \in \mathbb{Z}$, with a and b not both 0. Then* $\text{lcm}(a, b) = \frac{|ab|}{\gcd(a,b)}$.

3.2 The integers modulo n

We have already seen the group $\mathbb{Z}_n$ of integers modulo n under addition. We defined the addition of elements a and b as $(a + b) \bmod n$, where we add a and b and then take the remainder of dividing their sum by n. However, this is not the most common approach to defining $\mathbb{Z}_n$. Let's look briefly at the more common approach. This approach uses the relation defined on the integers below.

Definition 3.3. *We say $a \equiv b \pmod{n}$ if and only if $a - b$ is divisible by n.*

We read this as "a is congruent to b mod (or modulo) n" As an example, $14 \equiv 5 \pmod{3}$ because $14 - 5$ is divisible by 3. It's not too hard to show using the division algorithm that $a \equiv b \pmod{n}$ if and only if a and b leave the same remainder when divided by n. So, for example, 5, 8, 11, 14, 17, etc. are all congruent to 2 modulo 3. Similarly, 7, 11, 15, 19, 24, etc. are all congruent to 3 modulo 4.

We can rewrite the above definition as $a \equiv b \pmod{n}$ if and only if $a - b = nk$ for some integer k. This is very useful, as it gives us a way to turn a statement about modular arithmetic into a simple algebraic equation.

This relation is reflexive, symmetric, and transitive. That is, it is an equivalence relation. It partitions the integers into classes. For instance, when $n = 2$, this partitions the integers into evens and odds. The evens are denoted by $[0]$ and the odds by $[1]$. Specifically, $[0]$ is $\{\dots,-6,-4,-2,0,2,4,6,\dots\}$ and $[1]$ is $\{-5,-3,-1,1,3,5,\dots\}$. The customary names for these sets are $[0]$ and $[1]$, but we could use any item from the sets. For instance, $[0]$ is the same as $[2]$, $[4]$, and $[-12]$.

In general, modulo n, we define $[k] = \{k + in : i \in \mathbb{Z}\}$. Of these sets, there are n distinct sets, each of which is the same as one of $[0]$, $[1]$, ..., $[n-1]$. For example, modulo 3, here are the three sets:

$[0] = \{\dots,-9,-6,-3,0,3,6,9,\dots\}$
$[1] = \{\dots,-11,-8,-5,-2,1,4,7,10,\dots\}$
$[2] = \{\dots,-10,-7,-4,-1,2,5,8,11,\dots\}$

Note that, for example, $[3]$, $[6]$, $[9]$, etc. are all the same set as $[0]$. Writing the brackets becomes tiresome, so people often just write 0, 1, and 2 when they really mean $[0]$, $[1]$, and $[2]$.

We can define addition and multiplication on these sets via the rules $[a]+[b] = [a+b]$ and $[a][b] = [ab]$. For example, with $n = 5$, we have $[3]+[4] = [3+4] = [7]$. But $[7]$ and $[2]$ are the same when $n = 5$, so we can say $[3]+[4] = [2]$. This is all equivalent to saying $(3+4) \bmod 5 = 2$, though we won't go through the trouble of proving it. With these definitions, $\mathbb{Z}_n$ is defined as $\{[0],[1],\dots,[n-1]\}$ under the addition operation defined above. For reasons that will be clear later, people often use the notation $\mathbb{Z}/n\mathbb{Z}$ for this. In a similar way, U_n is defined using the multiplication operation defined above along with just those $[k]$ with k relatively prime to n.

These definitions make some proofs a little simpler. Also, they allow us to be a bit more precise with certain statements. For instance, in $\mathbb{Z}_5$, up to this point we have been writing $4+3=2$. That looks a little weird. Instead, people often will write $4+3 \equiv 2 \pmod 5$.

3.3 Cyclic groups

In Section 2.6, we briefly introduced the idea of generators and cyclic groups. Roughly speaking, a generator of a group "generates" all the elements of that group. Here is a formal definition.

Definition 3.4. *A generator g of a group G is an element such that every element of G is some power of g. That is, if $a \in G$, then $a = g^k$ for some $k \in \mathbb{Z}$. A group that has a generator is called a* cyclic group. *The notation $\langle g \rangle$ denotes the cyclic group $\{g^k : k \in \mathbb{Z}\}$ generated by g. In this expression, g^0 is taken to be the identity by definition, and the negative powers of g are positive powers of g^{-1}.*

The cyclic group $\langle g \rangle$ consists of all the powers of g. Every element is of the form g^k for some k. If the group's operation is addition, then these powers of g are actually multiples of g.

Example 3.1. The prototypical finite examples are the additive groups $\mathbb{Z}_n$. In these groups, $g = 1$ is always a generator. The operation is addition, so in the definition $\{g^k : k \in \mathbb{Z}\}$, the powers are actually repeated addition. In particular, $2 = 1+1$, $3 = 1+1+1$, etc. The identity, 0, is defined by the zero power of 1 (what we get by not adding any 1s at all).

Note that 1 isn't the only generator possible. For instance, in $\mathbb{Z}_6$, 5 is also a generator since 5, $5+5$, $5+5+5$, $5+5+5+5$, $5+5+5+5+5$, and $5+5+5+5+5+5$ give us 5, 4, 3, 2, 1, and 0, respectively. So $\mathbb{Z}_6$ can actually be written as $\langle 1 \rangle$ or $\langle 5 \rangle$. It's a nice exercise to try to show that m is a generator of $\mathbb{Z}_n$ if and only if $\gcd(m,n) = 1$.

The name "cyclic" comes from the fact that the powers of the generator continually cycle through the elements of the group. For instance, in $\mathbb{Z}_6$, the powers of 1 cycle in the pattern 1, 2, 3, 4, 5, 0, 1, 2, 3, 4, 5, 0, 1, ...

Example 3.2. The prototypical infinite example is $\mathbb{Z}$. Here 1 is a generator. Every positive integer can be gotten by adding 1 to itself enough times. Every negative integer can be gotten by adding the inverse of 1, which is -1, to itself enough times (these are the negative powers). And 0 comes from adding 1 to itself no times. There is one other generator of $\mathbb{Z}$, namely -1.

Example 3.3. The group U_{10} is cyclic. Its elements are $\{1,3,7,9\}$, and 3 is a generator since the powers 3^1, 3^2,

3^3, 3^4 are 3, 9, 7, and 1, respectively, which are all the elements of U_{10}.

Note that 7 also works as a generator since the powers of 7 come out to 7, 9, 3, and 1. However, neither 1 nor 9 are generators. They don't have order 4, so they can't generate everything in the group. Unlike with $\mathbb{Z}_n$, there is no simple rule that will tell us specifically what the generators of U_n are. If you're interested in this, look into primitive roots in number theory.

Also, it quite often happens that U_n is not cyclic. For instance, $U_8 = \{1, 3, 5, 7\}$ is not because all of its elements have order 1 or 2, so none are generators. It's not too hard to show that U_p is always cyclic if p is a prime. With more work, it can be shown that U_n is cyclic if and only if $n = 2$, $n = 4$, $n = p^k$, or $n = 2p^k$, where p is any odd prime and $k \geq 1$.

Example 3.4. The group $\mathbb{Z}_2 \times \mathbb{Z}_3$ is cyclic with generator $(1, 1)$. To see this, note that if we repeatedly add $(1, 1)$ to itself, we get $(1, 1)$, $(0, 2)$, $(1, 0)$, $(0, 1)$, $(1, 2)$, and $(0, 0)$, which are all the elements of the group.

However, $\mathbb{Z}_m \times \mathbb{Z}_n$ isn't always cyclic. For instance, $\mathbb{Z}_2 \times \mathbb{Z}_2$ has four elements, all of which have order 1 or 2, so there are no generators. It's a nice exercise to show that $\mathbb{Z}_m \times \mathbb{Z}_n$ is cyclic if and only if m and n are relatively prime. The basic idea is that if m and n are relatively prime, then when we repeatedly add $(1, 1)$ we won't get back to the identity $(0, 0)$ until we've gone through everything. However, if m and n are not relatively prime, then things will line up sooner, meaning $(1, 1)$ won't be able to generate everything, and, in fact, nothing will be able to generate the whole group.

Example 3.5. The rational numbers under addition are not a cyclic group. For $\mathbb{Q}$ to be cyclic, we would need a single rational number, call it p/q, such that every other rational is a repeated sum of p/q. However, repeated sums of p/q are of the form np/q, with n an integer. Anything that has a denominator with no factors in common with q would be impossible to get this way. For instance, if p/q were $1/12$, then would be no way to get $1/5$ or $1/7$.

Example 3.6. The dihedral group D_3 is not cyclic. Each item in that group is either a rotation or a reflection. The powers of a rotation are rotations themselves, and each reflection is its own inverse. So there is no way a single element can generate the whole group.

However, D_3 does have some cyclic subgroups. In particular, the rotations $\{r_0, r_1, r_2\}$ form a cyclic subgroup with generator r_1. Also, pairing any reflection with the identity will always give a cyclic subgroup with two elements, since each reflection has order 2.

Facts about cyclic groups

Every group has cyclic subgroups. Given any element a, the group $\langle a \rangle$ is a cyclic subgroup. For instance, in $\mathbb{Z}$, 2 generates the cyclic subgroup $\langle 2 \rangle$, which is all multiples of 2. Likewise, $\langle 3 \rangle$ is all the multiples of 3, $\langle 4 \rangle$ is all the multiples of 4, etc. Let's state this formally and prove it.

Proposition 3.4. *Given an element g in a group G, $\langle g \rangle$ is a subgroup of G.*

Proof. We'll use Proposition 2.11, where we take $a, b \in \langle g \rangle$ and show $ab^{-1} \in \langle g \rangle$. Given $a, b \in \langle g \rangle$, we have $a = g^i$ and $b = g^j$ for some integers i and j. Therefore, $ab^{-1} = g^i \left(g^j\right)^{-1} = g^{i-j}$, which is an element of $\langle g \rangle$. □

Note that two cyclic subgroups of a group can sometimes be the same. For instance, in $\mathbb{Z}_{12}$, the group $\langle 3 \rangle = \{0, 3, 6, 9\}$, all multiples of 3. This is the same as $\langle 9 \rangle$ since when working mod 12, $\{9, 9+9, 9+9+9, 9+9+9+9\}$ equals $\{9, 6, 3, 0\}$. Next, we have an important proposition.

Proposition 3.5. *Every subgroup of a cyclic group is cyclic.*

Proof. Let $\langle g \rangle$ be the cyclic group and let H be a subgroup of it. First, the trivial subgroup $\{e\}$ is cyclic, since it's generated by e, as all the powers of the identity are the identity. Otherwise, let H be a subgroup with more than one element. Since H contains a non-identity element and everything in G is a power of g, the subgroup H must

contain a nonzero power of g. Let g^m be the smallest nonzero power of g in H. We claim that $H = \langle g^m \rangle$. Since G is a cyclic group, any other element in H is of the form g^k for some integer $k \geq m$. By the division algorithm, we can write $k = mq + r$ for some integers q and r with $0 \leq r < m$. Thus, $g^k = g^{mq+r} = (g^m)^q g^r$. We can solve this for g^r to get $g^r = g^k (g^m)^{-q}$. Note that since g^k and g^m are in the subgroup H, the closure and inverse properties tell us $g^r = g^k (g^m)^{-q}$ must be in H. However, $r < m$ and g^m is the smallest power of g in H, so this g^r can only be in H if $r = 0$. Thus, we have shown that $g^k = g^{mq+r} = g^{mq} = (g^m)^q$, showing g^k is a power of g^m. Since g^k was an arbitrary element of H, that means that every element of H is a power of g^m, meaning $H = \langle g^m \rangle$, a cyclic group. □

To see the intuition behind the proof, look at the subgroup $\{1, 3, 4, 5, 9\}$ of U_{11}. The parent group U_{11} is cyclic with generator 2, and we can write the elements in the subgroup as $\{2^0, 2^8, 2^2, 2^4, 2^6\}$. The smallest positive power is $2^2 = 4$. Notice that every other element of the subgroup is a power of 2^2, showing that it is cyclic. The proof works by picking this smallest positive power and showing everything else is a power of it, making the subgroup cyclic with that smallest power being the generator. Letting g^m be the smallest power, we look at any other element g^k, writing it via the division algorithm as g^{mq+r} and using the closure and inverse properties to show that the remainder r is 0, meaning k is a multiple of m and thus that g^k is a power of g^m.

One immediate consequence of the proceeding proposition is the following:

Proposition 3.6. *In $\mathbb{Z}$, the only subgroups are $\langle 0 \rangle$, $\langle 1 \rangle$, $\langle 2 \rangle$, $\langle 3 \rangle$, etc. Likewise, in $\mathbb{Z}_n$, the only subgroups are $\langle 0 \rangle$, $\langle 1 \rangle$, …, $\langle n-1 \rangle$.*

That is, the only subgroups of $\mathbb{Z}$ and $\mathbb{Z}_n$ are all the multiples of a given integer. Nothing else, like primes, powers of 2, etc. has the right structure to be a subgroup.

Next, we come to an important fact that we've seen already, though we haven't formally proved it. We will need to prove it in order to prove an upcoming theorem. To get a feel for the result, let's look at the element 3 in U_{10}, which has order 4. The first few powers of 3 are $3^0 = 1$, $3^1 = 3$, $3^2 = 9$, and $3^3 = 7$. Then we get back to $3^4 = 1$. And the powers cycle in the repeating pattern 1, 3, 9, 7, 1, 3, 9, 7, 1, etc. forever. Notice that the 9s, for example, show up as 3^2, 3^6, 3^{10}, 3^{14}, etc. Those values are all congruent to 2 modulo 4. No other powers can come out to 9. This sort of thing will always happen in a cyclic group. If an element g has order n, then e, g, g^2, …, g^{n-1} must all be distinct. We then have $g^n = e$ and things follow in the same repeating pattern after that, with g_{n+1}, g_{n+2}, etc. being the same as g^1, g^2, etc. And then farther down the line g_{2n+1}, g_{2n+2}, etc. are the same as g^1, g^2, etc. We can make this precise using modular arithmetic. Below is the formal statement and proof.

Proposition 3.7. *Let g be an element with order n in a group. Then $g^i = g^j$ if and only if $i \equiv j \pmod{n}$.*

Proof. First, if $i \equiv j \pmod{n}$, then $j - i$ is a multiple of n. That is, $j - i = nk$ for some k. Then, using the fact that $g^n = e$, we have

$$e = (g^n)^k = g^{nk} = g^{j-i} = g^j g^{-i}.$$

Multiply both sides of this by g^i to get that $g^i = g^j$.

Next, assume $g^i = g^j$. Multiplying both sides of this by g^{-i} gives $g^{j-i} = e$. Use the division algorithm to write $j - i = nq + r$ for some quotient q and remainder r with $0 \leq r < n$. So, using the fact that $g^n = e$, we get

$$e = g^{j-i} = g^{nq+r} = (g^n)^q g^r = g^r.$$

So we have $g^r = e$ with $r < n$. Since n is the smallest positive power of g giving the identity, we must have $r = 0$. Thus, $j - i = nq$, showing that $i \equiv j \pmod{n}$. □

We come now to a very interesting result: Every cyclic group is isomorphic to either $\mathbb{Z}$ or $\mathbb{Z}_n$ for some n. Before proving it, let's look at a couple of cyclic groups. First, consider the powers of 2 under multiplication, $\{2^n : n \in \mathbb{Z}\}$. This groups consists of 2, 4, 8, 16, etc. along with 1 and the negative powers $\frac{1}{2}$, $\frac{1}{4}$, etc. The isomorphism between this group and $\mathbb{Z}$ is right there in the group's definition, being that there is one power of 2 for each integer. The exponentiation rule $2^{a+b} = 2^a 2^b$ corresponds to the homomorphism property, which turns multiplication in the group into addition of integer exponents.

For a finite group, take U_{10} as an example. It is $\langle 3\rangle$. The powers of 3 cycle as 3, 9, 7, 1, 3, 9, 7, 1, 3, ..., repeating every 4 powers, like the elements of $\mathbb{Z}_4$. Here is a formal statement of the theorem.

Theorem 3.1. *Every infinite cyclic group is isomorphic to $\mathbb{Z}$. Every finite cyclic group is isomorphic to $\mathbb{Z}_n$ for some positive integer n.*

Proof. First, let $\langle g\rangle = \{g^k : k \in \mathbb{Z}\}$ be an infinite cyclic group. Define $\phi : \mathbb{Z} \to \langle g\rangle$ by $\phi(n) = g^n$. This satisfies the homomorphism property because $\phi(n+m) = g^{n+m} = g^n g^m = \phi(n)\phi(m)$. It is also a bijection. To see why, note first that it is onto almost by definition, since every element in $\langle g\rangle$ is g^k for some integer k. To see why it is one-to-one, note that no two powers of g can be equal. This is because if we had $g^i = g^j$, with $i < j$, then multiplying both sides by g^{-i} would give $g^{j-i} = e$. But then there would only be finitely many distinct powers of g by Proposition 3.7. This can't happen since $\langle g\rangle$ is an infinite group.

Now, suppose $\langle g\rangle = \{g^k : k \in \mathbb{Z}\}$ is a finite cyclic group. By Proposition 2.10, g has a finite order n. Consider $\phi : \mathbb{Z}_n \to \langle g\rangle$ given by $\phi([k]) = g^k$. This is an isomorphism. To see why, first we have to show that the function is well defined. In $\mathbb{Z}_3$ for instance, we have $[1] = [4] = [7] = \ldots$, so we have to be careful that the value of the function doesn't change based on the representative we choose for $[k]$. If i and j are elements of $[k]$, then $i \equiv j \equiv k \bmod n$, so by Proposition 3.7, $g^i = g^j = g_k$. So the function is well defined. Next, by Proposition 3.7, the n powers g^0, g^1, g^2, ..., g^{n-1} are all the distinct elements of $\langle g\rangle$, with none of them equal to the others. Thus ϕ is a bijection. For the homomorphism property, note that in the formula, we would like to be able to say $\phi(i+j) = g^{i+j} = g^i g^j = \phi(i)\phi(j)$, like we did in the infinite case. However, we have to be a little careful because there is a mod happening here. However, working with the definition of $\mathbb{Z}_n$ in terms of equivalence classes, we have $\phi([i]+[j]) = \phi([i+j]) = g^{i+j} = g^i g^j = \phi([i])\phi([j])$. □

This tells us that any two finite cyclic groups of the same size n are isomorphic to $\mathbb{Z}_n$ and hence to each other. As we'll see a little later, if n is prime, then, up to isomorphism, $\mathbb{Z}_n$ is the only abelian group with n elements. That is, every abelian group with a prime number of elements n is isomorphic to $\mathbb{Z}_n$.

A little more about generators

Below is a useful fact that is helpful for identifying generators.

Proposition 3.8. *Let G be a finite cyclic group with n elements, and let $g \in G$. Then g is a generator of G if and only if the order of g is n.*

Proof. First, if g has order n, then by the Proposition 3.7, the n elements $g^0, g^1, \ldots, g^{n-1}$ are all distinct. Since G has n elements, each must be one of these powers, showing g is a generator.

Next, assume g is a generator. By Proposition 2.10, g must have finite order. If that order is some integer k, then by the Proposition 3.7, the powers $g^0, g^1, \ldots, g^{k-1}$ are distinct. But since g is a generator, every element in the group is a power of g. Thus $k = n$. □

Thus, when looking for generators of a finite group, we need to only focus on elements whose order matches the size of the group. Any element with a smaller order can't be a generator. Its powers will start repeating before it has a chance to generate everything.

We'll finish this section with a few facts about generators of cyclic groups. We won't prove any of them, though none of their proofs are particularly difficult.

Proposition 3.9. *Here are a few useful facts about generators.*

1. *The only generators of $\mathbb{Z}$ are 1 and -1.*
2. *An element k is a generator of $\mathbb{Z}_n$ if and only if k is relatively prime to n.*
3. *In a finite cyclic group $\langle g\rangle$ with n elements, all generators are of the form g^k, where k is relatively prime to n.*

4. *Let $\langle g \rangle$ be a cyclic group with n elements. For any j, the element g^j generates a cyclic subgroup with $n/\gcd(n,j)$ elements. This element g^j generates the same cyclic subgroup as another element g^i if and only if* $\gcd(j,n) = \gcd(i,n)$.

As an example of #2, the generators of $\mathbb{Z}_{12}$ are 1, 5, 7, and 11, the elements that are relatively prime to 12. As an example of #3, in U_{13}, one of its generators turns out to be 2. Since U_{13} has 12 elements, all of the generators are of the form 2^k with $k = 1$, 5, 7, or 11, the values relatively prime to 12. Modulo 13, these are $2^1 = 2$, $2^5 = 6$, $2^7 = 11$, and $2^{11} = 7$. On a related note, the following Python code is helpful for looking at powers in U_n and finding generators. It shows the powers from 1 to $n-1$ of all the elements of U_n.

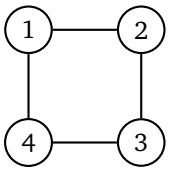
```
from math import gcd
def f(n):
    for j in range(1, n):
       if gcd(j, n) == 1:
           print(j, [pow(j,i,n) for i in range(1,n)])
```

Finally, as an example of #4, in $\mathbb{Z}_{10}$, $\langle 2 \rangle$ has $10/\gcd(10,2) = 5$ elements, and $\langle 4 \rangle$ has $10/\gcd(10,4) = 5$ elements as well. As another example, U_{13} has 12 elements and it has 2 as a generator. In this group, 3 is 2^4 and $\gcd(12,4) = 4$, so $\langle 3 \rangle$ will have $12/4 = 3$ elements.

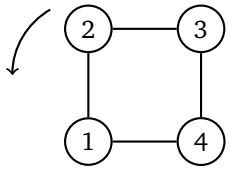
3.4 Dihedral groups

We have already briefly introduced the dihedral group D_3, which is the group of symmetries of an equilateral triangle. In general, the *dihedral group* D_n is the group of all symmetries of a regular n-gon. For instance, D_3 is all the symmetries of an equilateral triangle, D_4 is the all the symmetries of a square, and D_5 is all the symmetries of a regular pentagon.

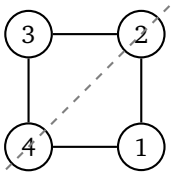
Symmetries of an n-gon are geometric transformations that pick up the n-gon put it back so that it fits exactly in the indentation. Except for putting it back exactly as we found it, all of these symmetries will move vertices around. Sometimes they are called rigid motions. For n-gons, there are two main types of symmetries: rotations and reflections. A rotation and a reflection of the square are shown below.



original



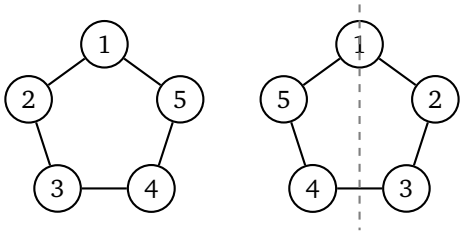
90° counterclockwise rotation



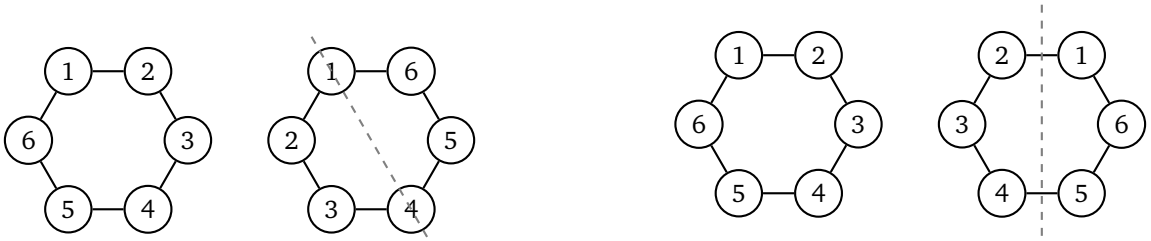
reflection about dotted line

Notice how each of these permutes the vertices but leaves the overall shape the same. In particular, each vertex's neighbors stay the same after a symmetry is applied.

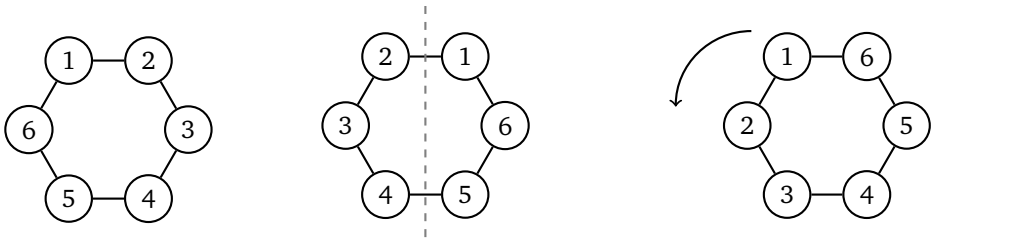
Rotations are always in multiples of $360/n$ degrees (so 120° for D_3, 90° for D_4, etc.). Reflections are always through a line that passes through the n-gon. Things are a little different based on whether n is odd or even. For odd n, the lines always pass through one vertex and the midpoint of the edge directly opposite it. Here is an example reflection with $n = 5$. Notice that vertices 2 and 5 swap places, as do 3 and 4, while 1 stays fixed.



For even n, the lines either pass through two vertices exactly opposite each other or through the midpoints of two edges exactly opposite each other. Here are examples of the two types of reflections for $n = 6$.



The group operation in D_n is function composition. Composition means applying one transformation and then applying the next. For instance, in D_6, if we let r denote rotation by 60° and s denote reflection about a vertical line, then rs is the result of first reflecting and then rotating (function composition happens from right to left). See below.

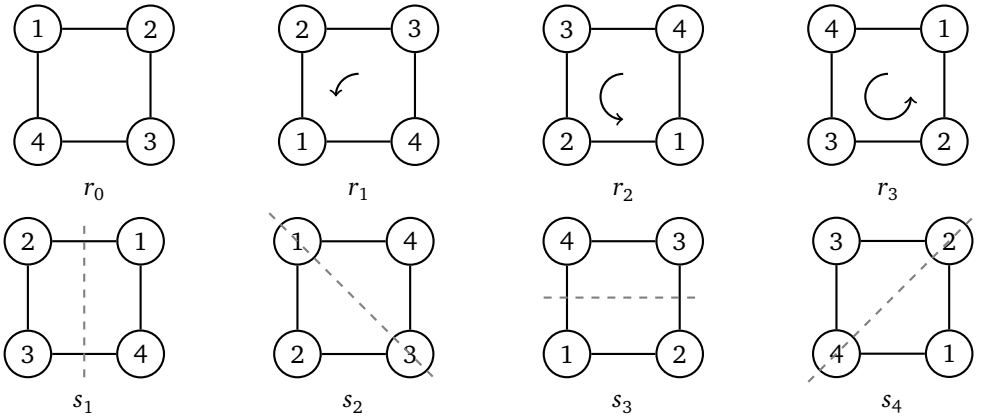


start after reflection (s) after rotation and reflection (rs)

Notice that the composite result is the same as if we had done a single reflection across the line through vertices 1 and 4. In general, whenever we compose two symmetries in D_n, the result will be another symmetry — either a reflection or a rotation. See Section 1.2 for more on D_3.

Structure of D_4

Here are all the elements of D_4.



r_0 r_1 r_2 r_3

s_1 s_2 s_3 s_4

Here is the Cayley table for D_4.

	r_0	r_1	r_2	r_3	s_1	s_2	s_3	s_4
r_0	r_0	r_1	r_2	r_3	s_1	s_2	s_3	s_4
r_1	r_1	r_2	r_3	r_0	s_2	s_3	s_4	s_1
r_2	r_2	r_3	r_0	r_1	s_3	s_4	s_1	s_2
r_3	r_3	r_0	r_1	r_2	s_4	s_1	s_2	s_3
s_1	s_1	s_4	s_3	s_2	r_0	r_3	r_2	r_1
s_2	s_2	s_1	s_4	s_3	r_1	r_0	r_3	r_2
s_3	s_3	s_2	s_1	s_4	r_2	r_1	r_0	r_3
s_4	s_4	s_3	s_2	s_1	r_3	r_2	r_1	r_0

From this we notice a few facts that turn out to be true in general:

- The product of two rotations is a rotation.
- The product of two reflections is a rotation.
- The product of a reflection and a rotation is a reflection.
- All reflections have order 2.
- The group is not abelian.
- Each reflection can be obtained from s_1 by multiplying with an appropriate rotation.

Structure of D_n in general

Let's formally define the elements of D_n. There are a variety of ways people define it, so if you're looking at other sources, be aware that they very likely do things a bit differently than this. We're looking at the symmetries of a regular n-gon with vertices labeled $1, 2, \ldots, n$. The rotations r_k for $k = 0, 1, \ldots n-1$ are rotations by $360k/n$ degrees counterclockwise. The reflections are defined differently based on whether n is odd or even. If n is odd, the reflection s_k for $k = 1, 2, \ldots n$ is the reflection about the line through vertex k and the midpoint of the edge diametrically opposite it on the polygon. If n is even, we break things into even and odd cases. For $k = 0, 1, \ldots, n/2-1$, we define s_{2k+1} to be the reflection across the line through the midpoint of the edge between vertices $2k+1$ and $2k+2$ and the midpoint of the edge on the diametrically opposite side of the polygon. And for $k = 1, 2, \ldots, n/2$, we define s_{2k} to be the reflection across the line through vertex $2k$ and the vertex on the diametrically opposite side of the polygon.

We define s_{2k+1} for $k = 0, 1, \ldots n/2$ to be the reflection about the line through vertex k and the midpoint of the side of the polygon opposite that vertex. We define s_{2k} for $k = 1, 2, \ldots n/2$ to be the reflection about the line between the midpoint of vertices $2k$ and $k+1$

In D_n, the rotation r_1 has order n. The order of the rotation r_k is $n/\gcd(n,k)$ for $k \geq 1$. Each reflection has order 2. The rules for composing symmetries are given by the following rules, with the arithmetic done mod n (where we treat s_0 as another name for s_n):

$$r_i r_j = r_{i+j} \qquad r_i s_j = s_{i+j} \qquad s_i r_j = s_{i-j} \qquad s_i s_j = r_{i-j}.$$

Further, for any rotation r and reflection s, we have $r^{-1} = srs$ and $s = rsr$.

The dihedral groups are not cyclic, but we can think of them as being generated by two elements, r_1 and s_1. Any element in D_n can be written as some combination of those two elements. In fact, we can generate D_n by any rotation of order n along with any reflection. One way people sometimes describe the group is via the notation $\langle r, s : r^n = e, s^2 = e, rs = sr^{-1} \rangle$. We can take $r = r_1$, $s = s_1$, and $e = r_0$ in this. We think of the group as being generated by two elements, r and s, and there are *relations*, $r^n = e$, $s^2 = e$ and $rs = sr^{-1}$, specifying properties of those elements.

Notational note: Notation is unfortunately not standard in math. Some authors use D_n to refer to the dihedral group of order n, meaning they would write D_6 to refer to the group of symmetries of a triangle, D_8 for the

square, etc. There are also various ways to describe the elements of the dihedral groups. Some authors use the term *reflection* to refer only to reflections where the reflecting line passes through a vertex.

Dihedral groups show up in chemistry, where the different symmetries of molecules can explain how different molecules interact with each other. They also show up in crystallography, where the symmetries can help explain the properties of different materials.

3.5 Symmetric groups

Permutations

A *permutation* of a set of items is a reordering of those items. For instance, given the set $\{1,2,3\}$, we can reorder those items in six different ways, namely as $\{1,2,3\}$, $\{1,3,2\}$, $\{2,1,3\}$, $\{2,3,1\}$, $\{3,1,2\}$, and $\{3,2,1\}$. In abstract algebra, a particular notation is used. As an example, the permutation $\{2,3,1\}$ would be written like this:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

The first row indicates the starting set and the second indicates what each item is moved to.

A permutation can be thought of as function whose input is a set and whose output is a rearrangement of that set. The permutation above could be represented by a function σ defined by $\sigma(1)=2$, $\sigma(2)=3$, $\sigma(3)=1$. We could also denote this using arrows as $1\to 2$, $2\to 3$, $3\to 1$.

Composition of permutations

To combine two permutations, we do one followed by the other. Since permutations are really functions, when we combine two permutations, we are actually composing two functions. So $\sigma\circ\tau$ is actually done as $\sigma(\tau(S))$, where S is the original set. This means that we work from right to left. That is, τ is done before σ.

As an example, suppose we want to compose the following permutations.

$$\sigma=\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{pmatrix} \qquad \tau=\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 3 & 5 & 1 \end{pmatrix}.$$

Start with 1 on the right. In τ, we see $1\to 2$. Then we see where 2 goes to in σ. There we have $2\to 1$. So overall, we have $1\xrightarrow{\tau}2\xrightarrow{\sigma}1$, showing that 1 maps to itself in the composition. Doing this for 1, 2, 3, 4, and 5 in order gives the following:

$$1\xrightarrow{\tau}2\xrightarrow{\sigma}1 \qquad 2\xrightarrow{\tau}4\xrightarrow{\sigma}5 \qquad 3\xrightarrow{\tau}3\xrightarrow{\sigma}4 \qquad 4\xrightarrow{\tau}5\xrightarrow{\sigma}3 \qquad 5\xrightarrow{\tau}1\xrightarrow{\sigma}2$$

So we have

$$\sigma\circ\tau=\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 4 & 3 & 2 \end{pmatrix}.$$

Symmetric groups

Permutations along with the composition operation have the structure of a group, namely they satisfy closure, associativity, identity, and inverses.

Closure If we compose two permutations, we end up with another permutation. Permutations just shuffle the order or things, and doing two shuffles in a row still results in shuffling things around. We could make this more formal if we like, but it's not worth the trouble.

Associativity The operation is function composition, which is associative, though we won't prove it.

Identity The permutation that doesn't shuffle anything at all, namely that sends 1 to 1, 2 to 2, 3 to 3, etc., is the identity. In permutation notation, it looks like below.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ 1 & 2 & 3 & \dots & n \end{pmatrix}$$

Inverses An inverse of an operation always undoes the result of that operation. For a permutation like σ below, the operation is $1 \to 2$, $2 \to 4$, $3 \to 1$, and $4 \to 3$. To get its inverse, we can reverse the arrows. For instance, since 1 got sent to 2, to undo that, we have to send 2 back to 1. We can write this as $1 \leftarrow 2$, $2 \leftarrow 4$, $3 \leftarrow 1$, and $4 \leftarrow 3$. Below are σ and its inverse.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} \qquad \sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$$

This process of reversing arrows always works.

Definition 3.5. *The set of all permutations of* $\{1, 2, \dots, n\}$ *with function composition is called the* symmetric group *on n elements. It is denoted by* S_n.

Usually instead of $\sigma \circ \tau$, we just write the operation as $\sigma\tau$. Remember that this is done in backwards order, with τ being done before σ. The composition operation is not commutative, so permutation groups are not abelian, except the very small groups S_0, S_1, and S_2. Note that S_3 turns out to be isomorphic to D_3. Larger symmetric groups are not isomorphic to dihedral groups, but they are isomorphic to generalizations of them. For instance, S_4 is isomorphic to the group of symmetries of a regular tetrahedron, and S_5 is isomorphic to the group of symmetries of a certain four-dimensional analog of a tetrahedron.

There are $n!$ elements in S_n since that is how many ways there are to rearrange the elements 1 to n.

Cycle notation

The notation given above for permutations takes up a lot of space, and there is a shorthand that is often used. It is called *cycle notation*, and it takes a little getting used to. Shown below is the same permutation in both notations:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 6 & 1 & 2 & 3 & 5 \end{pmatrix} \qquad (142653).$$

If we follow the permutation rules starting at 1, we see that 1 goes to 4, then 4 goes to 2, 2 goes to 6, 6 goes to 5, 5 goes to 3, and 3 goes back to 1. So we can write this as a cycle (142653). This notation contains all the information we need to recreate the full permutation.

For instance, given the cycle (324) in S_5, we know that $3 \to 2$, $2 \to 4$, and $4 \to 3$. Everything not shown is assumed to map to itself. Thus (324) is the same as

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 2 & 3 & 5 \end{pmatrix}.$$

Often a permutation can't be written as a single cycle, as in the example below:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 2 & 6 & 5 & 1 \end{pmatrix} \qquad (146)(23)(5).$$

Here, starting with 1, we see that 1 goes to 4, 4 goes to 6, and then 6 goes back to 1. This gives us the cycle (146). Then moving on to 2, we see that 2 goes to 3 and 3 goes back to 2, giving us the cycle (23). Finally 5

goes to itself, giving us (5). We write the entire thing as the product (146)(23)(5). This is actually a product of the following three permutations:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 2 & 3 & 6 & 5 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 3 & 2 & 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}.$$

Notice that the last one is just the identity permutation, so it can be left out. So (146)(23)(5) is often written as (146)(23).

Note: We can write same cycle in multiple ways. For instance, (12345) is same as (23451), (34512) (45123) and (51234). Notice that the starting point changes, while the order of the elements doesn't.

Disjoint cycles

Cycles are called *disjoint* if they share no values in common. That is, $(a_1\, a_2 \ldots a_m)$ and $(b_1\, b_2 \ldots b_n)$ are disjoint cycles provided $a_i \neq b_j$ for all $i = 1, 2, \ldots m$ and $j = 1, 2, \ldots, n$. For instance, (123) and (589) are disjoint cycles because they have nothing in common. But (123) and (539) are not disjoint since they share a 3.

We can always factor a permutation into disjoint cycles. Start with 1, see where it goes, see where that goes, etc. The process given in the preceding subsection will do that.

Note: While order matters for permutations in general, it *does not matter* for disjoint cycles. For instance, (146)(23) and (23)(146) are equal, while $(12)(23) \neq (23)(12)$.

Multiplying cycles

To multiply two cycles, work from right to left and trace out where each value 1, 2, etc. goes. Start off by seeing where 1 goes to in the right cycle and then seeing where that value goes in the left cycle. Repeat the process for 2, 3, etc.

Example 3.7. Multiply (3452)(215) in S_5.

Starting with 1, in the right cycle we see 1 goes to 5. Then move to the left cycle and notice that 5 goes to 2. So in the product, we will have 1 going to 2.

Then we see what happens with 2. Start on the right and see that 2 goes to 1. Then move to the left cycle and see where 1 goes. It doesn't appear in that cycle, which means 1 goes to itself. So in the product 2 goes to 1.

For 3, it goes to itself in the right cycle and 3 goes to 4 in the left, so 3 goes to 4 in the product.

For 4, it also goes to itself in the right cycle and then 4 goes to 5 in the left cycle, so 4 goes to 5 in the product.

For 5, it goes to 2 in the right cycle and then 2 goes to 3 in the left cycle, so 5 goes to 3 in the product.

Here is the finished product:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{pmatrix}.$$

If we like, we can rewrite this as a product of two disjoint cycles: (12)(345).

Note that the procedure for multiplying cycles really is the same thing as the procedure given earlier for composing permutations. The product we just did, (3452)(215), can be written as

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 5 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 3 & 4 & 2 \end{pmatrix}.$$

To see where 1 goes, we see 1 goes to 5 in the right permutation, and then 5 goes to 2 in the left permutation, so 1 goes to 2 in the product. The cycle notation way of doing things just streamlines things since it usually skips over things going to themselves.

Example 3.8. The same principle can be used to multiply more than two cycles. For instance, let's multiply (12)(213)(31) in S_3.

Moving right to left, 1 goes to 3 in the right cycle, then 3 goes to 2 in the middle cycle, and then 2 goes to 1 in the left cycle.

Similarly, for 2, we get 2 going to 2 in the right cycle, 2 going to 1 in the middle, and then 1 going to 2 in the left.

For 3, we have 3 going to 1 in the right, then 1 going to 3 in the middle, and 3 going to 3 in the left.

So the overall product has everything going to itself, making it the identity permutation.

Inverses

A simple way to find the inverse of a permutation written in cycle notation is to write it backwards, both within each cycle and in the order of the cycles. For instance, the inverse of (123)(45)(6789) is (9876)(54)(321).

This works because Proposition 2.7 tells us that $(ab)^{-1} = b^{-1}a^{-1}$ in general, and it's easy to check that multiplying a cycle by its reverse comes out to the identity.

Orders

The order of a single cycle is its length. For instance (13) has order 2 and (341) has order 3.

To find the order of a general permutation, rewrite it as a product of disjoint cycles and then the order is the least common multiple of their lengths.

The reason this works is that since disjoint cycles commute, if we write a permutation as a product $c_1c_2\dots c_k$ of disjoint cycles, then we can rewrite $(c_1c_2\dots c_k)^m$ as $c_1^m c_2^m \dots c_k^m$. This will come out to the identity when all of the individual c_i^m come out to the identity, which happens when m equals the least common multiple of all the lengths.

Note that it's very important that the permutation be written as *disjoint* cycles. If they are not disjoint, the least common multiple approach won't work.

Example 3.9. Let's find the order of the permutation below.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 7 & 2 & 6 & 1 & 4 & 8 & 3 \end{pmatrix}.$$

Start by rewriting it as a product of disjoint cycles, namely as (15)(2783)(46). The least common multiple of the cycle lengths 2, 4, and 2 is 4, so the order of the permutation is 4.

Even and odd permutations and alternating groups

Any given permutation can be written in many different ways as products of cycles. For instance, the identity permutation can be written as (12)(21), (312)(132), and (23)(32)(13)(31). One of the most important ways to write permutations is as products of cycles of length 2. Cycles of length 2 are called *transpositions*. It's possible to break any permutation down into a product of transpositions.

There are many ways to do this. We give two simple ones here. The permutation $(a_1a_2\dots a_n)$ can be rewritten as $(a_1a_2)(a_2a_3)(a_3a_4)\dots(a_{n-1}a_n)$ or as $(a_1a_n)\dots(a_1a_4)(a_1a_3)(a_1a_2)$. For example, the permutation (25431) in S_5 can be written as (25)(54)(43)(31) or as (21)(23)(24)(25).

All the many ways to break a permutation into transpositions share one thing in common: they must all have an even number of transpositions or they must all have an odd number of transpositions. A permutation is called *even* if every breakdown of it into transpositions contains an even number of transpositions. It is called *odd* if every breakdown contains an odd number of transpositions.

Proposition 3.10. *Every permutation is either even or odd. Moreover, exactly half of all permutations of S_n are even and the other half are odd.*

The proof is a little longer than others we've looked at thus far. The first step of the proof is to show the identity permutation is even. We then use that to show that if a permutation can be written as an even number of transpositions, then any other way of writing it as transpositions must also be an even number. We then do a similar thing for odds, and finally we prove the 50/50 split.

Proof. We will first show that the identity is an even permutation. We can't write the identity as a single transposition because that would not fix every element. So assume we have written the identity as a product of two or more transpositions. To show this must be an even amount, we will show that we can keep reducing the number of transpositions in this product by exactly two until we are left with no transpositions at all.

Assume the final transposition is called (ab). The second-to-last transposition can share none, one, or both elements in common with the last transposition. Keeping this in mind, these are the four possible formats for the last two transpositions:

1. $(cd)(ab)$
2. $(bc)(ab)$
3. $(ac)(ab)$
4. $(ab)(ab)$

In these a, b, c, and d are all distinct elements. For the first three cases, we will show how to remove a from the last transposition, For #1, the permutations are disjoint, so we can rearrange them into $(ab)(cd)$. For #2, $(bc)(ab)$ can be rewritten as $(ac)(bc)$, and for #3, it can be rewritten as $(ab)(bc)$. For #4, $(ab)(ab)$ cancels out to just the identity e since transpositions are their own inverses. We can then repeat this process with the second-to-last and third-to-last transpositions, then with the third-to-last and fourth-to-last, etc. At each stage, we either remove a from the rightmost transposition of the pair, or it cancels out like in case #4. So eventually, we will end up cancelling out two of the terms or else we'll end up writing the identity as a product of transpositions with a only appearing in the leftmost transposition, which is impossible for the identity permutation (since that would mean a does not get mapped to a.

All of this can be done more formally using induction, but we will not do that. Now that we know the identity is even, suppose we have an arbitrary permutation σ that we've written as a product of an even number of transpositions $t_1 t_2 \dots t_n$. Suppose we write it as transpositions in a different way as $u_1 u_2 \dots u_m$. Letting e denote the identity permutation, we have

$$e = \sigma^{-1}\sigma = (t_1 t_2 \dots t_n)^{-1}(u_1 u_2 \dots u_m) = \left(t_n^{-1} t_{n-1}^{-1} \dots t_1^{-1}\right)(u_1 u_2 \dots u_m).$$

Since e is an even permutation, that means there must be an even number of transpositions on the right, and since we know that n is even, that means m must be even as well.

A very similar argument can be used if σ is a product of an odd number of transpositions. We end up with the same equation as above, except in this case n is odd, and we must have m odd as well to come out with an even number of transpositions in total.

Finally, to show that half of the transpositions of S_n are even and the rest are odd, consider the function $f : S_n \to S_n$ given by $f(\sigma) = (12)\sigma$. There's nothing special about the choice of (12) here. Any transposition in S_n would do. If σ is an even permutation, then $f(\sigma)$ is an odd permutation, since $f(\sigma)$ contains one more transposition than σ. Likewise, if σ is odd, then $f(\sigma)$ is even.

The function f is one-to-one since if $f(\sigma) = f(\tau)$, then $(12)\sigma = (12)\tau$, and we can cancel (12) to get $\sigma = \tau$. Also, given any $\tau \in S_n$, we have $f((12)\tau) = (12)(12)\tau = \tau$. This shows f is onto. Thus, f is a bijection with every even permutation mapped to an odd and every odd mapped to an even. This can only happen if there are equal amounts of both. □

Definition 3.6. *The subset of all even permutations in S_n forms a subgroup called the* alternating group*, denoted by A_n.*

This definition requires a little work to show that the even permutations really do satisfy the properties of a subgroup. First, we have closure because if σ and τ are even permutations, then $\sigma\tau$ will be as well since the total number of transpositions in $\sigma\tau$ will be the sum of the transpositions from σ and τ (however we decide to write them as transpositions), and the sum of two even integers is even. We showed in the previous proof that the identity is even. Finally, if σ is even, then since $\sigma\sigma^{-1} = e$ and both σ and e have even numbers of transpositions, then σ^{-1} must as well.

Alternating groups are especially important in Galois theory, the key result of which is to show that there is no analog to the quadratic formula for finding roots of polynomials of degrees 5 and higher.

Applications to Puzzles A nice application of the alternating groups is to the 15 Puzzle. In the 15 Puzzle, you have a 4×4 grid of numbers 1 through 15, with one blank space. You can slide numbers around into the blank space, and the goal is to get the numbers in order, like to go from the state shown below on the left to the state in the middle.

6	3	15	8
5	11	4	7
9		2	13
10	1	14	12

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

1	2	3	4
5	6	7	8
9	10	11	12
13	15	14	

When the puzzle was first introduced in the late 1800s, a version looked like the one on the right above, with 14 and 15 swapped from the correct order, and the goal was to put things back in order, like in the middle above. This was actually impossible. The reason has to do with even and odd permutations. The solved state in the middle is the identity, which we know is an even permutation. The state on the right is equivalent to the permutation (1415), which swaps 14 and 15 and leaves everything else the same. This is an odd permutation. Since permutations are either even or odd, there is no way to go from (1415) back to the identity, and thus no way to solve the puzzle on the right.

A similar analysis applies to Rubik's Cubes. If you want to mess with someone, swap any two different-colored stickers and leave everything else alone. The resulting cube cannot be solved. There is a lot of other interesting group theory involving the Rubik's Cube.

Details about S_3

Let's look at S_3. Here are the six permutations in that group:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} = (1)$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = (23)$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = (12)$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = (123) = (13)(12)$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = (132) = (12)(13)$$

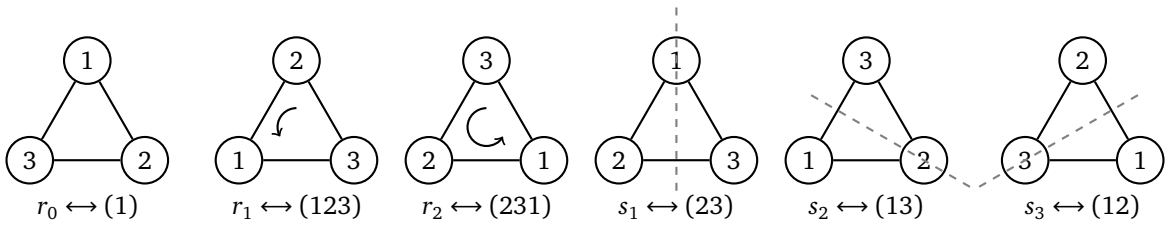
$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = (13).$$

There are three elements of order 2, two elements of order 3, and one element of order 1.

The even permutations are the identity, $(123) = (13)(12)$, and $(132) = (12)(13)$, making A_3 equal to $\{(1), (123), (132)\}$. This is the only subgroup of order 3 in S_3.

There are several subgroups of order 2, namely $\{(1), (23)\}$, $\{(1), (12)\}$, and $\{(1), (13)\}$.

As mentioned earlier, S_3 is isomorphic to D_3. We can see this below. Each symmetry is a permutation of the vertices 1, 2, 3 of the triangle, and there are six symmetries and six elements of D_3.



Cayley's theorem

Cayley's theorem is a remarkable little theorem that says that every group can be thought of as a collection of permutations, that is as a subgroup of a symmetric group. The theorem turns out to not be particularly useful, but it is still nice.

Theorem 3.2. *Every group is isomorphic to a subgroup of a symmetric group.*

We will just prove it for finite groups, since we haven't looked at infinite symmetric groups. The main idea of the proof uses Cayley tables. To help understand the idea of the proof, below is the table for U_5.

	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

Each row can be thought of as a permutation in S_4. The first row is the identity permutation. The second is the permutation $1 \to 2$, $2 \to 4$, $3 \to 1$, $4 \to 3$. The other rows work similarly. Note that there are no repeats within rows. So each row is a permutation. The proof will show this always happens. Next, suppose we take the permutations for rows 2 and 3 and multiply them, as done below:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}.$$

Notice in U_5 that $(2)(3) = 1$, the identity as well. This always happens, namely if we multiply the permutations corresponding to multiplication by a and b, we get the permutation corresponding to multiplication by ab. Let's use these ideas to write the formal proof.

Proof. Each row of a Cayley table corresponds to a permutation because there are no repeated entries in a row. This is because if the two entries in the row for multiplication by some value a were the same, then we would have $ab = ac$, which would mean that $b = c$. A similar argument shows there are no repeats within a column, so each row corresponds to a different permutation.

We can then define $\phi : G \to S_n$ by $phi(a) = \sigma_a$, where σ_a is the permutation of the elements of G gotten by multiplying by a. Note that our definition of S_n was permutations of $\{1, 2, \ldots, n\}$. To be consistent with this, we can think of the elements of G as $a_1, a_2, \ldots, a_n$ and use the subscripts to align with our notation.

The argument of the first paragraph shows that ϕ is one-to-one. It is an onto function if we restrict the codomain to be the image $\phi(G)$. To see why it satisfies the homomorphism property, let σ_a, σ_b, and σ_{ab} be the permutations corresponding to multiplication by a, b, and ab, respectively. Then, looking at what happens to any element x under their composition, we have $(\sigma_a \circ \sigma_b)(x) = \sigma_a(\sigma_b(x)) = \sigma_a(bx) = abx = \sigma_{ab}(x)$. Since x is an arbitrary element, this shows $\sigma_a\sigma_b = \sigma_{ab}$, which shows the homomorphism property. Thus G is isomorphic to its image $\phi(G)$.

Finally, let's look at why this image $\phi(G)$ is a subgroup of S_n. The preceding paragraph shows we have closure. For the identity property, the row of the Cayley table corresponding to multiplication by the group's identity is the identity permutation. For the inverse property, as we showed above, if we take the permutations corresponding to multiplication by a and a^{-1} and compose them, we get the permutation corresponding to multiplication by $aa^{-1} = e$, which we already said is the identity. Thus, each permutation has an inverse. □

Chapter 4

Cosets and quotient groups

4.1 Cosets and Lagrange's theorem

Cosets are an important topic in group theory with many applications in higher math. We'll start with a definition.

Definition 4.1. *Let H be a subgroup of a group G, and let $g \in G$. The set $gH = \{gh : h \in H\}$ is called a* left coset *of H. The set $Hg = \{hg : h \in H\}$ is called a* right coset *of H. If the group operation is addition, we write the left coset as $g + H$ and the right coset as $H + g$.*

To put this into words, given a subgroup H, the left coset gH is what we get by multiplying everything in the group on the left by g. Here are some examples.

Example 4.1. Let $G = U_7$ and $H = \{1, 2, 4\}$. Then the left cosets are

$$1H = \{1, 2, 4\}$$
$$2H = \{2, 4, 1\}$$
$$3H = \{3, 6, 5\}$$
$$4H = \{4, 1, 2\}$$
$$5H = \{5, 3, 6\}$$
$$6H = \{6, 5, 3\}.$$

Notice that some of these come out the same. In particular, $1H = 2H = 4H$, $3H = 5H = 6H$. It usually happens that some of the cosets come out equaling each other. We could also try computing the right cosets, but these will come out exactly the same as the left cosets since the group is abelian. That always happens for abelian groups.

Example 4.2. Let $G = \mathbb{Z}_6$ and $H = \{0, 3\}$. Since the operation in $\mathbb{Z}_6$ is addition, we use the notation $g + H$ for the left cosets. Here they are.

$$0 + H = \{0, 3\}$$
$$1 + H = \{1, 4\}$$
$$2 + H = \{2, 5\}$$
$$3 + H = \{3, 0\}$$
$$4 + H = \{4, 1\}$$
$$5 + H = \{5, 2\}$$

Here we have three distinct cosets, namely $0 + H = 3 + H$, $1 + H = 4 + H$, and $2 + H = 5 + H$.

Example 4.3. Let $G = \mathbb{Z}$ and $H = 4\mathbb{Z}$, the multiples of 4. We have the following:

$$0+H = \{\ldots,-8,-4,0,4,8,\ldots\}$$
$$1+H = \{\ldots,-7,-3,1,5,9,\ldots\}$$
$$2+H = \{\ldots,-6,-2,2,6,10,\ldots\}$$
$$3+H = \{\ldots,-5,-1,3,7,11,\ldots\}$$

With a little work, it would not be too hard to show that these are all the distinct cosets. $4+H$, $8+H$, $12+H$, etc. are all equal to $0+H$. Likewise, $5+H$, $9+H$, $13+H$, etc. are all equal to $1+H$; $6+H$, $10+H$, $14+H$, etc. are all equal to $2+H$; and $7+H$, $11+H$, $15+H$, etc. are all equal to $3+H$.

Example 4.4. Let $G = D_4$ and $H = \{r_0, s_1\}$. Below on the left are the left cosets and on the right are the right cosets. If you want to work these out yourself, refer to the group table in Section 3.4.

$$r_0H = \{r_0, s_1\} \qquad Hr_0 = \{r_0, s_1\}$$
$$r_1H = \{r_1, s_2\} \qquad Hr_1 = \{r_1, s_4\}$$
$$r_2H = \{r_2, s_3\} \qquad Hr_2 = \{r_2, s_3\}$$
$$r_3H = \{r_3, s_4\} \qquad Hr_3 = \{r_3, s_2\}$$
$$s_1H = \{s_1, r_0\} \qquad Hs_1 = \{s_1, r_0\}$$
$$s_2H = \{s_2, r_1\} \qquad Hs_2 = \{s_2, r_3\}$$
$$s_3H = \{s_3, r_2\} \qquad Hs_3 = \{s_3, r_2\}$$
$$s_4H = \{s_4, r_3\} \qquad Hs_4 = \{s_4, r_1\}$$

Here we get four distinct left cosets, with $r_0H = s_1H$, $r_1H = s_2H$, $r_2H = s_3H$, and $r_3H = s_4H$. Notice that we also get four right distinct cosets. However, they are mostly not the same as the left cosets. For instance, not only is $r_1H \neq Hr_1$, but r_1H is not equal to any right coset.

Example 4.5. Let's look at another subgroup of D_4, the rotations $H = \{r_0, r_1, r_2, r_3\}$. The left cosets are below.

$$r_0H = \{r_0, r_1, r_2, r_3\}$$
$$r_1H = \{r_1, r_2, r_3, r_0\}$$
$$r_2H = \{r_2, r_3, r_0, r_1\}$$
$$r_3H = \{r_3, r_0, r_1, r_2\}$$
$$s_1H = \{s_1, s_4, s_3, s_2\}$$
$$s_2H = \{s_2, s_1, s_4, s_3\}$$
$$s_3H = \{s_3, s_2, s_1, s_4\}$$
$$s_4H = \{s_4, s_3, s_2, s_1\}$$

In this case, we have only two distinct cosets, $r_0H = r_1H = r_2H = r_3H$ and $s_1H = s_2H = s_3H = s_4H$. It's a good exercise to work out the right cosets. In this case, even though D_4 is not abelian, the right cosets do turn out to be the same as the left cosets.

In general, if a group is abelian, then the left and right cosets will always turn out equal. If the group is not abelian, then sometimes they turn out equal and sometimes they don't. This turns out to be important later on.

Facts about cosets

Before going any further, here is a quick note about how to prove two sets S and T are equal. The typical approach is to pick $s \in S$ and show s must be in T, and then pick $t \in T$ and show t must be in S. This shows $S \subseteq T$ and $T \subseteq S$, which can only happen if $S = T$.

Let's look at some important facts about cosets. One of the first things you might have noticed from the examples is that the cosets of a given subgroup are all the same size. This is always true.

Proposition 4.1. *Let H be a subgroup of a group G. All the cosets have the same size as H.*

Proof. We will just prove this for left cosets. The proof for right cosets is very similar. Consider the left coset gH, where g is any element of G. First, since gH is just g multiplied by every element of H, it can't be larger than H. It might be smaller, however, if it ever happens that $gh_1 = gh_2$, for some $h_1, h_2 \in H$. But if $gh_1 = gh_2$, then multiplying by g^{-1} on the left gives $h_1 = h_2$. Thus each element of H leads to a different element in gH. □

Note that the proof above really only applies to finite cosets. In the infinite case, the more proper statement is that there is a bijection between each coset and H. The same ideas in the proof can be used to show that $\phi : H \to gH$ given by $\phi(h) = gh$ is a bijection.

Proposition 4.2. *The left cosets of H partition G. That is, every element from G is in exactly one left coset, and any two left cosets must either be equal or disjoint.*

Proof. First, since H is a subgroup, it contains the identity. Thus, for every $g \in G$, $g \in gH$ since we can write it as ge with $e \in H$. So, every element is in at least one coset. Now suppose two cosets g_1H and g_2H have an element x in common. This would mean $x = g_1h_1$ and $x = g_2h_2$ for some $h_1, h_2 \in H$. Setting these equal gives $g_1h_1 = g_2h_2$. Multiplying on the right by h_1^{-1} gives $g_1 = g_2(h_2h_1^{-1})$. But $h_2h_1^{-1} \in H$ since H is a subgroup, so we have shown $g_1 \in g_2H$. Now let $y = g_1h$ be any element in g_1H. Plugging in what we just found for g_1 gives $y = g_1h = g_2(h_2h_1^{-1}h)$, showing $y \in g_2H$. That is, every element y of g_1H is also in g_2H. So $g_1H \subseteq g_2H$. A similar argument shows $g_2H \subseteq g_1H$. Thus if g_1H and g_2H have any element in common, then they must be equal. So, different cosets must be disjoint. □

Proposition 4.3. *Let H be a subgroup of a group G, and let $g_1, g_2 \in G$. The following are equivalent:*

1. $g_1H = g_2H$
2. $g_1 \in g_2H$.
3. $g_2^{-1}g_1 \in H$

To prove n statements are equivalent, a typical approach is to prove $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow \cdots \Rightarrow n \Rightarrow 1$. We do this here.

Proof. First, suppose $g_1H = g_2H$. Since H is a subgroup, it contains e, and we have $g_1 = g_1e \in g_1H$. But $g_1H = g_2H$, so $g_1 \in g_2H$.

Next, suppose $g_1 \in g_2H$. Then $g_1 = g_2h$ for some $h \in H$. Multiply by g_2^{-1} to get $g_2^{-1}g_1 = h$, showing $g_2^{-1}g_1 \in H$.

Finally, suppose $g_2^{-1}g_1 \in H$. Then $g_2^{-1}g_1 = h$ for some $h \in H$. Multiply by g_2 on both sides to get $g_1 = g_2h$, showing $g_1 \in g_2H$. We also have $g_1 \in g_1H$ since we can write $g_1 = g_1e$ and $e \in H$ because H is a subgroup. Proposition 4.2 tells us if two cosets share an element, then they must be the same, so $g_1H = g_2H$. □

As an example of this, look at $G = \mathbb{Z}_{12}$ with $H = \{0, 3, 6, 9\}$. The coset $1 + H$ is $\{1, 4, 7, 10\}$ and this is the same as $4 + H$, $7 + H$, and $10 + H$. These are the only cosets equal to $\{1, 4, 7, 10\}$. The preceding proposition is useful when working with cosets algebraically. Below is another useful fact.

The left and right cosets don't always equal each other, but there are the same total amounts of left and right cosets, as stated below. We'll leave the proof as an exercise.

Proposition 4.4. *Let H be a subgroup of a finite group G. The number of left cosets of H equals the number of right cosets.*

Quick summary Here is a quick summary of facts about the cosets of a subgroup H in a group G.

1. All the cosets have the same size as H.
2. Cosets can't overlap unless they are exactly equal. That is, two cosets are either equal or disjoint.
3. Every element of G is in exactly one coset.
4. We have $x \in aH$ if and only if $xH = aH$. For instance, suppose the coset aH has elements $\{a, b, c, d\}$. Then $aH = bH = cH = dH$, and these are the only cosets that are equal to aH.
5. When working with cosets, if $g_1H = g_2H$, it's often easier to use the fact that this is equivalent to $g_1 g_2^{-1} \in H$.

Lagrange's theorem and consequences

One of the most important results in group theory, with wide-ranging consequences, is Lagrange's theorem. First, we need a definition.

Definition 4.2. *Let H be a subgroup of a group G. The* index *of H in G, denoted $[G : H]$ is the number of left cosets of H in G.*

By Proposition 4.4, the index could also be defined as the number of right cosets. In the theorem below and what follows, we will use the notation $|G|$ to refer to the number of elements of a group G.

The facts that the left cosets partition the group and each has the same size leads to a famous theorem in group theory, called *Lagrange's theorem*, that says the size of a subgroup of a finite group must be a divisor of the size of the group. As an example, look at $G = U_7$ with $H = \{1, 2, 4\}$. The two left cosets, as we saw earlier, are $\{1, 2, 4\}$ and $\{3, 5, 6\}$. So with 2 cosets of size 3, we get the 6 total elements of G. Below is a formal statement and proof.

Theorem 4.1 (Lagrange's theorem)**.** *Let H be a subgroup of a finite group G. Then $|G| = [G : H]|H|$. That is, the size of a subgroup must be a divisor of the size of the group.*

Proof. By Proposition 4.1 all the left cosets have the same size, $|H|$, and by Proposition 4.2, the left cosets partition G. That is, every element is in exactly one coset. Thus, the number of left cosets times the size of each must equal the size of the group. □

For example, the dihedral group D_6 has 12 elements, and Lagrange's theorem tells us the only possible sizes of subgroups would be 1, 2, 4, 6, and 12. The converse of Lagrange's theorem is not necessarily true. If $|G|$ is divisible by k, there is not necessarily a subgroup of size k. One example of this is A_4, which has order 12 but no subgroup of size 6, though that takes some work to show. However, the converse is true if the group is abelian. It is also true for any group, abelian or not, that has a prime number of elements. This is called Cauchy's theorem. We won't prove it. Below are a few useful corollaries of Lagrange's theorem.

Corollary 4.2. *The order of an element of a finite group must be a divisor of the size of the group.*

Proof. Let g be an element of the group G. Look at the subgroup $\langle g \rangle$, whose size is the order of g. That size must be a divisor of the size of G by Lagrange's theorem. □

Corollary 4.3. *Every group with p elements, where p is prime, must be isomorphic to $\mathbb{Z}_p$.*

Proof. By Corollary 4.2, the order of any element of the group must be a divisor of p, namely either 1 or p. Only the identity can have order 1, so every element of the group other than the identity is a generator. Thus the group is cyclic. By Theorem 3.1, it must be isomorphic to $\mathbb{Z}_p$. □

Corollary 4.4. *In any finite group G with element g, we have $g^{|G|} = e$.*

Proof. By Corollary 4.2, if n is the order of g, then n is a divisor of $|G|$. That is $nk = |G|$ for some integer k. So $g^{|G|} = g^{nk} = (g^n)^k = e^k = e$. □

Another really important corollary is Fermat's little theorem, which is very useful in number theory.

Corollary 4.5 (Fermat's little theorem)**.** *Let p be prime. If p is not a divisor of a, then $a^{p-1} \equiv 1 \pmod{p}$.*

An equivalent way to state the theorem is: If p is prime, then $a^p \equiv a \pmod{p}$ for any integer a. Rather than prove it, we will prove a more general fact, called Euler's theorem, of which Fermat's little theorem is a special case.

Corollary 4.6 (Euler's theorem)**.** *If $\gcd(a, n) = 1$, then $a^{\phi(n)} \equiv 1 \pmod{n}$, where $\phi(n)$ denotes the number of positive integers less than n that are relatively prime to n*

Proof. This is a statement about multiplication in U_n. Recall that U_n consists of all the integers relatively prime to n. So $|U_n| = \phi(n)$. By Corollary 4.4, we therefore have $a^{\phi(n)} \equiv 1 \pmod{n}$. □

Fermat's little theorem has a variety of applications. One simple one is for finding the inverse of an element in U_p, where p is prime. Since $a^{p-1} \equiv 1 \pmod{p}$, if we multiply through by a^{-1}, we get $a^{p-2} = a^{-1}$. For instance, the inverse of 2 in U_{17} is 2^{15} mod 17, which is 9.

Fermat's little theorem also has a nice application to testing if a number is prime. The simple approach to testing if a number is prime is to see if it is divisible by 2, 3, 4, 5, etc. Or to speed things up a bit, we can just test divisibility by primes 2, 3, 5, 7, etc. However, for huge prime numbers, hundreds of digits long, like those used in cryptography, this approach would take forever. The contrapositive of Fermat's little theorem says if $a^{p-1} \not\equiv 1 \pmod{p}$, then p is not prime. For instance, to test if 187 is prime, we could check if $2^{186} \equiv 1 \pmod{187}$. It turns out that $2^{186} \equiv 174 \pmod{187}$, so 187 is not prime. An issue with this approach is that 2^{186} is a pretty large number. However, there is a technique called *exponentiation by squaring* to find large powers mod n quickly. The basic idea is that by squaring a number and reducing mod n, you can get to large powers quickly, while keeping the overall result less than n.

A funny thing that happens is we can often use the $2^{p-1} \equiv 1 \pmod{p}$ not just to tell if something is not prime, but also to tell if it is prime. It works up until $p = 341$, which is the first place it gives a wrong answer. There turn out to be around 78,000 primes less than 1,000,000, and this test will give 245 false positives, which is not terrible for such a simple test. However, Fermat's little theorem can be used rather easily to develop another test, called the Miller-Rabin probabilistic probability test that involves testing just a few powers of random integers mod p to see if they come out to ± 1. By picking enough random integers, we can bring the probability of a mistake down to infinitesimally low, and get a really fast test that works to test incredibly large integers.

Euler's theorem is a key part of RSA cryptography, which is widely used for digital signatures and certificates in the internet. To do RSA, a person picks two large prime numbers p and q as well as an encryption key e relatively prime to both $p-1$ and $q-1$. They compute the inverse of e mod $(p-1)(q-1)$ to get the decryption key d. Encryption of a numerical message a is done via a^e mod pq and description of an encrypted value c is done via c^d mod pq. We have $\phi(pq) = (p-1)(q-1)$, and since $ed \equiv 1 \pmod{\phi(n)}$, we can write $ed = 1 + k\phi(n)$. By Euler's theorem $a^{\phi(n)} \equiv 1 \pmod{n}$, so $(a^e)^d \equiv a^{ed} \equiv a^{1+\phi(n)k} = a \cdot (a^{\phi(n)})^k \equiv a \pmod{n}$, showing that the decryption calculation recovers the original value.

4.2 Normal subgroups

We are building towards the definition of quotient groups in the next section. A key concept in their definition is that of *normal subgroups*.

Definition 4.3. *A subgroup N of a group G is said to be a* normal subgroup *of G if $gN = Ng$ for all $g \in G$. That is, for any given g, the left and right cosets are the same.*

Example 4.6. If G is an abelian group, then every subgroup is normal. The left coset $gN = \{gn : n \in N\}$ is the same as the right coset $Ng = \{ng : n \in N\}$ since $gn = ng$ for every $n \in N$ and $g \in G$.

Example 4.7. In Example 4.5 of the last section we saw that the left and right cosets of the group of all rotations in D_4 were equal. So that subgroup of rotations is a normal subgroup. On the other hand, we saw the left and right cosets of $\{r_0, s_1\}$ were not equal, so that subgroup is not normal.

Using the definition is a bit of a pain for showing that a subgroup is normal. Instead, the proposition below is typically used.

Proposition 4.5. *A subgroup N of a group G is normal if and only if $gng^{-1} \in N$ for every $n \in N$ and every $g \in G$.*

Proof. First suppose $gN = Ng$ for all $g \in G$. Let $g \in G$ and $n \in N$. We need to show that $x = gng^{-1}$ is in N. Rewrite this as $xg = gn$. We see that xg has been rewritten in the form of g times something in N, so $xg \in gN$. Since $gN = Ng$, we must have $xg \in Ng$ as well. But this means that $xg = n'g$ for some $n' \in N$. Rewrite this as $x = n'gg^{-1} = n'$. Thus $x \in N$, as desired.

For the other implication, we are assuming $gng^{-1} \in N$ for all $n \in N$ and $g \in G$ and we need to show $gN = Ng$ for all $g \in G$. To do this, we will show if $x \in gN$, then $x \in Ng$. A similar proof, which we will skip since it's nearly the same argument, can show that if $y \in Ng$ then $y \in gN$. Combined, these show $gN = Ng$. So let $x \in gN$. This means $x = gn$ for some $n \in N$. Multiply both sides by g^{-1} on the right to get $xg^{-1} = gng^{-1}$. By our hypothesis, $gng^{-1} \in N$, so we can write $xg^{-1} = n'$ for some $n' \in N$. Multiply both sides on the right by g to get $x = n'g$. This shows that $x \in Ng$, as desired. □

This proposition gives a much easier condition to use to see if a group is normal since it doesn't require computing all the cosets. Let's use it for a few examples.

Example 4.8. Recall that the alternating group A_k is the subgroup of S_k consisting of all the even permutations, all permutations that can be written only as even numbers of transpositions. All permutations are either even or odd. Let's show that A_k is a normal subgroup. To do this, we need it to be true that $gng^{-1} \in A_k$ for all $n \in A_k$ and all $g \in S_k$.

Recall that to find the inverse of a product of cycles, we reverse the order of the cycles and the order within each cycle. Thus if g is an even permutation, so is g^{-1} since its inverse can be written as the same transpositions, just in reverse order. Likewise, if g is odd, then g^{-1} will be as well. So, if g is an even permutation, then g, n, and g^{-1} will all correspond to even numbers of transpositions, and adding a bunch of even numbers gives another even number. Similarly, if g is an odd permutation, then so is g^{-1}, and combining the number of transpositions in g and g^{-1} will give an odd plus an odd, which equals an even number of transpositions. Add to this the even number of transpositions for n, and we see gng^{-1} is written with an even number of transpositions. Thus, whether g is even or odd, we see that gng^{-1} is even, meaning it belongs to A_k.

Example 4.9. Let $\phi : G \to H$ be a homomorphism, and let $K = \{g \in G : \phi(g) = e_H\}$. The set K is called the *kernel* of the homomorphism. It's all the elements that get mapped to the identity. It's not too hard to show that it is a subgroup. Assuming that, let's show that it is a normal subgroup of G.

To that end, let $g \in G$ and $n \in K$. We need to show $gng^{-1} \in K$. That is, given that $\phi(n) = e_H$, we need to show that $\phi(gng^{-1}) = e_H$, too. This follows because

$$\phi(gng^{-1}) = \phi(g)\phi(n)\phi(g^{-1}) = \phi(g)e_H\phi(g^{-1}) = \phi(g)\phi(g^{-1}) = \phi(gg^{-1}) = \phi(e_G) = e_H.$$

Example 4.10. The center of a group G is $Z(G) = \{z : za = az \text{ for all } a \in G\}$. Roughly speaking, it is the set of items of G that commute with every item of the group. It is a normal subgroup. To see this, let $g \in G$ and $n \in Z(G)$. Then, using the fact that n commutes with everything in the group, we have $gng^{-1} = ngg^{-1} = n \in Z(G)$.

For instance, the center of D_4 is $\{r_0, r_2\}$. It is a normal subgroup.

Notes

We'll close this section with a few notes. First, in other sources, you will often see the notation $H < G$ to indicate H is a subgroup of G, and $N \lhd G$ to indicate N is a normal subgroup of G. The element gng^{-1} is often called a *conjugate* of n, and multiplying by g and then g^{-1} is called *conjugation*. The idea is that we do g, then n, and then undo g by doing its inverse. In a non-abelian group, the resulting element is different than n, but it shares a special relationship with it. In particular, if N is normal, then the cosets of N consist of elements that are conjugate to each other. That is, cosets are conjugacy classes.

Of particular interest are what are called *simple groups*. These are groups that have no normal subgroups other than themselves and the trivial group. An important example of a simple group is A_n for $n \geq 5$. This fact, combined with results from the field of Galois theory, a branch of abstract algebra, are what is needed to show that there is no quintic analog of the quadratic formula. That is, there is no formula to find roots of polynomial equations of degree 5 (or higher) that uses only addition, subtraction, multiplication, division, and radicals. That there are quadratic, cubic, and quartic formulas, but no quintic or higher formulas, is related to the fact that A_n is simple and non-abelian for $n \geq 5$, while A_2, A_3, and A_4 are either non-simple or abelian.

A huge mathematical program that was finally finished in the 1980s was to classify all finite simple groups. It was eventually found that every finite simple group is either isomorphic to $\mathbb{Z}_p$ for prime p, A_n for $n \geq 5$, one of 16 families of what are called Lie groups, or one of 26 individual special cases, called sporadic groups. Most famous among these is the so-called Monster group, the last one to be found. It has 808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000 elements.

4.3 Quotient Groups

Quotient groups are of fundamental importance in group theory and are used widely in other subjects, like topology, that rely on group theory. To understand where they come from, let's look at an example first. Consider the group U_7 with subgroup $H = \{1, 2, 4\}$. Below are all the left cosets:

$$\begin{aligned}
1H &= \{1, 2, 4\} \\
2H &= \{2, 4, 1\} \\
3H &= \{3, 6, 5\} \\
4H &= \{4, 1, 2\} \\
5H &= \{5, 3, 6\} \\
6H &= \{6, 5, 3\}.
\end{aligned}$$

We can multiply two cosets by doing their "set product". The set product of sets S and T is $\{st : s \in S, t \in T\}$. That is, we take every element of S and multiply it by every element of T. Let's do this to multiply $(2H)(3H)$, remembering that all the multiplications are done modulo 7. We get

$$(2H)(3H) = \{2 \cdot 3, 2 \cdot 6, 2 \cdot 5, 4 \cdot 3, 4 \cdot 6, 4 \cdot 5, 1 \cdot 3, 1 \cdot 6, 1 \cdot 5\} = \{6, 5, 3, 5, 3, 6, 3, 6, 5\} = \{3, 6, 5\} = 6H.$$

Notice that $2 \cdot 3 = 6$, and $(2H)(3H)$ comes out to $6H$. This is not a coincidence. In fact, it always happens if the group is abelian. This is because if we have $x \in aH$ and $y \in bH$, then $xy = (ah_1)(bh_2)$ for some $h_1, h_2 \in H$, and we can rearrange this into $xy = ab(h_1h_2)$, showing xy is in $(ab)H$. On the other hand, if $z \in (ab)H$, then $z = (ab)h_3$ for some $h_3 \in H$, and we can rewrite this as $z = (ae)(bh_3)$, showing $z \in (aH)(bH)$.

A very important fact is that this can happen sometimes even if the group is not abelian. In particular, it happens whenever H is normal.

Proposition 4.6. *Let N be a normal subgroup of a group G, and let $a, b \in G$. Then $(aN)(bN) = (ab)N$.*

Proof. First, we let $x \in aN$ and $y \in bN$ and show xy must be in $(ab)N$. We must have $x = am$ and $y = bn$ for some $m, n \in N$. Then $xy = (am)(bn)$, which we can rewrite as $a(mb)n$. Since N is normal, the left and right cosets are the same, so we can write $mb = bm'$ for some $m' \in N$. Thus, $xy = (ab)(m'n)$, showing $xy \in (ab)N$.

Next, we let $x \in (ab)N$ and show x must be in $(aN)(bN)$. We have $x = (ab)n$ for some $n \in N$. We can rewrite this as $x = a(bn)$. Since N is normal, the left and right cosets are the same, so we can say $bn = n'b$ for some $n' \in N$. Then $x = an'b$, which we can rewrite as $x = (an')(be)$, showing $x \in (aN)(bN)$. □

It's a good exercise to show that the converse is true, namely that the product only works for normal subgroups and never anything else. The product $(aN)(bN) = (ab)N$ gives us an operation to use to operate on cosets, which turns the set of cosets into a group, whenever we have a normal subgroup.

Definition 4.4. *Let $(G, *)$ be a group, and let N be a normal subgroup of it. Define the operation $*$ on the cosets of N by $aN * bN = (a * b)N$. The set of cosets with this operation is called the* quotient group of G modulo N *and is denoted by G/N.*

People sometimes use the term *factor group* instead of quotient group. Note that the notation G/N is typically read aloud as "G mod N". Usually we won't write the operation. Instead, we'll just write $(aN)(bN) = (ab)N$. The one exception is if the operation is addition. In that case, we'll write the operation as $(a+N)+(b+N) = (a+b)+N$. There is a little work to do to show that this operation really does satisfy the properties of a group.

Proposition 4.7. *Let G be a group with a normal subgroup N. Then G/N is a group.*

Proof. First, we note that the operation really is well defined. That is, it doesn't matter which representatives we use. If we have $aN = a'N$ and $bN = b'N$, then $(a'N)(b'N)$ should also equal $(ab)N$. But this is true because aN and $a'N$ are equal sets, as are bN and $b'N$, so their set product should come out the same, which is $(ab)N$ by Proposition 4.6.

The set of cosets is closed under this operation since $(aN)(bN) = (ab)N$, and $(ab)N$ is a coset since $ab \in G$. The associativity of the operation follows directly from the associativity of the group operation of G. The identity is $eN = N$, where e is the identity in G. The inverse of aN is $a^{-1}N$, where a^{-1} is the inverse of a in G. □

Example 4.11. At the start of this section, we looked at $G = U_7$ and $N = \{1, 2, 4\}$. Note that $N = \langle 2 \rangle$.

We have only two distinct cosets, $1N = 2N = 4N$, and $3N = 5N = 6N$. We can just use the representatives 1 and 3, and we get the Cayley table below for $G/N = U_7/\langle 2 \rangle$:

	$1N$	$3N$
$1N$	$1N$	$3N$
$3N$	$3N$	$1N$

This is the same as the Cayley table for $\mathbb{Z}_2$, so $U_7/\langle 2 \rangle$ is isomorphic to $\mathbb{Z}_2$.

Example 4.12. Let's look at U_{13} with the subgroup $N = \langle 5 \rangle = \{1, 5, 8, 12\}$. One can quickly work out that there are three distinct cosets: $1N = 5N = 8N = 12N = \{1, 5, 8, 12\}$, $2N = 3N = 10N = 11N = \{2, 3, 10, 11\}$, and $4N = 6N = 7N = 9N = \{4, 6, 7, 9\}$. We will use 1, 2, and 4 as the representatives. We can work out the Cayley table as below:

	$1N$	$2N$	$4N$
$1N$	$1N$	$2N$	$4N$
$2N$	$2N$	$4N$	$1N$
$4N$	$4N$	$1N$	$2N$

Note in particular that $(2N)(4N) = 8N$, but $8N$ is the same as $1N$. Also, for $(4N)(4N)$, we have $4 \cdot 4 \equiv 3 \bmod 13$ and $3N = 2N$. This table is the same as the table for $\mathbb{Z}_3$, so $U_{13}/\langle 5 \rangle$ is isomorphic to $\mathbb{Z}_3$.

Example 4.13. Consider $G = \mathbb{Z}$ with $N = 4\mathbb{Z}$, all the multiples of 4. There are four distinct cosets:

$$0 + 4\mathbb{Z} = \{\ldots, -8, -4, 0, 4, 8, \ldots\}$$
$$1 + 4\mathbb{Z} = \{\ldots, -7, -3, 1, 5, 9, \ldots\}$$
$$2 + 4\mathbb{Z} = \{\ldots, -6, -2, 2, 6, 10, \ldots\}$$
$$3 + 4\mathbb{Z} = \{\ldots, -5, -1, 3, 7, 11, \ldots\}$$

In particular, $0+4\mathbb{Z}$ is all the multiples of 4, while $1+4\mathbb{Z}$ is all the integers congruent to 1 mod 4, i.e. that leave a remainder of 1 when divided by 4. Likewise $2+4\mathbb{Z}$ is all the integers congruent to 2 mod 4, and $3+4\mathbb{Z}$ is all the integers congruent to 3 mod 4. To keep the notation compact, write $[0]=0+4\mathbb{Z}$, $[1]=1+4\mathbb{Z}$, etc. The Cayley table is then

	[0]	[1]	[2]	[3]
[0]	[0]	[1]	[2]	[3]
[1]	[1]	[2]	[3]	[0]
[2]	[2]	[3]	[0]	[1]
[3]	[3]	[0]	[1]	[2]

This is exactly the Cayley table of $\mathbb{Z}_4$. This is no coincidence. Many authors take this as the definition of $\mathbb{Z}_4$ and use the notation $\mathbb{Z}/4\mathbb{Z}$ instead. Note that $\mathbb{Z}/n\mathbb{Z}$ works similarly for any positive integer n.

One thing to note about this is how all the multiples of 4 get "identified" to 0. Essentially, the numbers wrap back around to 0, like on a mod 4 clock, and we get an infinitely repeating cycle of 0, 1, 2, 3. This often happens when working with quotient groups. The next example will help illustrate this further.

Example 4.14. Take $G=\mathbb{R}$ and $N=\mathbb{Z}$. Every coset is equal to a coset of the form $a+\mathbb{Z}$ with $0\le a<1$. For example, the coset $0.03+\mathbb{Z}$ is the same as the cosets $1.03+\mathbb{Z}$, $2.03+\mathbb{Z}$, $126.03+\mathbb{Z}$, and $-47.03+\mathbb{Z}$, among many others. Like we saw above with $\mathbb{Z}/4\mathbb{Z}$, this has the effect of identifying all of the integers to 0. But instead of a discrete "clock", we get a continuous circle. That is, $\mathbb{R}/\mathbb{Z}$ essentially is a circle. We have identified the two ends of the interval $[0,1]$ together.

As a similar example, consider the group quotient $(\mathbb{R}\times\mathbb{R})/(\mathbb{Z}\times\mathbb{Z})$. Here, we are looking at the unit square $[0,1]\times[0,1]$ with the left and right sides being identified together, and the top and bottom sides being identified together. What shape does this give us? A torus.

Example 4.15. Recall that the alternating group A_n is the normal subgroup of S_n of all the even permutations. For $n\ge 2$, the quotient group S_n/A_n has two cosets, namely the even permutations A_n, and the odd permutations, which can be represented by $(12)A_n$. The quotient group is isomorphic to $\mathbb{Z}_2$. The fact that S_n has this quotient group tells us a little about S_n, namely that it can be broken up into "odds" and "evens". People study quotient groups of a group in order to understand more about the group.

Example 4.16. Consider the normal subgroup $R=\{r_0,r_1,r_2,r_3\}$ of rotations in D_4. We earlier worked out that there are two distinct cosets, namely R and $s_1R=\{s_1,s_2,s_3,s_4\}$. Let's denote this set by S. The Cayley table of D_4/R is below on the right. Shown on the left is the full Cayley table of D_4. Notice how it sort of collapses onto the table on the right, where the four quadrants of the big table all map to either R or S. The quotient group is giving us the information that a rotation composed with a rotation is a rotation, a rotation composed with a reflection is a reflection, and reflection composed with a reflection is a rotation.

	r_0	r_1	r_2	r_3	s_1	s_2	s_3	s_4
r_0	r_0	r_1	r_2	r_3	s_1	s_2	s_3	s_4
r_1	r_1	r_2	r_3	r_0	s_2	s_3	s_4	s_1
r_2	r_2	r_3	r_0	r_1	s_3	s_4	s_1	s_2
r_3	r_3	r_0	r_1	r_2	s_4	s_1	s_2	s_3
s_1	s_1	s_4	s_3	s_2	r_0	r_3	r_2	r_1
s_2	s_2	s_1	s_4	s_3	r_1	r_0	r_3	r_2
s_3	s_3	s_2	s_1	s_4	r_2	r_1	r_0	r_3
s_4	s_4	s_3	s_2	s_1	r_3	r_2	r_1	r_0

	R	S
R	R	S
S	S	R

Example 4.17. We'll finish with a non-example. The subgroup $H=\{r_0,s_1\}$ of D_4 is not normal. We earlier worked out the cosets and saw that the left and right ones didn't match. What also happens here is that the coset multiplication rule $(aN)(bN)=(ab)N$ breaks down. Consider $r_1H=\{r_1,s_2\}$ and $r_2H=\{r_2,s_3\}$. We have $(r_1r_2)H=r_3H=\{r_3,s_4\}$. But $(r_1H)(r_2H)=\{r_1r_2,r_1s_3,s_2r_2,s_2s_3\}=\{r_0,s_4,r_3\}$, which is not even a coset, much less equal to $(r_1r_2)H$. The problem here is that H is not normal. We are only guaranteed a quotient group when H is normal.

The First Isomorphism theorem

The First Isomorphism theorem, often called the Fundamental Homomorphism theorem, shows that there is a close relationship between homomorphisms from a group and cosets in the group. Let's look at a couple of examples first.

Example 4.18. Consider the homomorphism $\phi : \mathbb{Z}_{16} \to \mathbb{Z}_4$ given by $\phi(n) = n \bmod 4$. Below is a table of values of the function.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\phi(n)$	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3

Of particular interest is the kernel, all the elements of $\mathbb{Z}_{16}$ that get sent to the identity. It is $K = \{0, 4, 8, 12\}$. Related to the kernel are the *fibers* of the other elements, 1, 2, and 3. The fiber of 1 is all the elements mapping to 1, which is $\{1, 5, 9, 13\}$. The fiber of 2 is all the elements mapping to 2, which is $\{2, 6, 10, 14\}$, and the fiber of 3 is $\{3, 7, 11, 15\}$. The thing to notice here is that these are exactly the same as the cosets $1 + K$, $2 + K$, and $3 + K$. This is not a coincidence. It always happens. The fibers always correspond to cosets of the kernel. We will prove this soon, but let's look at another example first.

Example 4.19. Consider $\phi : U_{13} \to U_{13}$ given by $\phi(n) = n^4$. A table of values of the function is below.

n	1	2	3	4	5	6	7	8	9	10	11	12
$\phi(n)$	1	3	3	9	1	9	9	1	9	3	3	1

The kernel (which is the fiber of the identity 0) is $K = \{1, 5, 8, 12\}$. Below are the fibers and the cosets they correspond to:

fiber	values	coset
fiber of 1	$\{1, 5, 8, 12\}$	1K
fiber of 3	$\{2, 3, 10, 11\}$	2K
fiber of 9	$\{4, 6, 7, 9\}$	4K

There is always a one-to-one correspondence between the fibers and cosets. In fact, since the cosets form a group (a quotient group) and the image forms a group (by Proposition 2.15), this one-to-one correspondence is actually a group isomorphism. In this example, the isomorphism is between the image $\{1, 3, 9\}$, and the quotient group $U_{13}/K = \{1K, 2K, 4K\}$. The isomorphism is given by sending 1 to $1K$, 3 to $2K$, and 4 to $4K$. That is, we identify each item $\phi(g)$ of the image with the coset gK.

To see this further, let's look at the coset $2K = \{2, 3, 10, 11\}$. It is 2 times everything in the kernel. When we apply ϕ to each of these things, we are doing $\phi(2k)$ for some $k \in K$. By the homomorphism property, this is $\phi(2)\phi(k)$, which is just $\phi(2) = 3$ since $\phi(k)$ is the identity. So we see everything in $2K$ is part of the fiber of 3. Similarly, if we take an element g from the fiber of 3, we know we have $\phi(g) = 3$. Since $\phi(2) = 3$, we can say $\phi(g) = \phi(2)$, which we can rewrite as $\phi(g)\phi(2)^{-1} = 1$. By homomorphism properties, this is $\phi(2^{-1}g) = 1$, so $2^{-1}g$ is in the kernel K. That is, $2^{-1}g = k$ for some $k \in K$, or $g = 2k$, showing $g = 2K$. Thus, the cosets and the fibers are exactly the same things.

These examples lead us to the following theorem.

Theorem 4.7 (First Isomorphism Theorem)**.** *Let $\phi : G \to H$ be a homomorphism with kernel K. Then G/K is isomorphic to $\{\phi(g) : g \in G\}$.*

Proof. The isomorphism is $\theta : G/K \to \{\phi(g) : g \in G\}$ given by $\theta(aK) = \phi(a)$. We need to show that it really is an isomorphism – that it is one-to-one, onto, and that it satisfies the homomorphism property. We also need to show that it is well-defined. That is, if $aK = bK$, we must make sure that $\theta(aK) = \theta(bK)$. We have $\theta(aK) = \phi(a)$ and $\theta(bK) = \phi(b)$. This could be a problem if $\phi(a) \neq \phi(b)$. However, since $aK = bK$, we have $b = ak$ for some $k \in K$. And we have

$$\phi(b) = \phi(ak) = \phi(a)\phi(k) = \phi(a)e = \phi(a),$$

so things are okay. To show the homomorphism property, we have

$$\theta(xy) = \theta((aK)(bK)) = \theta((ab)K) = \phi(ab) = \phi(a)\phi(b) = \theta(aK)\theta(bK).$$

To show θ is one-to-one, suppose $\theta(aK) = \theta(bK)$. This means $\phi(a) = \phi(b)$. We can rewrite this as $\phi(a)\phi(b)^{-1} = e$. Using the homomorphism property, we can rewrite this as $\phi(ab^{-1}) = e$. Thus, $ab^{-1} \in K$. That is, $ab^{-1} = k$ for some $k \in K$. Thus, $a = kb$. This shows that a is an element of the right coset Kb, which is equal to the left coset bK since K is normal. A similar argument can be used to show $b \in aK$. This can only happen if $aK = bK$, since cosets must be either equal or disjoint by Proposition 4.2. Thus, θ is one-to-one.

Finally, To show θ is onto, note that if $y \in \{\phi(g) : g \in G\}$, then $y = \phi(a)$ for some a, and $\theta(aK) = \phi(a)$. □

Finally, we'll note without proof that each quotient group is associated with a homomorphism, namely $\phi : G \to G/H$ given by $\phi(g) = gH$ is a homomorphism with kernel H. It's good exercise to try to prove it.

4.4 Finitely-generated Abelian Groups

The direct product of two cyclic groups is sometimes cyclic and sometimes not. For instance $\mathbb{Z}_3 \times \mathbb{Z}_5$ is cyclic and isomorphic to $\mathbb{Z}_{15}$. But $\mathbb{Z}_2 \times \mathbb{Z}_2$ is not isomorphic to $\mathbb{Z}_4$. Here is the general fact:

Proposition 4.8. *The group $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \ldots \mathbb{Z}_{n_k}$ is isomorphic to $\mathbb{Z}_{n_1 n_2 \cdots n_k}$ if and only if the n_i are pairwise relatively prime.*

The "pairwise relatively prime" part of the theorem means that none of the n_i can share any factors besides 1 in common with any of the others.

As an example of the proposition, $\mathbb{Z}_2 \times \mathbb{Z}_2$ is not isomorphic to $\mathbb{Z}_4$ because 2 is not relatively prime to 2, but $\mathbb{Z}_3 \times \mathbb{Z}_5$ is isomorphic to $\mathbb{Z}_{15}$ because 3 and 5 are relatively prime. Similarly, $\mathbb{Z}_{120}$ is isomorphic to $\mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_8$. Note that $3 \cdot 5 \cdot 8 = 120$ and none of 3, 5, and 8 share any factors besides 1 in common with the others. But $\mathbb{Z}_{120}$ is not isomorphic to $\mathbb{Z}_3 \times \mathbb{Z}_4 \times \mathbb{Z}_{10}$ because 4 and 10 are not relatively prime.

We won't prove this proposition, but the proof is not too difficult. The basic idea is to show that the order of an element $(g_1, g_2, \ldots, g_k)$ of $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_k}$ is the least common multiple of the orders of g_1, g_2, etc. in $\mathbb{Z}_{n_1}$, $\mathbb{Z}_{n_2}$, etc. If the n_i are pairwise relatively prime, then $(1, 1, \ldots, 1)$ will work as a generator, as least common multiple of the individual orders will come out to the product $n_1 n_2 \cdots n_k$, which is the size of the group, giving us a generator. If they are not pairwise relatively prime, then every element has order less than the order of the group, meaning the group is not cyclic.

Fundamental Theorem of Finite Abelian Groups

There are two fundamental theorems that limit the possibilities for what abelian groups can look like. Here is the first theorem, for groups that have a finite number of elements. We will not prove it here, as the proof is more involved than what we want to get into.

Theorem 4.8 (Fundamental Theorem of Finite Abelian Groups)**.** *Every finite abelian group is isomorphic to a group of the following form, where p_1, p_2, etc. are primes, but with possible repeats:*

$$\mathbb{Z}_{p_1^{\alpha_1}} \times \mathbb{Z}_{p_2^{\alpha_2}} \times \ldots \mathbb{Z}_{p_k^{\alpha_k}}.$$

Example 4.20. The finite abelian groups of order 24 are all isomorphic to one of the following: $\mathbb{Z}_8 \times \mathbb{Z}_3$, $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3$, and $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$.

To get this, factor 24 into prime factors $2^3 \cdot 3$. Then use this to find all the ways to break 24 into products of factors, where each factor is a prime to a power. These are $2^3 \cdot 3$, $2 \cdot 2^2 \cdot 3$, and $2 \cdot 2 \cdot 2 \cdot 3$. We do not include something like $12 \cdot 2$ because 12 is not a prime to a power. Note that order doesn't matter. For instance, $\mathbb{Z}_8 \times \mathbb{Z}_3$ and $\mathbb{Z}_3 \times \mathbb{Z}_8$ are isomorphic. Note also, that using Proposition 4.8, people often write these products more

simply. For instance, $\mathbb{Z}_8 \times \mathbb{Z}_3$ is isomorphic to $\mathbb{Z}_{24}$, $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3$ is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_{12}$, and $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$ is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_6$.

Using the theorem and Proposition 4.8, here is a table of all the abelian groups of sizes 1 through 20. That is, every abelian group with 1 through 20 elements is isomorphic to one of the groups in the table.

n	Groups
1	$\mathbb{Z}_1$
2	$\mathbb{Z}_2$
3	$\mathbb{Z}_3$
4	$\mathbb{Z}_4,\ \mathbb{Z}_2 \times \mathbb{Z}_2$
5	$\mathbb{Z}_5$
6	$\mathbb{Z}_6$
7	$\mathbb{Z}_7$
8	$\mathbb{Z}_8,\ \mathbb{Z}_2 \times \mathbb{Z}_4,\ \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$
9	$\mathbb{Z}_9,\ \mathbb{Z}_3 \times \mathbb{Z}_3$
10	$\mathbb{Z}_{10}$
11	$\mathbb{Z}_{11}$
12	$\mathbb{Z}_{12},\ \mathbb{Z}_2 \times \mathbb{Z}_6$
13	$\mathbb{Z}_{13}$
14	$\mathbb{Z}_{14}$
15	$\mathbb{Z}_{15}$
16	$\mathbb{Z}_{16},\ \mathbb{Z}_4 \times \mathbb{Z}_4,\ \mathbb{Z}_2 \times \mathbb{Z}_8,\ \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4,\ \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$
17	$\mathbb{Z}_{17}$
18	$\mathbb{Z}_{18},\ \mathbb{Z}_3 \times \mathbb{Z}_6$
19	$\mathbb{Z}_{19}$
20	$\mathbb{Z}_{20},\ \mathbb{Z}_2 \times \mathbb{Z}_{10}$

Finitely generated groups

We have seen cyclic groups, in which every item is a power of some element g, called a generator. This idea can be extended to the idea of a finitely-generated group.

Definition 4.5. *A group is said to be* finitely-generated *if there is a collection of elements (called generators) of the group, $\{g_1, g_2, \ldots, g_n\}$, such that every element in the group is a product of powers of these elements. Specifically, any element of the group can be written as $g_{i_1}^{k_1} g_{i_2}^{k_2} \cdots g_{i_n}^{k_n}$, where each g_{i_j} is one of the generators and each k_{i_j} is an integer power.*

Put simply, every element of the group is gotten by multiplying a finite number of powers of the generators. Sometimes, particularly for non-abelian groups, the generators can appear multiple times, like $g_1^2 g_2^{-1} g_3 g_1^3$.

Example 4.21. The dihedral group D_n is generated by r_1 and s_1. In particular, every rotation r_n is $(r_1)^n$ and every reflection s_n can be written as $(r_1)^{n-1} s_1$.

Example 4.22. The group $\mathbb{Z}_2 \times \mathbb{Z}_2$ is not cyclic. However, it can be generated by the two elements $(1,0)$ and $(0,1)$. That is, any (a,b) in $\mathbb{Z}_2 \times \mathbb{Z}_2$ can be written as $a(1,0) + b(0,1)$. In fact, $(1,0)$ and $(0,1)$ always work as generators for $\mathbb{Z}_m \times \mathbb{Z}_n$. Sometimes, just $(1,1)$ works, in which case the group is cyclic. But this only happens if $\gcd(m,n) = 1$.

Example 4.23. The group of rationals $\mathbb{Q}$ under addition is not finitely generated. As an example to help us understand why, suppose $\mathbb{Q}$ has generators 1, $1/2$, and $1/5$. There is no way to write $1/3$ in terms of these elements, so they cannot form a set of generators. And in general, no finite list of elements could ever be enough to generate $\mathbb{Q}$, as the fraction $1/p$, for some prime p that's not a factor of any of the denominators in the finite list, would be impossible to write using the elements of the list.

Below is a theorem classifying abelian groups that may be infinite, but still have a finite number of generators. We will omit the proof.

Theorem 4.9 (Fundamental Theorem of Finitely-Generated Abelian Groups)**.** *Every finitely-generated abelian group is isomorphic to a group of the following form, where p_1, p_2, etc. are primes, but with possible repeats:*

$$\mathbb{Z}_{p_1^{\alpha_1}} \times \mathbb{Z}_{p_2^{\alpha_2}} \times \ldots \mathbb{Z}_{p_k^{\alpha_k}} \times \mathbb{Z} \times \cdots \times \mathbb{Z}.$$

The only difference between the finite and finitely-generated cases is the addition of a finite number of products of $\mathbb{Z}$ at the end.

Non-abelian groups

For non-abelian groups things are more tricky. There is no nice theorem like we had for abelian groups.

The smallest non-abelian group is D_3, with 6 elements. For every even integer $2n$ thereafter, D_n gives an example of a non-abelian group. The first non-abelian group that is not isomorphic to a dihedral group is the *quaternion group*, Q_8. Quaternions are a four-dimensional generalization of imaginary numbers. Its elements are $\{1,-1,i,-i,j,-j,k,-k\}$. The elements i, j, and k are thought of as three different square roots of -1. That is, we have $i^2 = j^2 = k^2 = -1$. We also have $ijk = -1$. The quaternion group also satisfies the following rules: $ij = k$, $jk = i$, and $ki = j$. If we reverse the orders of these, the signs reverse: $ji = -k$, $kj = -i$, and $ik = -j$. Quaternions have applications to computer graphics and physics.

After the quaternion group, the next new non-abelian group we meet is A_4, with 12 elements. There is one other non-abelian group of order 12, $S_3 \times \mathbb{Z}_2$. This is also described as the *semi-direct product* of $\mathbb{Z}_3$ and $\mathbb{Z}_4$. The definition of semi-direct products is more than we want to get into here, other than to say they are like direct products with a "twist" added. Many non-abelian groups are semi-direct products. The dihedral group D_n is actually a semi-direct product of Z_n with Z_2.

The next place we get new non-abelian groups is order 16, where there are 9 groups. One of them is D_8. Two are direct products with $\mathbb{Z}_2$, namely $D_4 \times \mathbb{Z}_2$ and $Q_8 \times \mathbb{Z}_2$. Another is Q_{16}, which is a generalization of the quaternion group. The rest are various semidirect products. At size 18, we pick up $S_3 \times \mathbb{Z}_3$ and another semidirect product. The first non-abelian group with an odd number of elements shows up at 21 elements. It's another semidirect product.

The most important infinite non-abelian group is the general linear group $GL(n, F)$ of invertible $n \times n$ matrices with entries coming a set F that is often $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}_p$, where p is prime. The group operation is matrix multiplication. The general linear group is widely used elsewhere in mathematics, as are a few important subgroups, such as the special linear group of matrices with determinant 1 and the orthogonal group of orthogonal matrices.

Chapter 5

Rings and Fields

5.1 Introduction to Rings

So far, we have explored one type of algebraic structure, a group. There are many other interesting structures. We turn our attention now to rings. The big difference between groups and rings is rings have two operations, inspired by addition and multiplication of integers and real numbers. Here is the formal definition of a ring.

Definition 5.1. *A* ring *is a set R together with two operations, addition and multiplication. Addition of elements a and b is denoted by $a+b$, while multiplication of a and b is denoted by placing a and b next to each other as ab. These operations must satisfy the following properties:*

1. *R is closed under addition. That is, if $a, b \in R$, then $a+b \in R$.*
2. *Addition is associative. That is, $(a+b)+c = a+(b+c)$ for all $a, b, c \in R$.*
3. *There exists an additive identity, denoted 0, with the property that $a+0 = a$ and $0+a = a$ for all $a \in R$.*
4. *For each $a \in R$, there exists an additive inverse, denoted $-a$, with the property that $a+-a = 0$ and $-a+a = 0$.*
5. *Addition is commutative. That is, $a+b = b+a$ for all $a, b \in R$.*
6. *R is closed under multiplication. That is, $ab \in R$ for all $a, b \in R$.*
7. *Multiplication is associative. That is, $(ab)c = a(bc)$ for all $a, b \in R$.*
8. *The operations satisfy the distributive laws $a(b+c) = ab+ac$ and $(b+c)a = ba+ca$ for all $a, b, c \in R$.*

The first five rules simply say that R is an abelian group under addition. However, for multiplication, we just require closure and associativity. There is not necessarily an identity or inverses. The distributive laws tell how the multiplication and addition operations interact. As a notational note, we will often write $a-b$ in place of $a+-b$ and refer to that operation as *subtraction*.

Example 5.1. The familiar sets of numbers $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all rings.

Example 5.2. The integers modulo n, $\mathbb{Z}_n$, are a ring with addition mod n and multiplication mod n as the operations.

Example 5.3. The set of all $n \times n$ matrices is a ring. Addition and multiplication are defined as they are in linear algebra.

Example 5.4. The set of all functions from $\mathbb{R}$ to $\mathbb{R}$ is a ring. The addition of functions f and g is defined such that $f+g$ is the function satisfying $(f+g)(x) = f(x)+g(x)$ for all $x \in \mathbb{R}$, and multiplication is defined such that fg is the function satisfying $(fg)(x) = f(x)g(x)$ for all $x \in \mathbb{R}$. For instance, if f and g are defined by $f(x) = x^2$ and $g(x) = x^3$, then $(f+g)(x) = x^2 + x^3$ and $(fg)(x) = x^5$.

Example 5.5. Given a ring R, the notation $R[x]$ denotes the ring of polynomials whose coefficients come from R. For instance, $\mathbb{Z}[x]$ is all polynomials with integer coefficients. These are things like $x+2$, $3x^2+4x+9$, and x^5-9x^2. Addition and multiplication are defined using the familiar rules from high school algebra. For instance, $(x^2+2x+2)+(x^3+2x^2+5)$ is x^3+3x^2+2x+7, and $(x^2+2x)(x^2+x+1) = x^4+x^3+x^2+2x^3+2x^2+2x = x^4+3x^3+3x^2+2x$.

It is important to remember that the coefficients come from the ring R, which affects the operations. For example, the ring $\mathbb{Z}_3[x]$ is polynomials whose coefficients come from $\mathbb{Z}_3$. That is, the coefficients are always 0, 1, or 2, and the operations on coefficients are all done mod 3. For instance, $(2x^2+x+2)+(2x^2+2x) = x^2+2$.

Example 5.6. The Gaussian integers, denoted $\mathbb{Z}[i]$, are all numbers of the form $a+bi$ for $a, b \in \mathbb{Z}$, with i denoting the imaginary number $\sqrt{-1}$. We define $(a_1+b_1i)+(a_2+b_2i)$ by adding the real and imaginary parts separately to get $(a_1+a_2)+(b_1+b_2)i$. We define $(a_1+b_1i)(a_2+b_2i)$ by FOIL-ing out the product to get $(a_1a_2-b_1b_2)+(a_1b_2+a_2b_1)i$.

For instance, $(2+9i)+(3+7i) = 5+16i$, and $(2+9i)(3+7i) = -57+41i$.

The Gaussian integers are a ring with these operations. It's a good exercise to try to verify each of the properties. Gaussian integers turn out to be important in number theory and have applications to cryptography.

Example 5.7. Consider the set $\mathbb{Z}[\sqrt{2}] = \{a+b\sqrt{2} : a, b \in \mathbb{Z}\}$, Define addition on this set by $(a_1+b_1\sqrt{2})+(a_2+b_2\sqrt{2}) = (a_1+a_2)+(b_1+b_2)\sqrt{2}$, and define multiplication by $(a_1+b_1\sqrt{2})(a_2+b_2\sqrt{2}) = (a_1a_2+2b_1b_2)+(a_1b_2+a_2b_1)\sqrt{2}$, which we get by FOIL-ing out the product. Note the similarity to the previous example. The only difference is using $\sqrt{2}$ in place of i. This set is a ring under these operations. It is a good exercise to go through and verify each of the properties.

Example 5.8. Given two rings A and B, their direct product is the ring whose set is the Cartesian product $A \times B$. The addition of two elements is $(a_1, b_1)+(a_2, b_2)$ defined as (a_1+a_2, b_1+b_2), with the addition from A used for the first half and the addition from B used for the second half. Similarly, multiplication is defined by $(a_1, b_1)(a_2, b_2) = (a_1a_2, b_1b_2)$, with the multiplication from A used for the first half and the multiplication from B used for the second half.

For example, in $\mathbb{Z}_4 \times \mathbb{Z}_7$, we have $(2,4)+(3,6) = (1,3)$ and $(2,4)(3,6) = (2,3)$.

Here are a few simple properties that rings satisfy. Their proofs follow directly from similar facts for groups (since a ring is a group under addition).

Proposition 5.1. *In a ring, the following are true.*

1. *The additive identity 0 is unique.*
2. *Additive inverses are unique.*
3. *If $a+c = b+c$, then $a = b$. Likewise, if $c+a = c+b$, then $a = b$.*
4. *$a+x = b$ has the unique solution $x = b+-a$. Likewise, $x+a = b$ has the unique solution $x = -a+b$.*
5. *$-(-a) = a$.*
6. *$-(a+b) = -a+-b$.*

Here are some more simple facts. We need to be careful in proving them. It's tempting to want to use facts from basic algebra in the proofs, but we can't since those basic algebra facts are what we're proving. All we have to

work with are the definition of a ring along with the previous proposition. This requires some cleverness to get to get things to work out.

Proposition 5.2. *In a ring, the following are true.*

1. $(a+b)(c+d) = ac + ad + bc + bd$
2. $0a = 0$ *and* $a0 = 0$.
3. $a(-b) = -(ab)$ *and* $(-a)b = -(ab)$.
4. $(-a)(-b) = ab$

Proof. For #1, working from the right side we have

$$ac + ad + bc + bd = a(c+d) + b(c+d) = (a+b)(c+d).$$

The first equality uses the left distributive rule twice to essentially factor out a from the first two terms and b from the second two. In the second equality, we use the right distributive law to factor out $c+d$ on the right.

For #2, let's prove that $0a = 0$. Since $0+0 = 0$, we can write $0a = (0+0)a$. Use the distributive law to write the right side as $0a + 0a$. Thus, we have $0a = 0a + 0a$. The inverse of $0a$ is $-(0a)$. Add this to both sides to get $0a + -(0a) = 0a + 0a + -(0a)$. By associativity and the definition of additive inverses, this simplifies into $0 = 0a$, as desired. A very similar proof can be used to show $a0 = 0$.

For #3, property #2 tells us that $a0 = 0$. We can rewrite $a0$ as $a(b + -b)$. Then use the distributive law to write the right side as $ab + a(-b)$. Thus $ab + a(-b) = 0$. This tells us $a(-b)$ must be the inverse of ab, which is $-(ab)$. That is, $a(-b) = -(ab)$. A similar proof shows $(-a)b = -ab$.

For #4, by property #2, we have $0 = (0)(0)$. We can write $0 = a + -a$ and $0 = b + -b$ and then use property #1 to get the following:

$$0 = (0)(0) = (-a + a)(-b + b) = (-a)(-b) + (-a)b + a(-b) + ab.$$

By property #3, we have $(-a)b = -(ab)$ and $a(-b) = -(ab)$. Thus the above can be rewritten as $0 = (-a)(-b) + -(ab) + -(ab) + ab$. We can use associativity to group the last two terms together. They are additive inverses, so they will add to 0. Then, since anything plus 0 is itself, we are left with $0 = (-a)(-b) + -(ab)$. This says that $(-a)(-b)$ and $-(ab)$ are additive inverses. But inverses are unique, and the additive inverse of ab is $-(ab)$. Thus, $(-a)(-b) = ab$, as desired. □

Notational note We end this section with a quick notational note. We will sometimes want to repeatedly add the same element. We will denote $a + a + \cdots + a$, where the addition is done n times, as $n \cdot a$. In this expression, n is an integer and *not* an element of the ring. We will use the dot notation to indicate that what happens here is not ring multiplication. It's unfortunate that this notation looks so much like ring multiplication, but that's just the way people do things. So just be careful of it.

5.2 Important types of rings

While general rings are interesting, certain rings have a multiplication operation that has stronger properties, making them useful for more things.

Definition 5.2. *Below are several important types of rings.*

1. *A* ring with identity *(also called a* ring with unity*) is a ring R that has a multiplicative identity. The multiplicative identity is denoted by 1, and it has the property that $a1 = a$ and $1a = a$ for all $a \in R$.*
2. *A* commutative ring *is a ring R in which the multiplication operation is commutative. That is, $ab = ba$ for all $a, b \in R$.*

3. *An* integral domain *is a commutative ring with identity in which, whenever $ab = 0$, then a or b must be 0.*
4. *A* division ring *is a ring with identity in which every nonzero element has a multiplicative inverse. That is, if $a \neq 0$, then there exists an element a^{-1} such that $aa^{-1} = 1$ and $a^{-1}a = 1$.*
5. *A* field *is a commutative ring with identity in which every nonzero element has a multiplicative inverse.*

The strongest of these are fields, in which the multiplication operation has an identity, is commutative, and has inverses for everything except 0. As we'll see later, fields are also integral domains. Division rings are nearly fields; they just lack commutativity. Below we will look at examples of each of these types of rings.

Rings with identity and commutative rings

The familiar rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all commutative rings with identity. So are the rings $\mathbb{Z}_n$ for all positive integers n. On the other hand, the ring of $n \times n$ matrices has an identity, but it is not commutative for $n > 1$.

The polynomial rings $R[x]$ are commutative if and only if R is. They are rings with identity, with the identity being the constant polynomial $p(x) = 1$.

Matrices are one very useful example of a non-commutative ring, but it's harder to find useful examples of rings that don't have a multiplicative identity. One example is the even integers. They do form a ring, but they exclude 1.

Integral domains

The definition of an integral domain involves the expression $ab = 0$. This brings up an important concept.

Definition 5.3. *In a ring R, an element a is called a* zero divisor *if $a \neq 0$ and $ab = 0$ for some non-zero $b \in R$.*

Zero divisors don't exist in the familiar rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. If two numbers multiply to give 0 in any of those rings, then at least one of them must be 0. Thus, those four rings are all integral domains.

A simple example of a ring that is not an integral domain is $\mathbb{Z}_6$. In this ring, we have $(2)(3) = 0$, since the multiplication is done modulo 6. We also have $(4)(3) = 0$. In general, $\mathbb{Z}_n$ is not an integral domain if n is composite since if $n = ab$ with $a, b \neq 1$, then ab will come out to 0 in $\mathbb{Z}_n$.

However, $\mathbb{Z}_p$ is an integral domain whenever p is prime. This is because the only way for $ab = 0$ in $\mathbb{Z}_p$ is if ab is a multiple of p, and this can't happen since a and b have no factors besides 1 in common with p.

As another example, $n \times n$ matrices for $n > 1$ aren't an integral domain. Part of the problem is they aren't commutative, and commutativity is a requirement for an integral domain. But they also have zero divisors. For instance, the product below comes out to 0:

$$\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

If R is an integral domain, then the ring of polynomials $R[x]$ is also an integral domain. The Gaussian integers $\mathbb{Z}[i]$ are an integral domain, as is $\mathbb{Z}[\sqrt{2}]$.

Fields

In a field, we have multiplicative inverses. Specifically, the following term is used.

Definition 5.4. *In a ring with identity R, a* unit *is an element that has a multiplicative inverse. That is, a is a unit if there exists an element $a^{-1} \in R$ such that $aa^{-1} = 1$ and $a^{-1}a = 1$.*

A ring such as $\mathbb{Z}_6$ has units, namely 1 and 5. But the other elements are not units. For instance, there is nothing to multiply 2 by that brings us back to 1. A field is a ring in which everything except 0 is a unit.

Of our familiar rings, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields since every nonzero element a has a multiplicative inverse $1/a$. However, $\mathbb{Z}$ is not a field since its only units are 1 and -1.

Looking at modular arithmetic, $\mathbb{Z}_p$ is a field if p is prime. If n is not prime, then not everything in $\mathbb{Z}_n$ has an inverse, so we don't get a field. Sometimes people refer to $\mathbb{Z}_p$ using the notation $\mathbb{F}_p$ or $GF(p)$. If n is not prime but is of the form p^k for some prime p, then there is a field with n elements, called a Galois field and denoted $GF(n)$. Up to isomorphism, these Galois fields are the only finite fields. These fields are defined in terms of polynomial rings. They are widely used in cryptography and coding theory. As another example, the ring $\mathbb{Q}[\sqrt{2}]$, that we looked at earlier, is a field.

Division rings

Division rings are like fields without commutativity. They are not as important as fields. Probably the most well-known example of a division ring is the quaternions. These are all numbers of the form $a+bi+cj+dk$, where i, j, and k are all square roots of -1, and $a, b, c, d \in \mathbb{R}$. The rules for multiplying i, j, and k were given at the end of Section 4.4.

Facts about integral domains and fields

As we move from rings to integral domains to fields, we get more and more properties that allow us to do more operations familiar from high school algebra. For example, one of the things we use all the time when solving equations in ordinary algebra is the fact that if $ab = 0$ then a or b must be 0. For instance, to solve the equation $x^2 - x = 0$, we factor it as $x(x-1) = 0$ and then conclude that $x = 0$ or $x - 1 = 0$ to get the solutions $x = 0$ and $x = 1$. This doesn't work in all rings, which makes solving equations trickier. However, if a ring is an integral domain then it does work.

Here is another familiar fact from high school algebra that doesn't work in every ring, but it does work in any integral domain.

Proposition 5.3. *In an integral domain, if $ca = cb$ and $c \neq 0$, then $a = b$. Likewise, if $ac = bc$ and $c \neq 0$, then $a = b$.*

Proof. We will just prove the first part. The second part has a similar proof. We are given that $ca = cb$. We can subtract cb from both sides to get $ca - cb = 0$. Factor out c to get $c(a-b) = 0$. Since we are in an integral domain, and c is given to not be 0, we must have $a - b = 0$. Add b to both sides to conclude that $a = b$. □

This is called a cancellation rule. Note that this doesn't necessarily work in non-integral domains. For instance, in $\mathbb{Z}_{12}$, we can have $2a = 2b$ but $a \neq b$ since $2(2) = 2(8)$ in $\mathbb{Z}_{12}$.

The reason why we categorize things into rings, integral domains, and fields is that each one has stronger properties than the others. Many systems of arithmetic fall into one of these categories, and anything we prove about one of those categories applies to all those systems. In particular, since cancellation works in an integral domain, we now know we can use cancellation when working with ordinary sets like $\mathbb{Z}$ or $\mathbb{R}$, but also for modular arithmetic modulo a prime and for the polynomial ring $\mathbb{Z}[x]$.

Proposition 5.4. *Fields are integral domains.*

Proof. Suppose we have the product $ab = 0$. To show our field is an integral domain, we need to show that we must have $a = 0$ or $b = 0$. If $a \neq 0$, then a must have a multiplicative inverse, a^{-1}, since we are in a field. Multiply both sides of the equation by a^{-1} to get $b = 0$. Thus either a is 0 or else b must be 0. □

Proposition 5.5. *If a is a unit in a ring, it can't also be a zero divisor.*

Proof. Suppose a is a unit and suppose $ab = 0$. Since a is a unit, its multiplicative inverse a^{-1} exists. Multiply both sides by a^{-1} to get $a^{-1}(ab) = 0$. This simplifies into $b = 0$. Thus, it can't happen that $ab = 0$ for $b \neq 0$, which means a is not a zero divisor. □

Proposition 5.6. *In a finite ring R, every nonzero element is either a unit or a zero divisor.*

Proof. First note by Proposition 5.5, no element can be both a zero divisor and a unit. Next, suppose a is not a zero divisor. We will show a must be a unit. This will complete the proof by showing it is not possible for something to be neither a unit nor a zero divisor. Let $b, c \in R$, and suppose $ab = ac$. Subtracting ac from both sides and factoring gives $a(b-c) = 0$. Since a is not a zero divisor, we must have $b - c = 0$ and hence $b = c$. Thus, the products of a by each element of R are all different from each other. Since the ring is finite, that means every element of R is some product of a. In particular, $ax = 1$ for some x. This means that $x = a^{-1}$ since inverses are unique. □

This fact is not necessarily true in infinite rings. For instance, in $\mathbb{Z}$, the only units are ± 1 and there are no zero divisors.

Note: If $ax = 1$, then we must also have $xa = 1$ even if the ring is not commutative. This is because if $ya = 1$ then $yax = x$ and using $ax = 1$ in this gives $y = x$.

Proposition 5.7. *Every finite integral domain is a field.*

Proof. An integral domain has no zero divisors by definition, and Proposition 5.6 tells us that in any finite ring, every nonzero element is either a zero divisor or a unit. So every nonzero element must be a unit. □

Proposition 5.8. *The set of all units in a ring with identity form a group with the multiplication operation.*

Proof. We have to show closure, associativity, identity, and inverses. The first two come right from the definition of a ring since multiplication is defined to have those properties. The identity 1 is a unit since $(1)(1) = 1$. Every element has an inverse by definition since the only thing in the group are units. □

Characteristic

People are interested in the following.

Definition 5.5. *The* characteristic *of a ring is the smallest positive integer n such that $n \cdot a = 0$ for all a. If it does not exist, the ring is said to be of characteristic 0.*

In the definition above, the notation $n \cdot a$ stands for $a + a + \cdots + a$ done n times.

The rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ all have characteristic 0. On the other hand, the ring $\mathbb{Z}_n$ has characteristic n. To see this, note that $n \cdot a \bmod n$ is 0 for any $a \in \mathbb{Z}_n$, so the characteristic must be n or smaller. However, if $k < n$, then $k \cdot 1 = k$, which is not 0 in $\mathbb{Z}_n$. So n is the smallest value that works. An interesting fact that we won't prove is that the characteristic of any integral domain or field must be either 0 or a prime number.

5.3 Polynomial Rings

Earlier, we briefly looked at the polynomial ring $R[x]$ of polynomials with coefficients from a ring R. They turn out to have a lot of applications, so we will look at them in more detail now. We will mostly not prove anything here, though the proofs are generally not difficult. Below we define some important terminology.

Definition 5.6. *Given a ring R, a* polynomial *$p(x)$ in R is a formal expression of the form $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$, where $a_0, a_1, \ldots, a_n \in R$. The set of all such polynomials is denoted $R[x]$. The a_i are called* coefficients*, with a_n being called the* leading coefficient *and a_0 being called the* constant *term. A*

polynomial is called monic *if its leading coefficient is 1. The* degree *of a polynomial $p(x)$, denoted* $\deg(p(x))$ *is the largest k such that the coefficient a_k is nonzero. A* constant polynomial *is one of the form a_0. The* zero polynomial *is a constant polynomial with $a_0 = 0$.*

Note that we leave the degree of the zero polynomial as undefined. Some texts define it to be -1 or $-\infty$. As an example of all the terms, $p(x) = x^4 + 3x^2 + 2x + 9$ is polynomial in $\mathbb{Z}[x]$. Its degree is 4 since x^4 is the highest nonzero power of x. This is a monic polynomial since the leading coefficient (the coefficient of x^4) is 1. The constant term of this polynomial is 9. In $\mathbb{Z}[x]$, the polynomials $f(x) = 3$ and $g(x) = 0$ are constant polynomials, the latter being the zero polynomial.

As noted earlier, $R[x]$ is a ring with the addition and multiplication operations being the usual ways of adding and multiplying polynomials. For instance, in $\mathbb{Z}[x]$, if we have $p(x) = 2x^2 + 3x + 4$ and $q(x) = 5x + 9$, then $p(x) + q(x) = 2x^2 + 8x + 13$ and $p(x)q(x) = (2x^2 + 3x + 4)(5x + 9) = 10x^3 + 33x^2 + 47x + 36$. The one trick is that the operations on the coefficients are done in the ring R. For instance, if we add $2x^2 + x + 1$ and $x^2 + 2$ in the ring $\mathbb{Z}_3[x]$, we get $3x^2 + x + 3$, which just simplifies to x. For the sake of completeness, below are the formal definitions of addition and multiplication, though it is not worth dwelling on them.

Definition 5.7. *Given polynomials $a_m x^m + a_{m-1}x^{m-1} + \cdots + a_2 x^2 + a_1 x + a_0$ and $b_n x^n + b_{n-1}x^{n-1} + \cdots + b_2 x^2 + b_1 x + b_0$.....*

It turns out that $R[x]$ is an integral domain if and only if R is. However, $R[x]$ is never a field since the only potential units of $R[x]$ are constant polynomials that are units of R. For instance, in $\mathbb{R}[x]$, there is no polynomial we can multiply $p(x) = x$ by in order to get 1. The inverse of x would have to be $\frac{1}{x}$, which is not a polynomial at all.

The x in the polynomial is just treated as a symbol. It doesn't actually have a value. However, we can plug in a value, called *evaluating* the polynomial. For instance, if $p(x) = 2x^2 + 9$, then $p(3) = 2(3)^2 + 9 = 171$. Evaluation is occasionally useful in abstract algebra, but usually we are more interested in the polynomials themselves as objects.

Division of polynomials

Recall the division algorithm for integers says that given any $a, b \in \mathbb{Z}$ with $b \neq 0$, there exists a unique quotient q and remainder r such that $a = bq + r$, with $0 < r < |q|$. For instance, if $a = 27$ and $b = 4$, we can write $27 = 4(6) + 3$. That is, the quotient is 6 and reminder is 3. An analog of this also works for polynomials.

Proposition 5.9. *Given any $f(x), g(x) \in R[x]$ with $g(x) \neq 0$, there exists a unique quotient $q(x)$ and remainder $r(x)$ such that $f(x) = g(x)q(x) + r(x)$ with either $r(x) = 0$ or $\deg(r(x) < \deg(g(x))$.*

We can find the quotient and remainder by doing the polynomial long division algorithm you may have learned in high school. It's the direct analog of the grade-school long-division algorithm for integers. Here is an example of dividing $2x^3 + x^2 + 9x + 4$ by $x + 5$.

$$
\begin{array}{r|l}
 & \quad 2x^2 - 9x + 54 \\
\hline
x+5 & \; 2x^3 + x^2 + 9x + 4 \\
 & \underline{-2x^3 - 10x^2} \\
 & \quad\; -9x^2 + 9x \\
 & \quad\; \underline{9x^2 + 45x} \\
 & \quad\quad\; 54x + 4 \\
 & \quad\quad\; \underline{-54x - 270} \\
 & \quad\quad\quad -266
\end{array}
$$

We see from this that the quotient is $2x^2 - 9x + 54$ and the remainder is -266.

Theorem 5.1 (Remainder theorem)**.** *When $f(x)$ in a polynomial ring $R[x]$ is divided by $x - a$, the remainder is $f(a)$.*

Proof. By the division algorithm, we can write $f(x) = (x-a)q(x) + r(x)$ for some polynomials $q(x)$ and $r(x)$. Since the degree of $x-a$ is 1, we know that $r(x)$ must equal some constant $c \in R$. Plugging in a to our expression from the division algorithm gives $f(a) = (a-a)q(a) + r(a) = c$. □

Irreducible polynomials

An important concept in the theory of polynomials is the idea of irreducibility, which is roughly analogous to the concept of prime numbers.

Definition 5.8. *Let F be a field. A polynomial in $F[x]$ is called* irreducible *if there do not exist polynomials $p(x), q(x) \in F[x]$ such that $f(x) = p(x)q(x)$ with both having a degree less than the degree of $f(x)$. If those polynomials do exist, we say $f(x)$ is* factorable.

Often people use the term *reducible* instead of *factorable*, and sometimes people use *unfactorable* instead of *irreducible*.

To make the analogy with primes, in $\mathbb{Z}$, we would say 6 is not prime because we can factor it into smaller integers $2 \cdot 3$, while 7 is prime because we can't factor it into two smaller integers. In $\mathbb{R}[x]$, we can factor $x^2 - 1$ into smaller polynomials $(x-1)(x+1)$, but we can't do so to $x^2 + 1$, so $x^2 + 1$ is irreducible. In $\mathbb{Z}$ we can write $7 = 7 \cdot 1$, but that is not a factorization into two *smaller* integers. Likewise, in $\mathbb{R}[x]$, we can factor $2x^2 + 3$ into $2(x^2 + 1.5)$, but that is not a factorization into two lower-degree polynomials. Degree-zero polynomials behave comparably to how ± 1 behave in $[Z]$. An alternate definition of irreducibility you'll sometimes see is that if $f(x)$ has degree 1 or higher, then $f(x)$ is irreducible if and only if whenever $f(x)$ is written as $p(x)q(x)$, then one of $p(x)$ and $q(x)$ has degree 0.

It's important to note that the ring matters for irreducibility. For instance, $x^2 + 1$ is irreducible in $\mathbb{R}[x]$, but it is not irreducible in $\mathbb{C}[x]$, since there it can be written as $(x-i)(x+i)$. It's also not irreducible in $\mathbb{Z}_5[x]$, since it can be factored into $(x+2)(x+3)$ (since $(x+2)(x+3) = x^2 + 5x + 6$ reduces to $x^2 + 1$ modulo 5).

Just like determining if something is prime or not is not easy, it is not easy to determine if a polynomial is irreducible or not. But there are some tricks people use. One useful idea uses factor theorem given below. It relies on the notion of a *root*. Roots are exactly what you might remember from high school algebra, values that you can plug in to a polynomial that make it 0. Here is the formal definition followed by the factor theorem.

Definition 5.9. *Let F be a field and let $f(x) \in F[x]$. A* root *of $f(x)$ is a value $a \in F$ such that $f(a) = 0$.*

Proposition 5.10 (Factor theorem)**.** *If F is a field and $f(x) \in F[x]$, then $f(x)$ is divisible by $x - a$ if and only if a is a root of $f(x)$.*

That is, linear factors correspond directly to roots. The proof of theorem follows directly from the remainder theorem (Theorem 5.1). The theorem shows that polynomials of degree 2 or 3 are irreducible if and only if they have no roots. This is because for a degree-2 polynomial to be factorable, the two factors would have to be linear factors, and for a degree-3 polynomial to be factorable, it would have to factor into one linear and one quadratic factor. However, degree-4 and higher polynomials can be factorable and still have no roots. For instance, $x^4 + 3x^2 + 2$ in $\mathbb{R}x$ factors into $(x^2 + 1)(x^2 + 2)$, and it has no roots in $\mathbb{R}$.

Example 5.9. Let's check if $f(x) = x^3 + 4x + 1$ is irreducible in $\mathbb{Z}_5[x]$. Since it has degree 3, we can just check to see if it has any roots in $\mathbb{Z}_5$. Since $\mathbb{Z}_5$ is $\{0, 1, 2, 3, 4\}$, we just have to check those five values. It is quick to calculate that $f(0) = 1$, $f(1) = 1$, $f(2) = 2$, $f(3) = 0$, and $f(4) = 1$. So this polynomial has a linear factor of $x - 3$, and we can use polynomial long division to factor $f(x)$ into $(x-3)(x^2 + 3x + 3)$.

Example 5.10. The polynomial $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$. It is degree 2, and its roots are $\pm\sqrt{2}$, neither of which are in $\mathbb{Q}$.

Below is another trick people use to check for irreducibility. It is sometimes taught in precalculus classes.

Proposition 5.11 (Rational roots theorem)**.** *Let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ be in $\mathbb{Z}[x]$, with a_0 not 0. If $f(x)$ has a root in $\mathbb{Q}$, then it has a root in $\mathbb{Z}$ and that root must be a divisor of a_0.*

Example 5.11. Let's use the theorem to show that $f(x) = x^4 - 2x^2 + 8x + 1$ is irreducible in $\mathbb{Q}[x]$. By the theorem, if it has a root in $\mathbb{Q}$, then one of its roots must be a divisor of constant term, 1. That is, it must be ± 1. But plugging in, we can see that neither $f(1)$ nor $f(-1)$ are 0.

Here is a bit of a more sophisticated trick people use.

Proposition 5.12 (Eisenstein criterion)**.** *Let $f(x) = a_n x^n + \cdots + a_1 x + a_0$ be in $\mathbb{Z}[x]$. Suppose there exists a prime p such that p is a divisor of all the coefficients except a_n, p is not a divisor of a_n, and p^2 is not a divisor of a_0, then $f(x)$ is irreducible in $\mathbb{Q}[x]$.*

Example 5.12. Consider $f(x) = 25x^3 - 9x^4 - 3x^2 - 12$. To use Eisenstein's criterion, we need a prime that is divisible by all the coefficients except the leading coefficient. The prime $p = 3$ fits that. We then need to make sure $p^2 = 9$ is not a divisor of the constant term -12, and it isn't. Thus by Eisenstein's criterion, the polynomial is irreducible.

Note that this is a one-way criterion only. If we can find the prime the Eisenstein criterion wants, then we know the polynomial is irreducible. But if no such prime exists, that doesn't mean the polynomial is factorable. If it doesn't work on $f(x)$, sometimes people shift to $f(x + a)$ for some constant a and try it on that. This is because $f(x)$ factors into $g(x)h(x)$ if and only if $f(x + a)$ factors into $g(x + a)h(x + a)$. For instance, Eisenstein's criterion won't work to show $f(x) = x^2 + x + 1$ is irreducible in $\mathbb{Z}[x]$. But if we try it on $f(x + 1)$, which is $x^2 + 3x + 3$, we see that the prime $p = 3$ works in the criterion to show $f(x + 1)$ is irreducible. Hence $f(x)$ is irreducible as well.

Proposition 5.13. *Let $f(x) = a_n x^n + \ldots a_1 x + a_0$ be in $\mathbb{Q}[x]$ with integer coefficients. Let $\overline{f}(x) = \overline{a}_n x^2 + \cdots + \overline{a}_1 x + \overline{a}_0$ be the polynomial in $\mathbb{Z}_p[x]$ where $\overline{a}_i = a_i \bmod p$ for $i = 0, 1, \ldots, n$. If $\overline{f}$ has the same degree as f and $\overline{f}$ is irreducible in $\mathbb{Z}_p[x]$, then f is irreducible in $\mathbb{Q}[p]$.*

Example 5.13. Let's use this to show $f(x) = x^4 + 3x^2 + 1$ is irreducible. Let's try mod 2 first. After reducing all the coefficients mod 2, we get $\overline{f}(x) = x^4 + x^2 + 1$. To see if it is irreducible, we first check for roots. In $\mathbb{Z}_2$, the only possible roots to check are 0 and 1. We get $\overline{f}(0) = 1$ and $\overline{f}(1) = 1$, so there are no linear factors. Since $\overline{f}$ has degree 4, we still have to check if it factors into a product of quadratics. The only possible quadratics mod 2 are x^2, $x^2 + 1$, and $x^2 + x + 1$. The first two are factorable, so we don't need to worry about them since if $\overline{f}$ factored into a product with at least one of those factors, then it would have a linear factor, and we already know there are no linear factors. So we just have to check if $\overline{f}(x)$ is divisible by $x^2 + x + 1$. We can use long division, and if we do so, we do find $\overline{f}(x) = (x^2 + x + 1)(x^2 + x + 1$. So our mod 2 test is inconclusive.

So move on to mod 3. Here $\overline{f}(x) = x^4 + 1$. First compute $\overline{f}(0) = 1$, $\overline{f}(1) = 2$, $\overline{f}(2) = 2$, so there are no linear factors. The three possible irreducible quadratics in $\mathbb{Z}_3[x]$ are $x^2 + 1$, $x^2 + x + 2$, and $x^2 + 2x + 2$. If we try long division by each of them, it turns out none go in evenly. Thus, $\overline{f}(x)$ is irreducible in $\mathbb{Z}_3$ and thus irreducible in $\mathbb{Q}[x]$.

There are other techniques people use, but for simplicity we've just given a small sampling. One reason people are interested in irreducibility is that irreducible polynomials can be used to construct finite fields of order p^k, where p is prime. These turn out to have many applications in cryptography and coding theory. We will look at this a bit later.

5.4 Subrings, Ring Homomorphisms, and Ideals

Subrings

Definition 5.10. *A* subring *of a ring R is a subset of R that is also a ring itself using the same operations as R.*

Because a subring uses the same operations as the parent ring, it inherits associativity, commutativity of addition, and the distributive property from it. Therefore, to show something is a subring, we don't have to show all eight properties, just the ones below:

- S is closed under addition
- S is closed under multiplication
- $0 \in S$
- If $a \in S$, then $-a \in S$ too.

Note that using the ab^{-1} criterion for groups, we could also show something is a subring just by showing $a - b \in S$ whenever $a, b \in S$ and that S is closed under multiplication.

Example 5.14. The set of even integers is a subring of $\mathbb{Z}$. It is closed under addition and multiplication since if $a = 2j$ and $b = 2k$ are even integers, then $a + b = 2j + 2k = 2(j + k)$ and $ab = (2j)(2k) = 2(2jk)$ are both even. The additive identity 0 is 2(0), so it is even. And the additive inverse of $a = 2j$ is $-2j$, which is $2(-j)$.

Ring Homomorphisms

Definition 5.11. *Let R and S be rings. A* ring homomorphism *is a function $\phi : R \to S$ satisfying $\phi(r+s) = \phi(r) + \phi(s)$ and $\phi(rs) = \phi(r)\phi(s)$ for all $r, s \in R$. If ϕ is one-to-one and onto, then it is a* ring isomorphism.

The definition of ring homomorphisms and isomorphisms almost directly parallels the group versions of those definitions. We now require the function to preserve both the addition and multiplication operations. Just like with group homomorphisms, the additive identity of R must get mapped to the additive identity of S. However, if both rings have a multiplicative identity, there is no guarantee that ϕ maps the one to the other. Unlike with groups, it isn't a consequence of the homomorphism formulas.

Example 5.15. The function $\phi : \mathbb{Z} \to \mathbb{Z}_n$ given by $\phi(k) = k \bmod n$ is a ring homomorphism.

Example 5.16. The function $\phi : \mathbb{Z}[x] \to \mathbb{Z}$ given by $\phi(p(x)) = p(1)$ is a homomorphism. It is an example of something called an *evaluation homomorphism*. As an example of it, $\phi(x^2 + 3x + 4)$ gets sent to $1^2 + 3(1) + 4 = 8$.

Isomorphisms Just like with groups, to show two rings are not isomorphic, find an algebraic property that one satisfies and the other doesn't. These could be things like commutativity, being an integral domain, orders of elements under one or the other operation, etc.

Example 5.17. As groups under addition $\mathbb{Z}$ and the even integers $2\mathbb{Z}$ are isomorphic, but they are not isomorphic as rings. The group isomorphism we used in an earlier example is $\phi(n) = 2n$. However, this fails the $\phi(mn) = \phi(m)\phi(n)$ rule as $\phi((1)(2)) = 4$, but $\phi(1)\phi(2) = 8$. There is in fact, no function that can act as an isomorphism since $\mathbb{Z}$ contains the multiplicative identity 1, but $2\mathbb{Z}$ does not.

Example 5.18. The groups $\mathbb{Z}_2 \times \mathbb{Z}_3$ and $\mathbb{Z}_6$ are isomorphic as rings. The function $\phi : \mathbb{Z}_6 \to \mathbb{Z}_2 \times \mathbb{Z}_3$ given by $\phi(n) = (n \bmod 2, n \bmod 3)$ is the isomorphism.

Kernel Just like with group homomorphisms, the kernel is important for ring homomorphisms. The kernel is all the elements sent to the additive identity 0.

Example 5.19. For the homomorphism $\phi : \mathbb{Z} \to \mathbb{Z}_n$ given by $\phi(k) = k \bmod n$, the kernel is all the multiples of n.

Example 5.20. The kernel of the evaluation homomorphism $\phi : \mathbb{Z}[x] \to \mathbb{Z}$ given by $\phi(p(x)) = p(1)$ is all the polynomials divisible by $x - 1$. Clearly if $p(x)$ is divisible by $x - 1$, then $\phi(p(x)) = 0$. On the other hand, by the factor theorem, if $p(1) = 0$, then $p(x)$ is divisible by $x - 1$.

Ideals

Normal groups are an important concept in group theory that allow us to define the notion of a quotient group. The analogous concept for rings are what are called *ideals*.

Definition 5.12. *A subset I of a ring R is called an* ideal *if it forms an additive subgroup and has the property that whenever $a \in R$ and $i \in I$, then $ai \in I$ and $ia \in I$.*

Note that since we will only be working with commutative rings for the rest of these notes, we only have to check if $ai \in I$ and we will not worry about checking $ia \in I$ since $ai = ia$ in a commutative ring.[1]

The $ai \in I$ property is about how ideals absorb elements. If we multiply a random element of the ring by an element in the ideal, the product gets sucked into the ideal. Let's look at a couple of examples.

Example 5.21. In $\mathbb{Z}$, the set I of all multiples of 3 is an ideal. To show this, note first that it is an additive subgroup since it contains 0, the sum of two multiples of 3 is another multiple of 3, and the additive inverse of $3k$ is $-3k$, which is also a multiple of 3. Next, take $a \in \mathbb{Z}$ and $i \in I$. Then $i = 3k$ for some k, and we have $ai = 3(ak)$ is a multiple of 3, hence in I.

Example 5.22. The kernel of a ring homomorphism is always an ideal. Let $\phi : R \to S$ be a ring homomorphism. The kernel is $K = \{r \in R : \phi(r) = 0\}$. By Proposition 2.17, it is an additive subgroup. To show the absorption property, suppose $a \in R$ and $k \in K$. Then $\phi(ak) = \phi(a)\phi(k) = (\phi(a))(0) = 0$, so $ak \in K$.

Example 5.23. Let I be the set of all polynomials in $\mathbb{Z}[x]$ whose constant term is even. Examples include $x^2 + 3x + 6$ and $x^3 + 10$ (whose constant terms 6 and 10 are even). First, we have that I is an additive subgroup because the identity polynomial, $p(x) = 0$, has an even constant term, if we add any two polynomials in I we get another one whose constant term is even, and the additive inverse will also have an even constant term. For the absorption property, when we multiply two polynomials, the constant term in the product is the product of their constant terms, so if one is even, then the product will be as well.

The first example, of all the multiples of 3, generalizes to the following:

Definition 5.13. *Let R be a ring with $a \in R$. The set $\langle a \rangle = \{ra : r \in R\}$ is called a* principal ideal.

That is, the set of all multiples of an element of a ring always forms an ideal. The proof is very similar to the argument used in the multiples of 3 example. Not all ideals are principal, as the polynomial example above shows, but the most common ideals you'll run into are often principal ideals. Note that many books will use the notation (a) instead of $\langle a \rangle$ for principal ideals.

Fields don't have any interesting ideals.

Proposition 5.14. *If F is a field, then the only ideals are $\langle 0 \rangle$ and F itself.*

Proof. It is pretty quick to check that $\langle 0 \rangle$ and F are ideals in any ring. Suppose I is an ideal that contains an element $a \neq 0$. Since F is a field, a has a multiplicative inverse, and by the absorption property $1 = aa^{-1} \in I$. But if $1 \in I$, then all of F is in I as well, since if $b \in F$, then $(b)(1) \in I$ by the absorption property. □

One other quick definition:

Definition 5.14. *A* maximal ideal *in a ring R is an ideal $I \neq R$ such that there does not exist another ideal J with $I \subseteq J$.*

That is, maximal ideals are as big as they can be. They can't be contained in another ideal. For instance, in $\mathbb{Z}$, $\langle 2 \rangle$ is maximal, but $\langle 4 \rangle$ is not since it is contained in $\langle 2 \rangle$.

[1] There are also concepts of *left ideal* and *right ideal* for when we are only guaranteed $ai \in I$ (left ideal) or $ia \in I$ (right ideal).

5.5 Quotient Rings

Recall that if N is a normal subgroup of a group G, then the quotient group G/N is the group whose elements are the cosets of N, with the group operation given by $(aN)(bN) = (ab)N$. The quotient group is only defined if N is normal. If it is not normal, then the group operation does not work. The idea of a quotient group is that everything in N gets identified together down to the identity of G/N, and the structure of G/N based on the remainders modulo N in a certain sense. The group Z_n for instance, is the quotient group $\mathbb{Z}/n\mathbb{Z}$, and its continuous analog $\mathbb{R}/\mathbb{Z}$ is the circle group, which turns out to be useful in Fourier analysis. Quotient groups can give information about the structure of the parent group, and they are often useful on their own. Recall also the first isomorphism theorem, which says that quotient groups are directly related to images of homomorphisms.

A key property of normal subgroups is if $g \in G$ and $n \in N$, then $gng^{-1} \in N$. That is, if we operate by g on n and then try to "undo" the operation of g by operating with its inverse, the result might not be n itself, but it still stays in N. This is a kind of absorption property, like the one we saw for ideals. Another key fact is that the kernel of a group homomorphism is always normal, and the kernel of a ring homomorphism is always an ideal. So, roughly speaking, ideals are the ring equivalent of normal subgroups. They allow us to define the notion of a quotient ring.

Definition 5.15. *Let R be a ring and let I be an ideal of R. The* quotient ring R/I *is the ring whose elements are the cosets $a + I = \{a + i : i \in I\}$ for $a \in R$ and whose operations are defined by $(a + I) + (b + I) = (a + b) + I$ and $(a + I)(b + I) = (ab) + I$.*

Note that when we defined quotient groups, we defined the product in terms of set multiplication. That doesn't quite work so well here since $(a + I)(b + I)$, when thought of as sets, don't quite multiply to give another coset. So instead, we just define $(a + I)(b + I)$ to be $(ab) + I$. This could potentially give a problem depending on what representatives we choose. For instance, if $a, c \in a + I$ and $b, d \in b + I$, we would want $(a + I)(b + I)$ and $(c + I)(d + I)$ to both give the same coset. It's not too hard to show that they do if and only if I is an ideal. The absorption property is what makes this work.

Example 5.24. The prototypical example is $\mathbb{Z}/n\mathbb{Z}$. We can use $0 + n\mathbb{Z}$, $1 + n\mathbb{Z}$, $\ldots$, $(n-1) + n\mathbb{Z}$ as representatives. The ring we get is the same thing as $\mathbb{Z}_n$. Notice that the multiples of n are identified to 0, and the remainders of division by n are what is left.

Example 5.25. Let's look at $\mathbb{Z}_3[x]/\langle x^2 \rangle$. Here $\langle x^2 \rangle$ is all multiples of the polynomial x^2. The quotient ring operation identifies all of them to 0, and what is left are the remainders upon division by x^2. These remainders are all of the form $ax + b$, so the elements of the quotient group are all elements of the form $(ax + b) + \langle x^2 \rangle$. Suppose we want to add $(2x + 1) + \langle x^2 \rangle$ and $(2x + 2) + \langle x^2 \rangle$. The result is $(4x + 3) + \langle x^2 \rangle$, which simplifies to $x + \langle x^2 \rangle$ since we're working mod 3.

Multiplication is more interesting. Suppose we want to multiply $(2x + 1) + \langle x^2 \rangle$ and $(2x + 2) + \langle x^2 \rangle$. The result is $(4x^2 + 5x + 2) + \langle x^2 \rangle$. Here is where the identification comes into play. Since x^2 is identified to 0, we replace the x^2 in the product with 0, and (after reducing mod 3), the result simplifies to $(2x + 2) + \langle x^2 \rangle$.

In general, when working in $R[x]/\langle p(x) \rangle$, we can compute the product of two cosets by multiplying their representatives and then reducing it using the rule that $p(x) = 0$. For instance, if $p(x) = x^3 - 2x + 1$, we can say $x^3 - 2x + 1 = 0$, so $x^3 = 1 - 2x$ and we could replace all the x^3 terms with $1 - 2x$. Let's look at an important application of this.

Example 5.26. Consider $\mathbb{R}[x]/\langle x^2 + 1 \rangle$. Here $\langle x^2 + 1 \rangle$ is all the multiples of the polynomial $x^2 + 1$. The quotient group operation identifies all of those to 0, and what is left are the remainders upon division by $x^2 + 1$. These will all be of the form $ax + b$, so the elements of the quotient group will be all elements of the form $(ax + b) + \langle x^2 + 1 \rangle$. Adding two elements $(a_1 x + b_1) + \langle x^2 + 1 \rangle$ and $(a_2 x + b_2) + \langle x^2 + 1 \rangle$ gives $((a_1 + a_2)x + (b_1 + b_2)) + \langle x^2 + 1 \rangle$. Multiplication is more interesting. We end up with $(a_1 a_2 x^2 + (a_1 b_2 + a_2 b_1)x + b_1 b_2) + \langle x^2 + 1 \rangle$. Now since we identified $x^2 + 1$ to 0, we can treat this as $x^2 + 1 = 0$ or $x^2 = -1$. Thus, everywhere we see an x^2, we can replace it with -1. Using this, our product simplifies to $((a_1 b_2 + a_2 b_1)x + (b_1 b_2 - a_1 a_2)) + \langle x^2 + 1 \rangle$.

The interesting thing is if we multiply two complex numbers $a_1 + b_1 i$ and $a_2 + b_2 i$, we get something very similar, specifically $(b_1 b_2 - a_1 a_2) + (a_1 b_2 + a_2 b_1)i$. Addition of complex numbers also gives something very similar to what we got for addition of cosets. In fact, we have nearly shown that the function $\phi(a + bi) = (a + bx) + \langle x^2 + 1 \rangle$ is an isomorphism between $\mathbb{R}[x]/\langle x^2 + 1 \rangle$ and $\mathbb{C}$.

Example 5.27. Another nice example is $\mathbb{R}[x]/\langle x^2 \rangle$. Here instead of identifying $x^2 + 1$, we identify x^2 to 0. The cosets are of the form $(ax + b) + \langle x^2 \rangle$. Addition is just like the previous example. For multiplication, the product of $(a_1 x + b_1) + \langle x^2 \rangle$ and $(a_2 x + b_2) + \langle x^2 \rangle$ is $(a_1 a_2 x^2 + (a_1 b_2 + a_2 b_1)x + b_1 b_2) + \langle x^2 \rangle$. Since x^2 is identified to 0, we just drop the first term and get $((a_1 b_2 + a_2 b_1)x + b_1 b_2) + \langle x^2 \rangle$.

The elements of this quotient ring are called *dual numbers*. They turn out to be important for something called automatic differentiation, which is a technique used to numerically compute derivatives of complicated functions to as much accuracy as a computer is able to provide.

Finite fields

We have already noted that $\mathbb{Z}_p$ is a field if and only if p is prime. We are interested in if there are any more finite fields. There are, but specifically only of sizes of the form p^k, where p is prime. For instance, there are fields of size 2, 4, 8, 16, 32, ..., fields of size 3, 9, 27, 81, ..., fields of size 5, 25, 125, 625, ..., and so on. There are no fields of sizes that are not a prime to a power, such as 6 or 12. The notation $\mathbb{F}_n$ or $GF(n)$ are commonly used for the finite field of order n. The latter notation is short for *Galois field* since these fields are important in Galois theory, a branch of abstract algebra built around proving that there is no way to write all solutions of all 5th-degree polynomials in terms of radicals, like we can do for quadratics via the quadratic formula.

These finite fields are widely used in cryptography and computing. They are an important component of AES, which is one of the most secure and widely used encryption algorithms. They are used for hash functions such as SHA-256, which are used all over. They are used in coding theory. For instance, almost every packet you send over the internet uses a CRC-32 checksum to check for errors, and that checksum makes use of arithmetic over a finite field.

Let's look at how to construct them. The construction of the finite field of order p^k comes from the quotient ring $\mathbb{Z}_p[x]/\langle g(x) \rangle$ where $g(x)$ is an irreducible polynomial in $\mathbb{Z}_p[x]$ of degree k. Let's look at some examples.

Example 5.28. Let's construct the finite field of order $2^2 = 4$. We need an irreducible polynomial of degree 2 in $\mathbb{Z}_2$. There is one such polynomial: $x^2 + x + 1$. The field will be $\mathbb{Z}_2[x]/\langle x^2 + x + 1 \rangle$. The elements are all of the form $(ax + b) + \langle x^2 + x + 1 \rangle$ with $a, b \in \mathbb{Z}_2$. So these are $(0) + \langle x^2 + x + 1 \rangle$, $(1) + \langle x^2 + x + 1 \rangle$, $(x) + \langle x^2 + x + 1 \rangle$, and $(x + 1) + \langle x^2 + x + 1 \rangle$. For simplicity, we will just write these as 0, 1, x, and $x + 1$. We add them mod 2. For instance, $x + (x + 1) = 2x + 1$, which simplifies to 1. For multiplication, we use the fact that $x^2 + x + 1 = 0$, which tells us $x^2 = -x - 1$, which becomes $x^2 = x + 1$ since we are working mod 2. Thus to multiply $x(x + 1)$, we get $x^2 + x$, and plugging in $x + 1$ for x^2, this becomes $x + 1 + x$, which simplifies to 1. We can construct the following addition and multiplication tables:

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

$\cdot$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	$x+1$	1
$x+1$	0	$x+1$	1	x

Notice that as an additive group, $\mathbb{F}_4$ has the same structure as $\mathbb{Z}_2 \times \mathbb{Z}_2$ or the Klein four-group. As a multiplicative group, the non-zero elements have the same structure as $\mathbb{Z}_3$. This actually generalizes to $\mathbb{F}_{p^k}$ in general. The additive group will be isomorphic to $\mathbb{Z}_p \times \mathbb{Z}_p \times \cdots \times \mathbb{Z}_p$ and the multiplicative group of units will be isomorphic to $\mathbb{Z}_{p^k - 1}$.

Example 5.29. Let's construct a finite field of size 81. Since $81 = 3^4$, we will be working in $\mathbb{Z}_3[x]$, and we want an irreducible polynomial of degree 4. After a little trial and error, one can find the polynomial $p(x) = x^4 + x + 2$. It is irreducible because it has no roots in $\mathbb{Z}_3$ as $p(0) = 2$, $p(1) = 1$, and $p(2) = 2$, and long division shows it is not divisible by any of the three irreducible quadratics of $\mathbb{Z}_3[x]$ (namely $x^2 + 1$, $x^2 + x + 2$, and $x^2 + 2x + 2$).

So our finite field of size 81 is $\mathbb{Z}_3[x]/\langle x^4 + x + 2\rangle$. The elements are all the cosets of the form $(ax^3 + bx^2 + cx + d) + \langle x^4 + x + 2\rangle$, with $a, b, c, d \in \mathbb{Z}_3$. With 3 possibilities each for a, b, c, and d, we see there are 81 elements in total. For simplicity, we will leave off the ideal and just write elements in the form $ax^3 + bx^2 + cx + d$. Addition is done by adding the polynomials. Multiplication is done by multiplying and simplifying via the rule $x^4 + x + 2 = 0$, which we can rewrite as $x^4 = 2x + 1$. For instance, we can multiply x^3 and $2x^2 + 1$ to get $2x^5 + x^3$. We rewrite this as $2(x^4)(x) + x^3$ and plug in our rule to get $2(2x + 1)(x) + x^3$, which simplifies mod 3 to $x^3 + x^2 + 2x$.