Notes on Abstract Algebra

2.1.2 Groups

At this point we've basically beat associativity to death, so let's get on with defining a group in a precise way. As we've mentioned before, the benefit of working in such generality is the fact that we will be able to unify all of our examples under one umbrella. We can prove results about many different examples at once, rather than having to consider many different cases.

Definition 2.1.7. A **group** is a set G together with a binary operation $* : G \times G \to G$ satisfying.

1. **Associativity:** For all $a, b, c \in G$, we have
$$a * (b * c) = (a * b) * c.$$
2. **Identity:** There exists an element $e \in G$ with the property that
$$e * a = a * e = a$$
for all $a \in G$.
3. **Inverses:** For every $a \in G$, there is an element $a^{-1} \in G$ with the property that
$$a * a^{-1} = a^{-1} * a = e.$$

Remark 2.1.8. To be concise, we'll often write $\langle G, * \rangle$ to distinguish the operation on G. If the operation is understood, we'll just write G for the group.

Remark 2.1.9. This seems like a good place to make some remarks about how one should read math. Reading a math book is an active process—you need to stop and think about the material frequently, and it helps to have a pen and paper handy to check things on your own. In particular, when you come across a definition, it is *extremely* helpful to think about two things:

- Immediately try to think of examples that satisfy that definition.
- Think about why the definition is useful. Why are we bothering to make this definition, and why does it say what it says?

We'll address both of these points presently.

Before we start investigating properties of groups, it will be nice to have some examples of groups to fall back on. We've already seen some in our motivating discussion, and we'll add in some others that may also be familiar (or perhaps less familiar).

Example 2.1.10. Here are some examples of groups.

1. $\langle \mathbb{Z}, + \rangle$ is a group, as we have already seen.
2. $\langle \mathrm{M_n}(\mathbb{R}), + \rangle$ is a group.
3. $\langle \mathbb{Z}_n, +_n \rangle$ is a group.

Here are some nonexamples.

4. $\langle \mathbb{Z}, \cdot \rangle$ is *not* a group, since inverses do not always exist. However, $\langle \{1, -1\}, \cdot \rangle$ is a group. We do need to be careful here—the restriction of a binary operation to a smaller set need not be a binary operation, since the set may not be closed under the operation. However, $\{1, -1\}$ is definitely closed under multiplication, so we indeed have a group.
5. $\langle \mathrm{M_n}(\mathbb{R}), \cdot \rangle$ is not a group, since inverses fail. However, $\langle \mathrm{GL_n}(\mathbb{R}), \cdot \rangle$ is a group. We already saw that it is closed, and the other axioms hold as well.
6. $\langle \mathbb{Z}_n, \cdot_n \rangle$ is not a group, again because inverses fail. However, $\langle \mathbb{Z}_n^\times, \cdot_n \rangle$ will be a group. Again, we just need to verify closure: if $a, b \in \mathbb{Z}_n^\times$, then a and b are both relatively prime to n. But then neither a nor b shares any prime divisors with n, so ab is also relatively prime to n. Thus $ab \in \mathbb{Z}_n^\times$.

Here are some other examples that might be less familiar.

Example 2.1.11.

1. Just as $\mathbb{Z}$ is a group under addition, so are

$$\begin{aligned} \mathbb{Q} &= \{\text{rational numbers}\} \\ \mathbb{R} &= \{\text{real numbers}\} \\ \mathbb{C} &= \{\text{complex numbers}\}. \end{aligned}$$

   These are all groups under addition, but not multiplication. Since $\mathbb{Z}$ (and each of these groups) is only a group under addition, we'll usually write $\mathbb{Z}$ for the additive group $\langle \mathbb{Z}, + \rangle$.

2. Let $\mathbb{Q}^\times = \mathbb{Q} - \{0\} = \{a/b \in \mathbb{Q} : a \neq 0\}$, and consider $\langle \mathbb{Q}^\times, \cdot \rangle$. I claim that this is a group. Multiplication is certainly associative, and the identity is 1. Given a rational number $a/b \in \mathbb{Q}^\times$, the inverse is just a/b:

$$\frac{a}{b} \cdot \frac{a}{b} = 1.$$

   Similarly, we could define $\mathbb{R}^\times$ and $\mathbb{C}^\times$, and these would be groups under multiplication (the inverse is simply $a^{-1} = 1/a$).

3. Let $\mathbb{R}^n = \{(a_1, \ldots, a_n) : a_1, \ldots, a_n \in \mathbb{R}\}$ (the addition is done coordinatewise). Then $\langle \mathbb{R}^n, + \rangle$ is a group. Associativity follows from that for addition of real numbers, the identity is the zero vector, and the inverse is just the negative. More generally, any vector space is a group—we simply "forget" about scalar multiplication.

We'll introduce many more examples as we move through the course.

Now that we've done some examples, let's address the second point that we made above. Why does the definition of a group look the way it does? We want to come up with a general sort of object that subsumes the objects we're used to (such as number systems, vector spaces, and matrices), so the "set with a binary operation" part makes sense. We already talked about why associativity is important. But why do we require that there be an identity and inverses? It all comes down to solving equations (after all, that is what "algebra" originally meant). Suppose we are given a group G, and we write down the equation

$$a * x = b,$$

for $a, b, x \in G$. How could we solve for x? Multiply on the left by a^{-1}:

$$\begin{aligned} a^{-1} * (a * x) = a^{-1} * b &\underset{\text{assoc.}}{\Longrightarrow} (a^{-1} * a) * ba^{-1} * b \\ &\underset{\text{inverse}}{\Longrightarrow} e * b = a^{-1} * b \\ &\underset{\text{identity}}{\Longrightarrow} x = a^{-1} * b. \end{aligned}$$

In short, groups are algebraic structures, and it is good to be able to solve equations within them.

We'll often break our study of groups down into different collections of groups that satisfy certain properties. Therefore, let's introduce a couple of adjectives for describing groups. We mentioned before that commutativity is "optional" when it comes to groups. Groups that have commutative operations are special, and therefore they have a special name.

Definition 2.1.12. A group $\langle G, * \rangle$ is said to be **abelian** if $*$ is commutative, i.e.

$$a * b = b * a$$

for all $a, b \in G$. If a group is not commutative, we'll say that it is **nonabelian**.

Abelian groups are named in honor of Niels Henrik Abel, who is mentioned in the historical discussion at the beginning of these notes. Except for $\mathrm{GL_n}(\mathbb{R})$, all the groups we've seen so far are abelian. Soon we'll see two interesting examples of finite nonabelian groups, the symmetric and dihedral groups.

Another useful way to describe groups is via their size. Of course we have a special name for it.

Definition 2.1.13. The **order** of a group G, denoted by $|G|$, is the number of elements in G.

If a group G has infinitely many elements, we will write $|G| = \infty$.[3]

Example 2.1.14. Most of the examples that we have seen so far are infinite groups. In particular, $|Z| = \infty$.

We know lots of examples of infinite groups, but they will actually be hard to understand, with the exception of $\mathbb{Z}$. We'll be more interested in **finite groups**.

Definition 2.1.15. A group G is said to be **finite** if $|G| < \infty$.

One of the reasons that we study finite groups is that it is much easier to analyze their structure, and we'll be able to classify many types of finite groups. Therefore, we'd like to have lots of examples at our disposal. Of course, we've already seen two examples of finite groups, namely the groups that arise from the study of modular arithmetic.

Example 2.1.16.

1. For any n, the additive group $\mathbb{Z}_n$ is a finite group, with $|\mathbb{Z}_n| = n$.
2. Similarly, $\langle \mathbb{Z}_n^\times, \cdot \rangle$ is a finite group. Its order is a little harder to determine. We know that $|\mathbb{Z}_n{}^\times|$ will be the number of $a \in \mathbb{Z}_n$ which are relatively prime to n.

We will produce two more families of finite groups very soon, namely the symmetric and dihedral groups to which we have already alluded.

2.1.3 Group Tables

Now we'll make a slight detour to talk about a tool for working with finite groups. One of the things that makes finite groups easier to handle is that we can write down a table that completely describes the group. We list the elements out and multiply "row by column."

Example 2.1.17. Let's look at $\mathbb{Z}_3$, for example. We'll write down a "multiplication table" that tells us how the group operation works for any pair of elements. As we mentioned, each entry is computed as "row times column":

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

(Of course we have to remember that "times" really means "plus" in this example.) This is called a **group table** (or a **Cayley table**).

Group tables can get pretty tedious when the order of the group is large, but they are handy for analyzing fairly small groups. In particular, let's look at groups of order 1, 2, and 3. If we have a group of order one, there is really only one thing to do. The lone element must be the identity e, and $e * e = e$, so the group table must be:

$*$	e
e	e

Suppose next that we have a group of order two, so that the group contains the identity e and some other element a with $a \neq e$. Then $e * a = a * e = a$, and since a must have an inverse (and the only option is a), we must have $a * a = e$. Therefore, the group table must look like:

$*$	e	a
e	e	a
a	a	e

Finally, suppose we have a group of order three. Then there are three distinct elements: the identity e, and two other elements a and b. The first row and first column are easy to fill in, since e is the identity. We are then left to determine $a*a$, $a*b$, $b*a$, and $b*b$. We claim first that it is impossible to have $a*a = e$. If $a*a = e$, then $b * b = e$ as well, since b must have an inverse. The reason for this is that if $a = b^{-1}$, so $a * b = e$, then

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

If $a = b$, then the group no longer has three elements, so this can't be the case. Therefore, $b * b = e$. But now we must have either $a * b = a$ or $a * b = b$. In the first case,

$$b = e * b = (a * a) * b = a * (a * b) = a * a = e,$$

which can't happen either. Similarly, $a * b \neq b$. The hypothesis that got us into all this trouble was that $a * a = e$, so we can't have $a = a^{-1}$. Also, $a * a \neq a$; if it were, then

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

which is not true. Therefore, we must have $a * a = b$ instead. Since a must have an inverse (and it must be b),

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

Finally, $b * b = a$, since

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

Consequently, the group table must be:

$*$	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

Note that, in particular, there is really only one group of each order (1, 2, or 3), since there was only one way that we could fill in each group table. (In language that we'll develop later, there is only one group of each of these orders *up to isomorphism.*) Also, any group of order 1, 2, or 3 is abelian. To see this, we can simply note that the group table is symmetric about the diagonal. It is an exercise to show that groups of order 4 and 5 are abelian. The result does not extend to order 6—we will see two groups of order 6 which are not abelian. As you can see, group tables can be quite useful, and we will look at them again once we've defined the symmetric and dihedral groups.

Exercise 2.1. Show that any group of order 4 is abelian. [Hint: Compute all possible Cayley tables. Up to a reordering of the elements, there are two possible tables.]

Exercise 2.2. Show that any group of order 5 is abelian. [Hint: There is only one possible Cayley table, up to relabeling.]

2.1.4 Remarks on Notation

Now we'll make one last note on working with general groups. We will get really sick of writing $*$ all the time, so we will oftentimes suppress the operation $*$, and simply write

$$ab = a * b.$$

when working with arbitrary groups. If we are dealing with a specific example (such as the integers under addition), we'll still write the operation to make things

clear. In keeping with this convention, we'll also have a shorthand way of denoting a product of an element a with itself. We'll write

$$a^2 = a * a,$$

and in general

$$a^n = \underbrace{a * a * \cdots * a}_{n \text{ times}}.$$

We can also define negative powers by

$$a^{-n} = \left(a^{-1}\right)^n.$$

These notational conventions will make it much easier to write down computations when we are working with groups. The point of all this is that computations in an arbitrary group will behave very much like usual multiplication, except that things may not always commute. (In fact, matrix multiplication might be the best operation to think of.) Also we should note that if we are dealing with the group $\mathbb{Z}$, a^n really means

$$a^n = \underbrace{a + a + \cdots + a}_{n \text{ times}} = na.$$

This goes to show that one has to be careful when writing down computations in a group, since we usually write things multiplicatively.

2.2 The Symmetric and Dihedral Groups

We are now ready to introduce our first really interesting examples of (finite) nonabelian groups, which are called the **symmetric** and **dihedral groups**. (These are actually *families* of groups, one for each positive integer.) Not only will they give examples of finite nonabelian groups, but we will see that their elements are quite different from the examples that we have been considering so far.

2.2.1 The Symmetric Group

The symmetric group is one of the most important examples of a finite group, and we will spend quite a bit of time investigating its properties. It will arise as a special case of the set S_X of bijections from a set X back to itself, so let's begin there.

Before we can proceed, we need some preliminaries on functions. We'll need some of these facts later on when we discuss homomorphisms, so we will work a little more generally than is absolutely necessary right now.

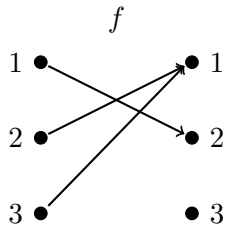
Let X and Y be nonempty sets. Recall that a function $f : X \to Y$ is, at least informally, a rule which assigns to each element $x \in X$ a *unique* element $f(x) \in T$:

$$x \mapsto f(x).$$

We're eventually going to look at a particular class of functions from a set to itself. To do this, we'll need to know what it means for a function $f : X \to Y$ to be **one-to-one** (or **injective**) and **onto** (or **surjective**).

Definition 2.2.1. A function $f : X \to Y$ is **one-to-one** or **injective** if whenever $x_1, x_2 \in X$ with $f(x_1) = f(x_2)$, then $x_1 = x_2$. Equivalently, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$, or different inputs always yield different outputs.

Example 2.2.2. 1. Let $X = \{1, 2, 3\}$ (or any set with three elements), and define $f : X \to X$ by $f(1) = 2$, $f(2) = 1$, and $f(3) = 1$. We can represent this pictorially as



This function is not injective, since 2 and 3 both map to 1.

2. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. This function is *not* injective, since $f(1) = f(-1) = 1$, for example.

3. On the other hand, define $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = x$. This function is injective.

Definition 2.2.3. A function $f : X \to Y$ is **onto** or **surjective** if for any $y \in Y$, there exists an $x \in X$ such that $f(x) = y$.

An intuitive way of describing surjectivity is to say that every element of Y is "hit" by f, or lies in the image of f.

Example 2.2.4. 1. The function $f : \{1, 2, 3\} \to \{1, 2, 3\}$ described in the previous example is not onto, since no element is mapped to 3 by f.

2. The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not onto.

3. If we define $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = x$, then this is a surjective function.

Any function which is both one-to-one and onto is of particular importance, so we have a special name for such objects.

Definition 2.2.5. A function $f : X \to Y$ is **bijective** (or simply a **bijection**) if f is both one-to-one and onto.

The proofs of the following statements would all be good exercises to try on your own, but we will reproduce them here anyway.

Proposition 2.2.6. *Let X, Y, and Z be sets, and let $f : X \to Y$ and $g : Y \to Z$ be functions.*

1. *If f and g are both one-to-one, then so is $g \circ f$.*
2. *If f and g are both onto, then so is $g \circ f$.*
3. *If f and g are both bijections, then so is $g \circ f$.*

Proof. Let's start with the first one. Suppose that f and g are one-to-one, and suppose that $x_1, x_2 \in X$ with $g \circ f(x_1) = g \circ f(x_2)$. This means that

$$g(f(x_1)) = g(f(x_2)),$$

and since g is one-to-one, $f(x_1) = f(x_2)$. But f is also one-to-one, so $x_1 = x_2$. Thus $g \circ f$ is one-to-one.

Now let's handle the question of surjectivity. Suppose f and g are both onto, and let $z \in Z$. We need to produce an $x \in X$ such that $g \circ f(x) = z$, i.e., $g(f(x)) = z$. Since g is onto, there exists a $y \in Y$ such that $g(y) = z$. Also, f is onto, so there exists an $x \in X$ such that $f(x) = y$. Putting this together, we get

$$g \circ f(x) = g(f(x)) = g(y) = z,$$

so $g \circ f$ is onto.

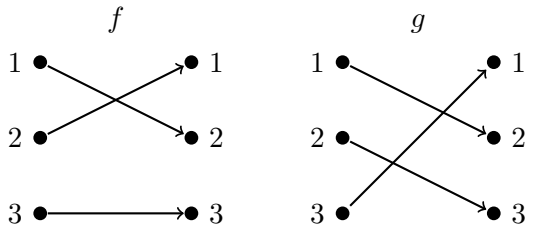
The third statement follows from the first two. If f and g are bijective, then they are both one-to-one and onto. But then $g \circ f$ is one-to-one by (a), and $g \circ f$ is onto by (b), so $g \circ f$ is bijective. □

Now let's formally define S_X, the set of bijections from X to itself. We will then proceed prove that S_X is a group under composition.

Definition 2.2.7. Let X be a set. We define:

$$S_X = \{f : X \to X : f \text{ is a bijection}\}.$$

Note that Proposition 2.2.6(c) tells us that S_X is closed under composition of functions. In other words, composition defines a binary operation on S_X. But what kind of operation is it? Why, it's associative! We already verified this in Example 2.1.6(5). Unfortunately, it is not a commutative operation. To see this, let $X = \{1, 2, 3\}$, and define f and g by the following diagrams:



Then

$$g \circ f(1) = g(f(1)) = g(2) = 3,$$

but

$$f \circ g(1) = f(g(1)) = f(2) = 1,$$

so $g \circ f \neq f \circ g$. Thus S_X will provide a new example of a *nonabelian* group.

To finish checking that S_X is a group, we need to verify the existence of an identity and inverses. For the first one, recall that any set X has a special bijection from X to X, namely the identity function id_X:

$$\mathrm{id}_X(x) = x$$

for all $x \in X$. Note that for any $f \in S_X$, we have

$$f \circ \mathrm{id}_X(x) = f(\mathrm{id}_X(x)) = f(x)$$

and

$$\mathrm{id}_X \circ f(x) = \mathrm{id}_X(f(x)) = f(x)$$

for all $x \in S$. Thus $\mathrm{id}_X \circ f = f \circ \mathrm{id}_X = f$ for all $f \in S_X$, so id_X serves as an identity for S_X under composition. Finally, what do we know about any bijection? Why, it has an inverse, in the sense that there is another function that "undoes" it. More specifically, if $f \in S_X$ and $y \in X$, there is an $x \in X$ such that $f(x) = y$ (since f is onto). But f is also one-to-one, so this x is unique. Therefore, we can define $f^{-1}(y) = x$. You can then check that

$$f \circ f^{-1}(y) = f(f^{-1}(y)) = f(x) = y = \mathrm{id}_X(y)$$

and

$$f^{-1} \circ f(x) = f^{-1}(f(x)) = f^{-1}(y) = x = \mathrm{id}_X(x),$$

so f^{-1} really is an inverse for f under composition. Therefore, by making all of these observations, we have established the following result:

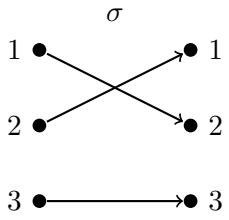
Proposition 2.2.8. *S_X forms a group under composition of functions.*

If X is an infinite set, then S_X is fairly hard to understand. One would have to be either very brave or very crazy to try to work with it. Things are much more tractable (and interesting) when X is finite. Suppose then that X is finite, say with n elements. It doesn't really matter what X is; it only matters that X has n elements. That is, we can label the elements of X to be whatever we want (say, numbers, people, flavors of ice cream, etc.), so we could simply assume that X is the set

$$X = \{1, 2, \ldots, n\}.$$

In this case we call S_X the **symmetric group on** n **letters**, and we denote it by S_n.

An element of S_n is a bijection from $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. In other words, it rearranges the numbers $1, \ldots, n$. Pictorially, we could represent the bijection σ of $\{1, 2, 3\}$ defined by $\sigma(1) = 2$, $\sigma(2) = 1$, and $\sigma(3) = 3$ with the diagram



for representing a particular permutation. It would be annoying if we had to specify $\sigma(1), \sigma(2), \ldots, \sigma(n)$ each time, so we'll often write an element of S_n using **two-line notation**:

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}$$

For example, suppose that $\sigma \in S_3$ is given by the picture that we considered earlier, i.e., $\sigma(1) = 2$, $\sigma(2) = 1$, and $\sigma(3) = 3$. Then we have

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

Of course if we are going to represent permutations in this way, it would help to know how multiplication works in this notation. As an example, let

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

Then remember that multiplication is really just composition of functions:

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

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

On the other hand, what is $\tau\sigma$?

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

In other words, one moves right to left when computing the product of two permutations. First one needs to find the number below 1 in the rightmost permutation, then find this number in the top row of the left permutation, and write down the number directly below it. Repeat this process for the rest of the integers 2 through n.

In the example above, note that $\sigma\tau \neq \tau\sigma$, so what have we shown? We have actually verified that S_3 is nonabelian. This holds more generally:

Proposition 2.2.10. *For $n \geq 3$, S_n is a nonabelian group.*

Proof. Let $\sigma, \tau \in S_3$ be defined as in the example, and suppose that $n > 3$. Define $\tilde{\sigma}, \tilde{\tau} \in S_n$ by

$$\tilde{\sigma}(i) = \begin{cases} \sigma(i) & \text{if } 1 \leq i \leq 3 \\ i & \text{if } i > 3, \end{cases}$$

and similarly for $\tilde{\tau}$. Then the computation that we performed in S_3 shows that $\tilde{\sigma}\tilde{\tau} \neq \tilde{\tau}\tilde{\sigma}$, so S_n is nonabelian. □

It would be entirely feasible to compute the group table for S_n, at least for relatively small n. We'll display the group table for S_3 here, though we won't perform any of the computations explicitly. (There are $6 \cdot 6 = 36$ different products to compute, and these are left to the interested reader.) Label the elements of S_3 as follows:

$$\iota = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \qquad \mu_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

$$\rho_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \qquad \mu_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

$$\rho_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \qquad \mu_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

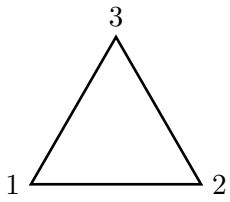
Then the group table for S_3 is:

$*$	ι	ρ_1	ρ_2	μ_1	μ_2	μ_3
ι	ι	ρ_1	ρ_2	μ_1	μ_2	μ_3
ρ_1	ρ_1	ρ_2	ι	μ_3	μ_1	μ_2
ρ_2	ρ_2	ι	ρ_1	μ_2	μ_3	μ_1
μ_1	μ_1	μ_2	μ_3	ι	ρ_1	ρ_2
μ_2	μ_2	μ_3	μ_1	ρ_2	ι	ρ_1
μ_3	μ_3	μ_1	μ_2	ρ_1	ρ_2	ι

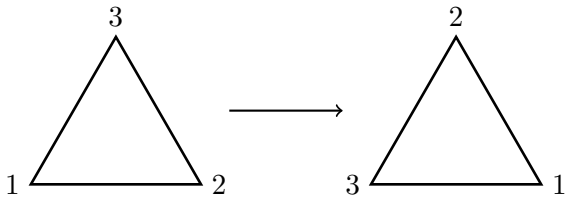
The labeling of the elements may look odd at this point, but we will see a good reason for it quite soon.

2.2.2 The Dihedral Group

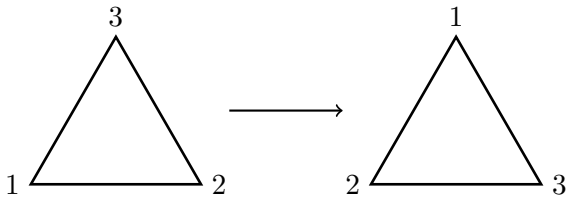
Now it's time to talk about another interesting nonabelian group, the **dihedral group**. Again, we're actually going to be dealing with a family of groups, one for each positive integer. Suppose we have a triangle, and we label the vertices 1, 2, and 3:



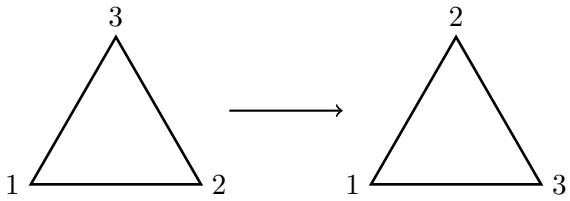
We can rotate the triangle counterclockwise by 120° and obtain a triangle with the same orientation, albeit with the vertices relabeled:



We could also rotate by 240°:



Call these two transformations r_1 and r_2, respectively. We could also reflect the triangle across any of the angle bisectors to obtain a relabeling of the vertices:



Label these reflections m_1, m_2, and m_3. Define the **identity transformation** i to be the one that simply leaves the triangle unmoved. The set

$$\{i, r_1, r_2, m_1, m_2, m_3\}$$

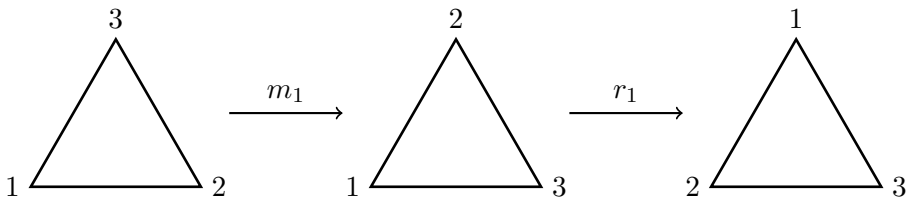
is called the set of **symmetries** of the triangle. They are exactly the transformations of the triangle that reorder the vertices in a way that exploits the symmetry of the equilateral triangle. This set forms a group under composition, called D_3, or the **third dihedral group**. We can define more generally the n**th dihedral group** D_n to be the set of symmetries of a regular n-gon.

Definition 2.2.11. The group D_n is the set of all symmetries of the regular n-gon under composition of transformations.

Naturally, we could ask ourselves what the order of D_n should be. In general, there will be n rotations (including the identity transformation) and n reflections, so

$$|D_n| = 2n.$$

Of course we should talk about how to multiply two symmetries of an n-gon. Again, we need to think of multiplication as composition of functions. For example, suppose we wanted to compute $r_1 m_1$ in D_3. We would first have to apply m_1 to the triangle, and then apply r_1:



We see from this picture that $r_1 m_1 = r_2$. By doing this for all possible pairs of elements, we could write down the group table for D_3:

$*$	i	r_1	r_2	m_1	m_2	m_3
i	i	r_1	r_2	m_1	m_2	m_3
r_1	r_1	r_2	i	m_3	m_1	m_2
r_2	r_2	i	r_1	m_2	m_3	m_1
m_1	m_1	m_2	m_3	i	r_1	r_2
m_2	m_2	m_3	m_1	r_2	i	r_1
m_3	m_3	m_1	m_2	r_1	r_2	i

Observe that this table is the same as that of S_3. (This is why we labeled the elements of S_3 in the way that we did.) This essentially shows that S_3 and D_3 are *isomorphic groups*: they are the same group dressed up in different disguises. In this case, it is not too hard to see how one could identify D_3 with S_3, since the elements of D_3 can be viewed as permutations of the vertices. In general, we can view elements of D_n as permutations of the vertices of the regular n-gon, but we

cannot realize *all* permutations in this way. In other words, we will see that D_n and S_n are not the same. One way to see this at this point is to notice that $|D_n| = 2n$ but $|S_n| = n!$, and these are only equal when $n = 3$.

2.3 Basic Properties of Groups

Now that we've defined groups and we have some interesting examples in our toolkits, it's time to start investigating properties of groups. We'll start of with some simple properties before we really get into the bigger questions regarding the structure and classification of groups. At the very least, this will simplify some of the routine calculations that we need to make when working with groups.

Proposition 2.3.1. *Let G be a group. The identity element $e \in G$ is* unique*, i.e., there is only one element e of G with the property that*

$$ae = ea = a$$

for all $a \in G$.

Proof. For this proof, we need to use the standard mathematical trick for proving uniqueness: we assume that there is another gadget that behaves like the one in which we're interested, and we prove that the two actually have to be the same. Suppose there is another $f \in G$ with the property that

$$af = fa = a$$

for all $a \in G$. Then in particular,

$$ef = fe = e.$$

But since e is an identity,

$$ef = fe = f.$$

Therefore,

$$e = ef = f,$$

so e is unique. □

The next result has to do with *solving equations*, which was our original motivation for requiring that inverses exist. It's called **cancellation**.

Proposition 2.3.2 (Cancellation laws)**.** *Let G be a group, and let $a, b, c \in G$. Then:*

(a) *If $ab = bc$, then $b = c$.*

(b) *If $ba = ca$, then $b = c$.*

Proof. (a) Suppose that $ab = ac$. Multiply both sides on the left by a^{-1}:

$$a^{-1}(ab) = a^{-1}(ac).$$

By associativity, this is the same as

$$(a^{-1}a)b = (a^{-1}a)c,$$

and since $a^{-1}a = e$, we have

$$eb = ec.$$

Since e is the identity, $b = c$. The same sort of argument works for (b), except we multiply the equation on the right by a^{-1}. □

The cancellation laws actually give us a very useful corollary. You may have already guessed that this result holds, but we will prove here that inverses in a group are unique.

Corollary 2.3.3. *Let G be a group. Every $a \in G$ has a* unique *inverse, i.e. to each $a \in G$ there is exactly one element a^{-1} with the property that*

$$aa^{-1} = a^{-1}a = e.$$

Proof. Let $a \in G$, and suppose that $b \in G$ has the property that $ab = ba = e$. Then in particular,

$$ab = e = aa^{-1},$$

and by cancellation, $b = a^{-1}$. Thus a^{-1} is unique. □

While we are on the topic of inverses, we will prove two more results about them. The first tells us what happens when we "take the inverse twice," and the second gives us a rule for determining the inverse of a product of two group elements.

Proposition 2.3.4. *If $a \in G$, then $(a^{-1})^{-1} = a$.*

Proof. By definition, $a^{-1}(a^{-1})^{-1} = e$. But $a^{-1}a = aa^{-1} = e$ as well, so by uniqueness of inverses, $(a^{-1})^{-1} = a$. □

Recall from linear algebra that if $A, B \in \mathrm{GL_n}(\mathbb{R})$, we have a formula for $(AB)^{-1}$; it is simply $B^{-1}A^{-1}$. In general there is a rule for determining the inverse of a product of two group elements, which looks just like the rule for matrices—it is the product of the inverses, but in the reverse order.

Proposition 2.3.5. *For any $a, b \in G$, $(ab)^{-1} = b^{-1}a^{-1}$.*

Proof. We'll explicitly show that $b^{-1}a^{-1}$ is the inverse of ab by computing:

$$\begin{aligned}(ab)(b^{-1}a^{-1}) &= ((ab)b^{-1})a^{-1} \\ &= (a(bb^{-1}))a^{-1} \\ &= (ae)a^{-1} \\ &= aa^{-1} \\ &= e.\end{aligned}$$

Of course we also need to check that $(b^{-1}a^{-1})(ab) = e$, which works pretty much the same way:

$$\begin{aligned}(b^{-1}a^{-1})(ab) &= b^{-1}(a^{-1}(ab)) \\ &= b^{-1}((a^{-1}a)b) \\ &= b^{-1}(eb) \\ &= b^{-1}b \\ &= e.\end{aligned}$$

Thus $(ab)^{-1} = b^{-1}a^{-1}$. □

In order to be thorough and rigorous, we needed to check that $b^{-1}a^{-1}$ was a *two-sided* inverse in the previous proof. It would be quite annoying if we had to do this all the time, and you may be wondering if there is a shortcut. There is one, which allows us to check that two group elements are inverses of each other simply by multiplying them in only one of the two possible orders. To prove it, we'll use the following proposition.

Proposition 2.3.6. *The equations $ax = b$ and $xa = b$ have unique solutions in G.*

Proof. The solution to $ax = b$ is $x = a^{-1}b$, and for $xa = b$ it is $x = ba^{-1}$. These are unique since inverses are unique. □

Proposition 2.3.7. *Let G be a group, and let $a, b \in G$. If either $ab = e$ or $ba = e$, then $b = a^{-1}$.*

Proof. This really amounts to solving the equation $ax = e$ (or $xa = e$). We know from Proposition 2.3.6 that there is a unique solution, namely $x = a^{-1}e = a^{-1}$ (in either case). Therefore, if $ab = e$ or $ba = e$, then b is a solution to either $ax = e$ or $xa = e$, so $b = a^{-1}$. □

Exercise 2.3. Prove that if G is a group and $a, b \in G$ with $ab = a$, then $b = e$.

2.4 The Order of an Element and Cyclic Groups

Recall that we've developed some shorthand notation for writing out computations in groups. For one, we have been suppressing the binary operation $*$, and we simply write ab in place of $a*b$. We also developed some notation for powers of an element. We write

$$a^n = \underbrace{a * a * \cdots * a}_{n \text{ times}}$$

for $n \in \mathbb{Z}^+$, and

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

In addition, we define $a^0 = e$. Note that this is written for arbitrary groups—if we're dealing with an *additive* group (like $\mathbb{Z}$ or $\mathbb{Z}_n$), then we would really write a^n as

$$\underbrace{a + \cdots + a}_{n \text{ times}} = na.$$

We should always keep this convention in mind when we are working with specific examples.

Now let G be a group, and let $a \in G$. We're going to investigate the things that we can do with powers of a. Therefore, we'll begin by giving a name to the set of all powers of a.

Definition 2.4.1. Given a group G and an element $a \in G$, we define

$$\langle a \rangle = \{a^n : n \in \mathbb{Z}\} = \{\ldots, a^{-2}, a^{-1}, e, a, a^2, a^3, \ldots\}.$$

The first observation that we will make is that the familiar "exponent rules" hold for an arbitrary group element.

Proposition 2.4.2. *Let G be a group, and let $a \in G$.*

(a) $a^n a^m = a^{n+m}$ *for all* $n, m \in \mathbb{Z}$.

(b) $(a^n)^{-1} = a^{-n}$ *for all* $n \in \mathbb{Z}$.

(c) $(a^n)^m = a^{nm}$ *for all* $n, m \in \mathbb{Z}$.

Exercise 2.4. Prove Proposition 2.4.2.

Now we'll investigate what happens when we keep taking powers of an element. Let's use a specific example to get started.

Example 2.4.3. Let $G = \mathbb{Z}_{12}$. Let's compute the "powers" of some elements. (Recall that in this group "power" really means "multiple.") We'll calculate the powers of 2 first:

$$\begin{aligned}
1 \cdot 2 &= 2 \\
2 \cdot 2 &= 2 +_{12} 2 = 4 \\
3 \cdot 2 &= 2 +_{12} 2 +_{12} 2 = 6 \\
4 \cdot 2 &= 8 \\
5 \cdot 2 &= 10 \\
6 \cdot 2 &= [12]_{12} = 0 \\
7 \cdot 2 &= [14]_{12} = 2 \\
8 \cdot 2 &= [16]_{12} = 4
\end{aligned}$$

and so on. What about powers of 3?

$$\begin{aligned}
1 \cdot 3 &= 3 \\
2 \cdot 3 &= 3 +_{12} 3 = 6 \\
3 \cdot 3 &= 3 +_{12} 3 +_{12} 3 = 9 \\
4 \cdot 3 &= [12]_{12} = 0 \\
5 \cdot 3 &= [15]_{12} = 3 \\
6 \cdot 3 &= [18]_{12} = 6
\end{aligned}$$

and so on. Notice that the lists repeat after a while. In particular, we reach 0 (i.e., the identity) after a certain point. We quantify this phenomenon by saying that these elements have **finite order**.

Definition 2.4.4. Let G be a group. We say that an element $a \in G$ has **finite order** if there exists $n \in \mathbb{Z}^+$ such that $a^n = e$. The smallest such integer is called the **order** of a, denoted by $o(a)$ (or $|a|$). If no such integer exists, we say that a has **infinite order**.

Example 2.4.5.

1. The identity element in any group has order 1.
2. In $\mathbb{Z}_{12}$, we have seen that $o(2) = 6$ and $o(3) = 4$.
3. In D_3, the order of the reflection m_1 is 2. (This is true of all reflections in D_3.) Also, $o(r_1) = o(r_2) = 3$.
4. In $\mathbb{Z}$, 5 has infinite order, as do all other elements.

Let's ask ourselves a question. Do all the elements of $\mathbb{Z}_{12}$ have finite order? Yes—if $a \in \mathbb{Z}_{12}$, then $12 \cdot a = 0$, for example. What is the real reason for this? The group is finite, so there are only so many places to put the powers of an element. We can quantify this with the following proposition.

Proposition 2.4.6. *Let G be a finite group. Then every element $a \in G$ has finite order.*

Proof. Consider the set

$$\{a^n : n \geq 0\} = \{e, a, a^2, \ldots\}.$$

Since G is finite, this list of powers can't be infinite. (This follows from the Pigeonhole principle, for instance. We have an infinite list of group elements that need to fit into only finitely many slots.) Therefore, two different powers of a must coincide, say $a^i = a^j$, with $j \neq i$. We can assume that $j > i$. Then

$$a^{j-i} = a^j a^{-i} = a^i a^{-i} = e,$$

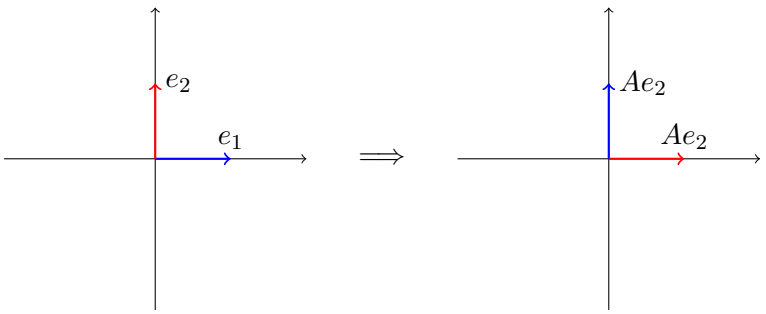
so a has finite order. (In particular, $o(a) \leq j - i$.) Since $a \in G$ was arbitrary, the result follows. □

Remark 2.4.7. There are two questions that we could ask here. First, you may be wondering if the converse of Proposition 2.4.6 holds. That is, if G is a group in which every element has finite order, is G necessarily finite? The answer to this question is no—there are examples of infinite groups in which every element has finite order, but we do not have the tools yet to describe them.

On the other hand, if G is an infinite group, can we have elements of finite order? Yes—there are even examples in $\mathrm{GL_n}(\mathbb{R})$:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

so this matrix has order 2. It's also easy to see this if we think about the linear transformation that this matrix represents—it reflects the plane $\mathbb{R}^2$ across the line $y = x$. That is, it interchanges the two standard basis vectors:



Let's get on with proving some facts about order. First, we'll relate the order of an element to that of its inverse.

Proposition 2.4.8. *Let G be a group and let $a \in G$. Then $o(a) = o(a^{-1})$.*

Proof. Suppose first that a has finite order, with $o(a) = n$. Then

$$(a^{-1})^n = a^{-n} = (a^n)^{-1} = e^{-1} = e,$$

so $o(a^{-1}) \leq n = o(a)$. On the other hand, if we let $m = o(a^{-1})$, then

$$a^m = ((a^{-1})^{-1})^m = (a^{-1})^{-m} = ((a^{-1})^m)^{-1} = e,$$

so $n \leq m$. Thus $n = m$, or $o(a) = o(a^{-1})$.

Now suppose that a has infinite order. Then for all $n \in \mathbb{Z}^+$, we have $a^n \neq e$. But then

$$(a^{-1})^n = a^{-n} = (a^n)^{-1} \neq e$$

for all $n \in \mathbb{Z}^+$, so a^{-1} must have infinite order as well. □

Let's continue with our investigation of basic properties of order. The first one says that the only integers m for which $a^m = e$ are the multiples of $o(a)$.

Proposition 2.4.9. *If $o(a) = n$ and $m \in \mathbb{Z}$, then $a^m = e$ if and only if n divides m.*

Proof. If $n \mid m$, it is easy. Write $m = nd$ for some $d \in \mathbb{Z}$. Then

$$a^m = a^{nd} = (a^n)^d = e^d = e.$$

On the other hand, if $m \geq n$, we can use the Division Algorithm to write $m = qn + r$ with $0 \leq r < n$. Then

$$e = a^m = a^{qn+r} = a^{qn}a^r = (a^n)^q a^r = ea^r = a^r,$$

so $a^r = e$. But $r < n$, and n is the smallest positive power of a which yields the identity. Therefore r must be 0, and n divides m. □

Note that this tells us something more general about powers of a: when we proved that elements of finite groups have finite order, we saw that $a^i = a^j$ implied that $a^{j-i} = e$. This means that $n = o(a)$ divides $j - i$. In other words, i and j must be congruent mod n.

Proposition 2.4.10. *Let G be a group, $a \in G$ an element of finite order n. Then $a^i = a^j$ if and only if $i \equiv j (\mathrm{mod} n)$.*

Along the same lines, we observed that if $a^i = a^j$ with $j > i$, then $a^{j-i} = e$, so a has to have finite order. Taking the contrapositive of this statement, we get the following result.

Proposition 2.4.11. *Let G be a group, $a \in G$ an element of infinite order. Then all the powers of of a are all distinct. (That is, $a^i = a^j$ if and only if $i = j$.)*

Finally, If we know that a has order n, then how can we find the orders of other powers of a? It turns out that there is a nice formula.

Theorem 2.4.12. *Let G be a group, $a \in G$ an element of finite order, say $o(a) = n$, and let $m \in \mathbb{Z}$. Then*

$$o(a^m) = \frac{n}{\gcd(m, n)}.$$

Proof. Let $d = \gcd(m, n)$. Then

$$(a^m)^{n/d} = a^{mn/d} = (a^n)^{m/d} = e^{m/d} = e.$$

However, we need to check that n/d is the smallest positive integer which gives the identity. Suppose that $k \in \mathbb{Z}^+$ and $(a^m)^k = e$. Then

$$a^{mk} = e,$$

so $n \mid mk$ by the previous proposition. It follows that n/d divides $(m/d)k$. Now we observe that $\gcd(n/d, m/d) = 1$: we use Bézout's lemma to write

$$d = nx + my$$

for some $x, y \in \mathbb{Z}$. Dividing by d, we get

$$1 = (n/d)x + (m/d)y,$$

so n/d and m/d have to be relatively prime. (Any common divisor would have to divide this linear combination , hence divide 1.) This means that n/d divides k. In particular, $n/d \leq k$, so $o(a^m) = n/d$. □

2.4.1 Cyclic Groups

Let's take a few steps back now and look at the bigger picture. That is, we want to investigate the structure of the set $\langle a \rangle$ for $a \in G$. What do you notice about it?

- **Closure:** $a^i a^j = a^{i+j} \in \langle a \rangle$ for all $i, j \in \mathbb{Z}$.
- **Identity:** $e = a^0 \in \langle a \rangle$
- **Inverses:** Since $(a^j)^{-1} = a^{-j}$, we have $(a^j)^{-1} \in \langle a \rangle$ for all $j \in \mathbb{Z}$.

It other words, $\langle a \rangle$ is itself a group. That is, the set of all powers of a group element is a group in its own right. We will investigate these sorts of objects further in the next section, but let's make the following definition now anyway.

Definition 2.4.13. For $a \in G$, the set $\langle a \rangle$ is called the **cyclic subgroup** generated by a.

For now, let's look at a particular situation. Is G ever a cyclic subgroup of itself? That is, can a "generate" the whole group G? Yes, this does happen sometimes, and such groups are quite special.

Definition 2.4.14. A group G is called **cyclic** if $G = \langle a \rangle$ for some $a \in G$. The element a is called a **generator** for G.

Cyclic groups are extremely well-understood. We'll see that they have very nice properties, and we can completely classify them almost immediately. First, let's do some examples, many of which we have already seen.

Example 2.4.15.

1. One of our first examples of a group is actually a cyclic one: $\mathbb{Z}$ forms a cyclic group under addition. What is a generator for $\mathbb{Z}$? Both 1 and -1 generate it, since every integer $n \in \mathbb{Z}$ can be written as a "power" of 1 (or -1):
$$n = n \cdot 1 = \underbrace{1 + 1 + \cdots + 1}_{n \text{ times}}.$$
Thus,
$$\mathbb{Z} = \langle 1 \rangle = \langle -1 \rangle.$$
These are actually the only two generators.

2. How about a finite cyclic group? For any n, $\mathbb{Z}_n$ is cyclic, and 1 is a generator in much the same way that 1 generates $\mathbb{Z}$. There are actually plenty of other generators, and we can characterize them by using our knowledge of greatest common divisors. We'll postpone this until we've made a couple of statements regarding cyclic groups.

3. The group $\langle \mathbb{Q}, + \rangle$ is not cyclic. (This is proven in Saracino.)

4. The dihedral group D_3 is not cyclic. The rotations all have order 3, so
$$\langle r_1 \rangle = \langle r_2 \rangle = \{i, r_1, r_2\}.$$
On the other hand, all of the reflections have order 2, so
$$\langle m_1 \rangle = \{i, m_1\}, \quad \langle m_2 \rangle = \{i, m_2\}, \quad \langle m_3 \rangle = \{i, m_3\}.$$

Now let's start making some observations regarding cyclic groups. First, if $G = \langle a \rangle$ is cyclic, how big is it? It turns out that our overloading of the word "order" was fairly appropriate after all, for $|G| = o(a)$.

Theorem 2.4.16. *If $G = \langle a \rangle$ is cyclic, then $|G| = o(a)$.*

Proof. If a has infinite order, then $|G|$ must be infinite. On the other hand, if $o(a) = n$, then we know that $a^i = a^j$ if and only if $i \equiv j \mod n$, so the elements of G are

$$e, a, a^2, \ldots, a^{n-1},$$

of which there are $n = o(a)$. □

If we pair this result with Theorem 2.4.12, we can characterize the generators of any finite cyclic group.

Proposition 2.4.17. *The generators of a finite cyclic group $G = \langle a \rangle$ of order n are precisely the elements a^r for which $\gcd(r, n) = 1$.*

Proof. By the theorem,

$$o(a^r) = \frac{n}{\gcd(r, n)}.$$

Note that a^r generates G if and only if $o(a^r) = n$. If a^r has order n, then $|\langle a^r \rangle| = n$ and $\langle a^r \rangle \subseteq \langle a \rangle$. Since both sets have the same (finite) number of elements, they must be the same. On the other hand, if $\langle a^r \rangle = \langle a \rangle$, then

$$o(a^r) = |\langle a^r \rangle| = |\langle a \rangle| = n.$$

Now note that $o(a^r) = n$ precisely when $\gcd(r, n) = 1$. □

Remark 2.4.18. This corollary allows us to characterize the generators of $\mathbb{Z}_n$. Recall that $\mathbb{Z}_n = \langle 1 \rangle$. Remember that we have to be careful here—when we write something like a^m in an arbitrary group, we mean "multiply a by itself m times," where "multiply" should be interpreted as whatever the group operation happens to be. In particular, the operation in $\mathbb{Z}_n$ is addition, so a^m really means

$$\underbrace{a + a + \cdots + a}_{m \text{ times}} = m \cdot a.$$

Therefore, the generators of $\mathbb{Z}_n$ are precisely the elements of the form $r \cdot 1$, where $\gcd(r, n) = 1$. In other words, an element $a \in \mathbb{Z}_n$ is a generator if and only if $\gcd(a, n) = 1$.

On a related note, let $G = \langle a \rangle$ be a finite cyclic group with $n = |G| = o(a)$. If we take any element $b \in G$ and compute b^n, what do we get? Well, $b = a^i$ for some $i \in \mathbb{Z}$, so

$$b^n = (a^i)^n = a^{in} = (a^n)^i = (a^{o(a)})^i = e^i = e.$$

Therefore:

Theorem 2.4.19. *Let G be a finite cyclic group. Then for any element $b \in G$, we have $b^{|G|} = e$.*

If we combine this result with Proposition 2.4.9, If $b^{|G|} = e$, then what can we say about $o(b)$ in relation to $|G|$? We must have that $o(b)$ divides $|G|$.

Proposition 2.4.20. *Let G be a finite cyclic group. Then for any $b \in G$, we have $o(b) \mid |G|$.*

These last two results are not as serendipitous as they may seem at first glance. These phenomena for cyclic groups actually hold for *any* finite group. Once we have established Lagrange's theorem, we'll see why this is true.

Finally, the astute reader will notice a key similarity between all of the examples of cyclic groups that we've listed above: they are all *abelian*. It is not hard to see that this is always the case.

Theorem 2.4.21. *Every cyclic group is abelian.*

Proof. Let G be a cyclic group and let a be a generator for G, i.e. $G = \langle a \rangle$. Then given two elements $x, y \in G$, we must have $x = a^i$ and $y = a^j$ for some $i, j \in \mathbb{Z}$. Then

$$xy = a^i a^j = a^{i+j} = a^{j+i} = a^j a^i = yx,$$

and it follows that G is abelian. □

Remark 2.4.22. The converse to Theorem 2.4.21 is not true. That is, there are abelian groups that are not cyclic. Saracino gives the example of the non-cyclic group $\langle \mathbb{Q}, + \rangle$. However, this is a good place to introduce a different group—the **Klein 4-group**, denoted V_4.[7] The Klein 4-group is an abelian group of order 4. It has elements $V_4 = \{e, a, b, c\}$, with

$$a^2 = b^2 = c^2 = e$$

and

$$ab = c, bc = a, ca = b.$$

Note that it is abelian by a previous exercise (Exercise 2.1). However, it is not cyclic, since every element has order 2 (except for the identity, of course). If it were cyclic, there would necessarily be an element of order 4.

2.4.2 Classification of Cyclic Groups

As one final word on cyclic groups, we should mention that these groups are very easy to classify. It will turn out that they have a very rigid structure, and we have already seen some examples of this. Specifically, we will see that there are really

[7]

only two flavors of cyclic group: finite and infinite. Let's think about what happens in the infinite case first. If we have an infinite cyclic group $G = \langle a \rangle$, then a has infinite order, and the group elements are simply

$$G = \left\{ a^j : j \in \mathbb{Z} \right\}.$$

How does the group operation work? We have

$$a^j a^k = a^{j+k},$$

so we can multiply two elements by simply adding the exponents as integers. Therefore, we'll see that this will allow us to set up an isomorphism between any infinite cyclic group G and $\mathbb{Z}$. That is, there is only one infinite cyclic group, namely $\mathbb{Z}$.

Things are almost as easy in the finite case as well. Suppose that $G = \langle a \rangle$ is a finite cyclic group of order n. Then to add two elements, we have

$$a^j a^k = a^{j+k}.$$

However, we have to remember that in doing this, $j + k$ may have exceeded the order of the group, and we can reduce it. That is, we can write

$$j + k = nq + r,$$

so

$$a^j a^k = a^{j+k} = a^{nq+r} = a^{nq} a^r = e a^r = a^r.$$

Therefore, multiplication corresponds to addition of the exponents modulo n. We'll see that this allows us to identify G with the elementary cyclic group $\mathbb{Z}_n$. In other words, there is only one finite cyclic group of each order $n \in \mathbb{Z}^+$.

2.5 Subgroups

Recall that we saw in the last section that if G is a group and $a \in G$, then the set $\langle a \rangle$ is itself a group which sits inside G. That is, we have a subset of G which is itself a group with respect to the same operation as G. A more general version of this situation will be of great interest to us. Let's first take a moment to say why this is so.

We now want to begin studying groups from a much broader perspective. So far, we have only studied groups as standalone objects, without any relations between them. This viewpoint will begin to change now. We will start to consider more interesting questions, like how groups are connected to each other via functions—the concept of homomorphism and isomorphism—and how they are built out of smaller groups. This leads naturally to the questions of *classification* and *structure*:

- **Classification:** When are two groups the same or different? If we are dealing with a mysterious new example of a group, perhaps it is just a more familiar group in disguise.

- **Structure:** How is a group built out of smaller pieces? When working with a big, complicated group, we might be able to break the group up into smaller pieces, which are then easier to analyze.

We've already seen some primitive examples that foreshadow the classification problem. At this point we will start laying the foundation for the structure problem. Eventually we'll want to tear down groups into smaller bits, which will make the larger group easier to understand. To do this, we need to know what these "bits" really are.

Obviously we want to look at subsets of a group G, but we don't just want any old subset of G. We would like to consider only subsets that encode information about the group structure on G. These subsets are exactly what are called *subgroups* of G.

Definition 2.5.1. Let $\langle G, * \rangle$ be a group. A **subgroup** of G is a nonempty subset $H \subseteq G$ with the property that $\langle H, * \rangle$ is a group.

Note that in order for H to be a subgroup of G, H needs to be a group with respect to the operation that it inherits from G. That is, H and G *always* carry the same binary operation. Also, we'll write

$$H \leq G$$

to denote that H is a subgroup of G. Finally, if we want to emphasize that $H \leq G$ but $H \neq G$, we will say that H is a **proper** subgroup of G.

In order to think about how to show that a subset of a group is actually a subgroup, let's work with an example. This will really be a specific example of the "cyclic subgroups" that we introduced earlier, but let's rehash the details here anyway.

Example 2.5.2. Let's look at the group $\mathbb{Z}$ (under addition, of course). Define

$$2\mathbb{Z} = \{\text{even integers}\} = \{2n : n \in \mathbb{Z}\}.$$

Is $2\mathbb{Z}$ a subgroup of $\mathbb{Z}$? We need to check that $2\mathbb{Z}$ itself forms a group under addition:

- **Closure:** If $a, b \in 2\mathbb{Z}$, then $a = 2n$ and $b = 2m$ for some $n, m \in \mathbb{Z}$. Then

  $$a + b = 2n + 2m = 2(n + m) \in 2\mathbb{Z},$$

  so $2\mathbb{Z}$ is indeed closed under addition.

- **Associativity:** Is there even anything to check here? No—the operation on $\mathbb{Z}$ is already associative, so nothing changes when we pass to a subset of $\mathbb{Z}$.
- **Identity:** The identity for addition on $\mathbb{Z}$ is 0, which is even: $0 = 2 \cdot 0 \in 2\mathbb{Z}$.
- **Inverses:** If $a \in 2\mathbb{Z}$, then $a = 2n$ for some $n \in \mathbb{Z}$, and $-a = -2n = 2(-n) \in 2\mathbb{Z}$.

Therefore, $\langle 2\mathbb{Z}, + \rangle$ is a group, hence a subgroup of $\mathbb{Z}$.

In general, we will want to be able to check whether a subset of a group is actually a subgroup. Fortunately, this example tells us exactly how to do it.

To check that $H \leq G$, one needs to verify the following:

1. H is **closed** under the operation on G.
2. The **identity** element $e \in G$ is in H.
3. For every $a \in H$, its **inverse** a^{-1} is in H.

Example 2.5.3. Naturally, it would be helpful to look at some examples of subgroups.

1. Every group G has two special subgroups, namely

   $$\{e\} \text{ and } G.$$

   These are called the **trivial subgroups** of G.[9]

2. We saw earlier that $2\mathbb{Z}$ is a subgroup of $\mathbb{Z}$. There is nothing special about 2 in this example: for any $n \in \mathbb{Z}^+$,

   $$n\mathbb{Z} = \{na : a \in \mathbb{Z}\}$$

   is a subgroup of $\mathbb{Z}$. The exact same computations that we performed for $2\mathbb{Z}$ will show that $n\mathbb{Z} \leq \mathbb{Z}$.

3. The rational numbers $\mathbb{Q}$ form an additive subgroup of $\mathbb{R}$.

4. Here is an example from linear algebra. Consider the n-dimensional vector space $\mathbb{R}^n$. Then $\mathbb{R}^n$ is, in particular, an abelian group under addition, and any

[9] Many books will reserve the phrase "trivial subroup" only for the identity subgroup $\{e\}$. The group G is sometimes referred to as the **improper subgroup**.

vector subspace of $\mathbb{R}^n$ is a subgroup of $\mathbb{R}^n$.[10] If H is a subspace of $\mathbb{R}^n$, then it is closed under addition, and closure under scalar multiplication guarantees that $0 \in H$ and for $v \in H$, $-v = -1 \cdot v \in H$.

5. Let D_n be the nth dihedral group, and let

$$H = \{i, r_1, r_2, \ldots, r_{n-1}\}$$

be the set of rotations in D_n. Then $H \leq D_n$. It is closed, since the composition of two rotations is another rotation, the identity $i \in H$, and for any $1 \leq j \leq n-1$,

$$r_j^{-1} = r_{n-j},$$

which is again in H.

6. Let $\mathrm{GL_n}(\mathbb{R})$ be our usual group of invertible $n \times n$ matrices under matrix multiplication, and define

$$\mathrm{SL_n}(\mathbb{R}) = \{A \in \mathrm{GL_n}(\mathbb{R}) : \det(A) = 1\}.$$

Then $\mathrm{SL_n}(\mathbb{R}) \leq \mathrm{GL_n}(\mathbb{R})$. To see that it is closed, we recall that $\det(AB) = \det(A)\det(B)$, so if $A, B \in \mathrm{SL_n}(\mathbb{R})$,

$$\det(AB) = \det(A)\det(B) = 1 \cdot 1 = 1,$$

and $AB \in \mathrm{SL_n}(\mathbb{R})$. The identity matrix surely has determinant 1, and if $A \in \mathrm{SL_n}(\mathbb{R})$, then

$$\det(A^{-1}) = \frac{1}{\det(A)} = \frac{1}{1} = 1,$$

so $A^{-1} \in \mathrm{SL_n}(\mathbb{R})$. Therefore, $\mathrm{SL_n}(\mathbb{R})$ is a subgroup of $\mathrm{GL_n}(\mathbb{R})$, called the **special linear group**.[11]

7. Here's our last example, and a more interesting one at that. Exercise 2.1 leads to toward the observation that there are two groups of order 4, and both are abelian. One of them is of course $\mathbb{Z}_4$, and the other is the Klein 4-group V_4,

[10] A word of caution about this statement: any subspace of $\mathbb{R}^n$ is necessarily a subgroup, but there are plenty of subgroups that are not vector subspaces. For example, the *rational* vector space $\mathbb{Q}^n$ is an additive subgroup of $\mathbb{R}^n$, but it is not a *real* subspace. It is a $\mathbb{Q}$-subspace, however, which goes to show that the field of scalars is critical when talking about subspaces of vector spaces. Even worse, the set

$$\mathbb{Z}^n = \{(a_1, a_2, \ldots, a_n) : a_i \in \mathbb{Z}\}$$

is a subgroup of $\mathbb{R}^n$, called a **lattice**, but it is not a vector space of any kind. (If you're feeling extra brave/curious, it is an example of a more general object, called a **module**.)

[11] You might wonder what makes the special linear group "special." If we view the matrices in $\mathrm{GL_n}(\mathbb{R})$ as invertible linear transformations from $\mathbb{R}^n$ to itself, then the elements of $\mathrm{SL_n}(\mathbb{R})$ are precisely the transformations which preserve volume and orientation.

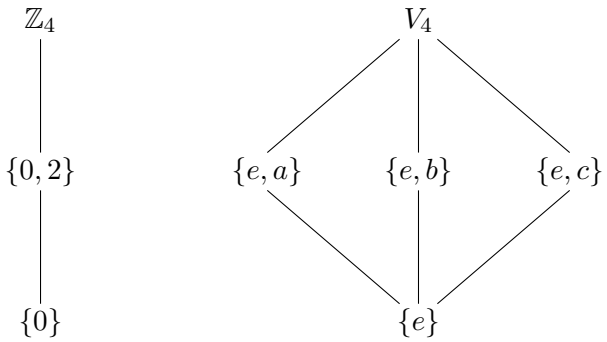
which was introduced in Remark 2.4.22. You saw in that exercise that $\mathbb{Z}_4$ and V_4 really are different groups, since they have completely different Cayley tables. We will now observe this in a different way, by checking that they have different subgroup structures. First, we claim that the only nontrivial subgroup of $\mathbb{Z}_4$ is $H = \{0, 2\}$. It's easy to check that this is a subgroup, but why is it the only one? We will prove a result to this effect quite soon, or we could observe that the other possible proper subgroups are

$$\{0, 1\}, \{0, 3\}, \{0, 1, 2\}, \{0, 1, 3\}, \{0, 2, 3\},$$

and it is easy to check that none of these are closed under addition. On the other hand, V_4 has three subgroups (in addition to $\{e\}$ and V_4 itself):

$$\{e, a\}, \{e, b\}, \{e, c\}.$$

We therefore know all the subgroups of these two groups, and we can represent this pictorially with something called a **subgroup lattice** diagram:



As we have already insinuated, we will sometimes be able to tell when two groups are different by studying these sorts of lattice diagrams. For example, $\mathbb{Z}_4$ has only one subgroup of order 2, while V_4 has three such subgroups. This is one indication that these are indeed different groups. If you go on to study Galois theory at all, you will see that subgroup diagrams are quite important in that context.

2.5.1 Cyclic Subgroups

Some of the examples that we have mentioned are actually cases of very special kinds of subgroups, called **cyclic subgroups**. We already introduced cyclic subgroups en route to our study of cyclic groups. To recap, let G be a group and let $a \in G$. We defined the set

$$\langle a \rangle = \left\{ a^j : j \in \mathbb{Z} \right\},$$

and we observed that $\langle a \rangle$ is itself a group which sits inside G. That is, every element of a group G generates a whole subgroup of G, to which we attach a special name.

Definition 2.5.4. The group $\langle a \rangle$ is called the **cyclic subgroup** generated by a.

When we say that a "generates" $\langle a \rangle$, we mean that that $\langle a \rangle$ is created entirely out of the element a. In a certain sense, $\langle a \rangle$ is the *smallest* possible subgroup of G which contains a. Let's try to make this more precise. If $H \leq G$ and $a \in H$, then H must contain the elements

$$a, a^2, a^3, \ldots,$$

since H is closed. It also must contain e and a^{-1}, hence all of the elements

$$\ldots, a^{-2}, a^{-1}, e, a, a^2, \ldots,$$

i.e. all powers of a. That is, $\langle a \rangle \subset H$, and we have proven the following fact:

Theorem 2.5.5. *Let G be a group and let $a \in G$. Then $\langle a \rangle$ is the smallest subgroup of G containing a, in the sense that if $H \leq G$ and $a \in H$, then $\langle a \rangle \subseteq H$.*

Of course we've already encountered several examples of cyclic subgroups in our studies thus far.

Example 2.5.6.

1. Our first example of a subgroup, $2\mathbb{Z} \leq \mathbb{Z}$, is a cyclic subgroup, namely $\langle 2 \rangle$. Similarly, $n\mathbb{Z}$ is cyclic for any $n \in \mathbb{Z}$.
2. The subgroup consisting of rotations on D_n,
$$H = \{i, r_1, r_2, \ldots, r_{n-1}\} \leq D_n,$$
is cyclic since $H = \langle r_1 \rangle$.
3. All the proper subgroups of $\mathbb{Z}_4$ and V_4 that we listed are cyclic. In addition, $\mathbb{Z}_4$ is a cyclic subgroup of itself, but V_4 is not.
4. The trivial subgroup $\{e\}$ is always a cyclic subgroup, namely $\langle e \rangle$.

Cyclic subgroups are useful because they will allow us to gather information about elements of G by studying the cyclic subgroups that they generate. Properties of the subgroup are reflected in those of the element, and vice versa. One example is the relationship between the order of an element and the order of the associated cyclic subgroup:

$$|\langle a \rangle| = o(a).$$

2.5.2 Subgroup Criteria

Now we'll give a couple of other criteria for showing that a subset of a group is actually a subgroup. When dealing with specific examples, it is often easiest to simply verify the axioms that we have been using all along. However, when proving things about subgroups it can be useful to use one of the following characterizations. The first one shows that we can collapse the usual subgroup axioms into a single condition.

Theorem 2.5.7. *Let G be a group. A nonempty subset $H \subseteq G$ is a subgroup if and only if whenever $a, b \in H$, $ab^{-1} \in H$.*

Proof. Suppose that $H \leq G$, and let $a, b \in H$. Then $b^{-1} \in H$, so $ab^{-1} \in H$ since H is closed.

Conversely, suppose that $ab^{-1} \in H$ for all $a, b \in H$. Then for any $a \in H$, we can take $a = b$ and conclude that

$$e = aa^{-1} \in H,$$

so H contains the identity. Since $e \in H$, for any $a \in H$ we have

$$a^{-1} = ea^{-1} \in H,$$

so H is closed under taking inverses. Finally, we claim that H is closed under the group operation. If $a, b \in H$, then $b^{-1} \in H$, so $b^{-1}a^{-1} \in H$, and therefore

$$ab = \left((ab)^{-1}\right)^{-1} = (b^{-1}a^{-1})^{-1} \in H.$$

Thus H is closed, hence a subgroup of G. □

The next criterion is quite interesting. It obviously reduces the number of things that one needs to check, but it only works for a *finite* subset of a group G.

Theorem 2.5.8. *Let G be a group and H a nonempty **finite** subset of G. Then H is a subgroup if and only if H is closed under the operation on G.*

Proof. If H is a subgroup, then it is obviously closed by hypothesis.

On the other hand, we are assuming that H is closed, so we need to verify that $e \in H$ and that for every $a \in H$, $a^{-1} \in H$ as well. Since $\{e\} \leq G$, we may assume that H is nontrivial, i.e. that H contains an element a distinct from the identity. Since H is closed, the elements

$$a, a^2, a^3, \ldots$$

are all in H, and since H is finite, this list cannot go on forever. That is, we must eventually have duplicates on this list, so

$$a^i = a^j$$

for some $1 \leq i < j \leq |H|$. Since $i < j$, $j - i \geq 0$ and we have

$$a^i = a^j = a^{j-i}a^i,$$

and using cancellation, we get

$$a^{j-i} = e.$$

Therefore, $e \in H$. Now observe that $j - i - 1 \geq 0$, so $a^{j-i-1} \in H$, and

$$aa^{j-i-1} = a^{j-i} = e,$$

so $a^{-1} = a^{j-i-1} \in H$. Therefore, H is a subgroup of G. □

This theorem has an easy corollary, which is useful when the group is finite.

Corollary 2.5.9. *If G is a **finite** group, a subset $H \subseteq G$ is a subgroup of G if and only if it is closed under the operation on G.*

2.5.3 Subgroups of Cylic Groups

Let's return now to the cyclic case. There is one very important thing that we can say about cyclic groups, namely that their subgroups are always cyclic.

Theorem 2.5.10. *A subgroup of a cyclic group is cyclic.*

Proof. Let $G = \langle a \rangle$ be a cyclic group and let H be a subgroup of G. We may assume that $H \neq \{e\}$, since $\{e\}$ is already known to be cyclic. Then H contains an element other than e, which must have the form a^m for some $m \in \mathbb{Z}$ since G is cyclic. Assume that m is the *smallest* positive integer for which $a^m \in H$. We claim that $H = \langle a^m \rangle$. To do this, we need to show that if $a^n \in H$, then a^n is a power of a^m.

Suppose that $a^n \in H$, and use the Division Algorithm to write $n = qm + r$, where $0 \leq r < m$. Then

$$a^n = a^{qm+r} = a^{qm}a^r = (a^m)^q a^r.$$

Since H is a subgroup, $(a^m)^{-q} \in H$, hence $(a^m)^{-q}a^n \in H$, and it follows that

$$a^r = (a^m)^{-q}a^n$$

is in H. But $r < m$ and we have assumed that m is the smallest positive integer such that $a^m \in H$, so we must have $r = 0$. In other words, $a^n = (a^m)^q$, so $a^n \in \langle a^m \rangle$. Since a^n was an arbitrary element of H, we have shown that $H \subseteq \langle a^m \rangle$. Since $a^m \in H$, we also have $\langle a^m \rangle \subseteq H$, so $H = \langle a^m \rangle$, and H is cyclic. □

This theorem has a particularly nice corollary, which tells us a lot about the structure of $\mathbb{Z}$ as an additive group.

Corollary 2.5.11. *The only subgroups of $\mathbb{Z}$ are the cyclic subgroups $n\mathbb{Z}$, where $n \in \mathbb{Z}$.*

Proof. The cyclic subgroups of $\mathbb{Z}$ are simply $n\mathbb{Z} = \langle n \rangle$ for any $n \in \mathbb{Z}$. By the theorem, the only subgroups of $\mathbb{Z}$ are the cyclic ones, so we are done. □

The next corollary is quite interesting. Recall that if $G = \langle a \rangle$ is a finite cyclic group of order n and $a^j \in G$, we have a formula for the order of a^j:

$$o(a^j) = \frac{n}{\gcd(j, n)}.$$

This tells us something in particular, which we mentioned earlier: if $a^j \in \langle a \rangle$, then $o(a^j)$ divides $|\langle a \rangle|$. In particular, the order of the cyclic subgroup generated by a^j must divide the order of $\langle a \rangle$. This holds more generally—in fact, it is true for any finite group, as we will see when we prove Lagrange's theorem. However, a sort of converse holds for cyclic groups: if m divides the order of the group, then there is a subgroup of that order. Moreover, if a cyclic group has a subgroup of a given order, then that subgroup is unique.[12]

Corollary 2.5.12. *Let $G = \langle a \rangle$ be a cyclic group of order n. If m is a positive divisor of m, then G has exactly one subgroup of order m.*

Proof. First we need to show that G even has a subgroup of order m whenever $m \mid n$. Well, suppose that $m \mid n$, and put $b = a^{n/m}$. Then by Theorem 2.4.12,

$$o(b) = \frac{n}{\gcd(n/m, n)} = \frac{n}{n/m} = m.$$

Therefore, $\langle b \rangle$ has m elements, so G has a subgroup of order m.

Now we need to show that G only possesses one subgroup of order m. Suppose that a^j is another element of order m, so that $\langle a^j \rangle$ has m elements. Then

$$\left|a^j\right| = \frac{n}{\gcd(j, n)} = m = \frac{n}{\gcd(n/m, n)} = |b|.$$

Therefore, $\gcd(j, n) = \gcd(n/m, n) = n/m$. In particular, n/m divides j, so we can write

$$j = r(n/m)$$

for some $r \in \mathbb{Z}$. Then

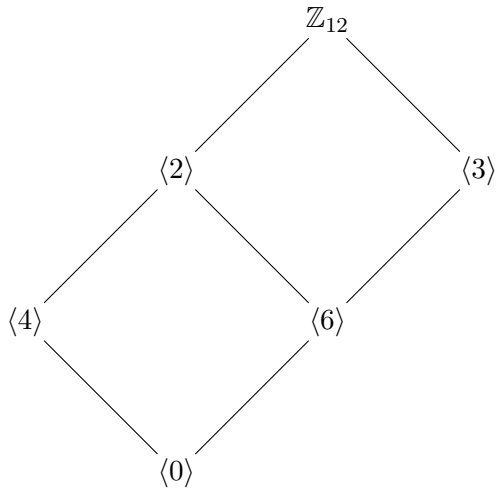
$$a^j = a^{rn/m} = (a^{n/m})^r = b^r,$$

so $a^j \in \langle b \rangle$. This forces $\langle a^j \rangle \subset \langle b \rangle$, and since both sets have the same (finite) number of elements, $\langle a^j \rangle = \langle b \rangle$. □

Let's put this to work in an example.

Example 2.5.13. Let's write down all of the subgroups of $\mathbb{Z}_{12}$. We know that there will be one of each order that divides 12, and these divisors are 1, 2, 3, 4, 6, and 12. The order 1 subgroup is just $\{0\}$, and for order 12 we have the whole group $\mathbb{Z}_{12}$. For the others, we need to find an element of that order. We can see that 6 has order 2, so $\{0, 6\}$ is our order 2 subgroup. For order 3, we can use 4 as the generator, so $\{0, 4, 8\}$ is the subgroup. For order 4, we have $3 \in \mathbb{Z}_{12}$, and the subgroup is $\{0, 3, 6, 9\}$. Finally, 2 has order 6, and the subgroup is $\{0, 2, 4, 6, 8, 10\}$. Therefore, the subgroup lattice of $\mathbb{Z}_{12}$ looks like:

[12] These latter two statements fail miserably for arbitrary finite groups. If m divides the order of the group, there need not be a subgroup of order m. (It is true for abelian groups, however.) Also, if a subgroup of order m exists, it need not be unique. (Look at the Klein 4-group, for example.)



2.6 Lagrange's Theorem

In our investigation of cyclic groups, we noticed one thing—that if $G = \langle a \rangle$ is finite, the order of any element divides the order of the group. This implies something else: if $H \leq \langle a \rangle$, then we know that $H = \langle a^m \rangle$ for some $m \in \mathbb{Z}$, and

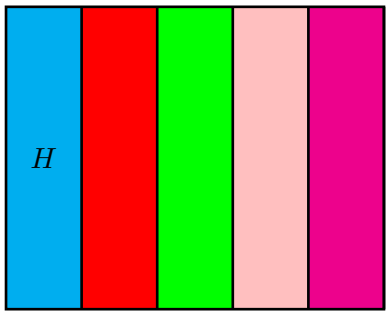
$$|H| = |\langle a^m \rangle| = |a^m| \text{ divides } |a| = |G|.$$

Thus the order of any subgroup divides the order of the group. This is no accident—it holds much more generally. In fact, it is true not just for cyclic groups, but for any finite group and any subgroup.

Theorem 2.6.1 (Lagrange)**.** *Let G be a finite group and H be a subgroup of G. Then $|H|$ divides $|G|$.*

We're not going to prove Lagrange's theorem yet. The proof isn't hard, but we don't quite have the right tools for writing it down yet. We need to develop some new language in order to properly prove it. The concepts that we'll talk about will be really important for the rest of the course, but the first thing they will buy us is a proof of Lagrange's theorem.

The idea of the proof will be the following: given a finite group G and $H \leq G$, we want to "carve up" G into a collection of subsets, all of which are determined by the subgroup H. We will first need to figure out what these subsets should be.



The proper way to say this is that we want to *partition* G. In order to do this, we need to talk about **equivalence relations**.

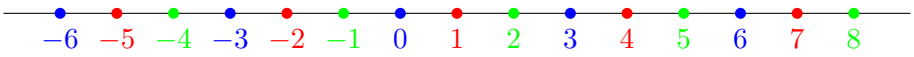
Remark 2.6.2. Lagrange's theorem is a statement about *finite* groups, but we are going to look at arbitrary groups for this part. A lot of the techniques (specifically the idea of cosets) will be important in a more general setting.

Before we formally introduce equivalence relations, let's do an example for motivation. It should be eerily familiar, or it will be once we're done with it.

Example 2.6.3. Let's look at our favorite group—$\mathbb{Z}$. We just saw that the only subgroups of $\mathbb{Z}$ are those of the form $n\mathbb{Z}$, for $n \in \mathbb{Z}^+$. Let's take $3\mathbb{Z}$, for example. What if we start "translating" this subgroup by integers?

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

This is depicted on the color-coded number line below:



What if we keep translating? We just get the same sets all over again:

$$3\mathbb{Z} + 3 = 3\mathbb{Z} + 0$$
$$3\mathbb{Z} + 4 = 3\mathbb{Z} + 1$$

and so on. Also, do these "translates" of $3\mathbb{Z}$ have any overlap? No—if two of the sets are distinct, then they are actually completely disjoint. Finally, they fill up all of $\mathbb{Z}$, in the sense that their union is $\mathbb{Z}$. In other words, every integer belongs to one of the sets on this list.

What do you notice about the elements of each of these sets? They are all congruent mod 3:

$$3\mathbb{Z} + 0 = \{a \in \mathbb{Z} : a \equiv 0 \bmod 3\}$$
$$3\mathbb{Z} + 1 = \{a \in \mathbb{Z} : a \equiv 1 \bmod 3\}$$
$$3\mathbb{Z} + 2 = \{a \in \mathbb{Z} : a \equiv 2 \bmod 3\}$$

In fact, we could say that these sets are determined by the relationship of "congruence mod 3"—two integers a and b fall into the same "translate" if and only if they are congruent mod 3, i.e., if and only if 3 divides $a - b$.

Before moving on, let's recap the key observations that we've made in this example. By shifting around the subgroup $3\mathbb{Z}$, we obtain translates of the subgroup such that

1. They are determined by the relation $a \equiv b \bmod 3$.
2. Distinct translates are disjoint.
3. Their union is all of $\mathbb{Z}$.

The latter two conditions say that we have **partitioned** $\mathbb{Z}$. This is the situation that we would like to generalize.

2.6.1 Equivalence Relations

In mathematics, when one wants to partition a set, the natural way to do it is with something called an *equivalence relation*. These objects are important throughout mathematics, and not just in the field of algebra. This means that we are about to step away from algebra for a moment and talk about some even more abstract ideas (if you can believe such a thing). The primary examples will be from algebra, of course, but these sorts of concepts are ubiquitous throughout mathematics. They are used in algebra, topology, analysis, combinatorics, and other fields. Therefore, this will be a good place for you to encounter them.

Even though equivalence relations are fairly abstract, we'll use congruence mod n as a guiding example. In that case, we have some very nice things going on. There are three things in particular that we can single out:

1. If $a \in \mathbb{Z}$, then $a \equiv a \bmod n$.
2. If $a, b \in \mathbb{Z}$ and $a \equiv b \bmod n$, then $b \equiv a \bmod n$.
3. If $a, b, c \in \mathbb{Z}$ with $a \equiv b \bmod n$ and $b \equiv c \bmod n$, then $a \equiv c \bmod n$.

These are the properties that we will eventually want equivalence relations to possess. Before we can properly define them, we need to make the word "relation" precise.

Definition 2.6.4. Let S be a set. A **relation** on S is a set $R \subseteq S \times S$ of ordered pairs.

Remark 2.6.5. The definition above is the formal, precise way of describing a relation. It is not a particularly helpful way in which to think about relations. You should simply think of a relation as a way to "pair off" elements of S. To this end,

we will usually write $x \sim y$ to mean $(x, y) \in R$, and we will say that "x is related to y." We will usually refer to $\sim$ as the relation, and not worry too much about the proper definition.

We will not concern ourselves with general relations. Our main interests will lie in relations which behave like "congruence mod n." These are the equivalence relations.

Definition 2.6.6. Let S be a set. An **equivalence relation** on S is a relation $\sim$ on S satisfying the following three properties:

1. **Reflexivity:** $a \sim a$ for all $a \in S$.
2. **Symmetry:** If $a, b \in S$ and $a \sim b$, then $b \sim a$.
3. **Transitivity:** If $a, b, c \in S$ with $a \sim b$ and $b \sim c$, then $a \sim c$.

We will usually think of an equivalence relation as a way of pairing off elements of S by equivalence. If $a \sim b$, we will usually say that "a is equivalent to b."

The whole point of introducing equivalence relations is to obtain a method of partitioning a group. Therefore, we will of course be interested in subsets of S that are somehow "determined" by an equivalence relation.

Definition 2.6.7. Given $a \in S$, define

$$[a] = \{b \in S : b \sim a\}.$$

We call $[a]$ the **equivalence class** of a, and a is called a **representative**[13] of the equivalence class.

Example 2.6.8. Here are some examples of equivalence relations.

1. This is one of the most important examples, and it's the one we've already seen. Fix $n \in \mathbb{Z}$, and define $\sim$ on $\mathbb{Z}$ by $a \sim b$ if and only if $a \equiv b \bmod n$. We have already checked that this is an equivalence relation. What are the equivalence classes? For $a \in \mathbb{Z}$, we have

$$[a] = \{b \in \mathbb{Z} : b \equiv a \bmod n\} = n\mathbb{Z} + a.$$

2. Define $\sim$ on $\mathbb{R}$ by $a \sim b$ if $a = b$. Then $\sim$ is an equivalence relation:
   - **Reflexive:** For any $a \in \mathbb{R}$, $a = a$, so $a \sim a$.
   - **Symmetric:** If $a, b \in \mathbb{R}$ and $a = b$, then $b = a$, so $a \sim b$ implies that $b \sim a$.

[13] Note that an equivalence class may have many representatives. In fact, a representative is just a chosen element of the equivalence class, so any element of the class can be taken as a representative.

- **Transitive:** If $a, b, c \in \mathbb{R}$ with $a = b$ and $b = c$, then $a = c$.

What are the equivalence classes? Each class consists of a single element:

$$[a] = \{a\}.$$

3. Define $\sim$ on $\mathrm{GL_n}(\mathbb{R})$ by $A \sim B$ if $\det(A) = \det(B)$. It's easy to see that this is an equivalence relation, and that

$$[A] = \{B \in \mathrm{GL_n}(\mathbb{R}) : \det(B) = \det(A)\}.$$

4. This example is really the prototypical equivalence relation, in the sense that any equivalence relation can be viewed in this way. Let S be any set, and let $\{S_i\}$ be a **partition** of S. That is, $S_i \subset S$ for all i, $S = \bigcup S_i$, and the S_i are pairwise disjoint:

$$S_i \cap S_j = \varnothing$$

if $i \neq j$. We call the S_i the **cells** of the partition. We can define an equivalence relation $\sim$ on S by $a \sim b$ if and only if a and b belong to the same cell S_i. It is easy to check that $\sim$ is reflexive, symmetric, and transitive, hence an equivalence relation.

Exercise 2.5. Verify that the relation $\sim$ defined in Example 2.6.8(4) is an equivalence relation.

Aside 2.6.9. While we're on the topic of specific examples, let's make a couple of comments regarding $\mathbb{Z}$ with the relation of congruence mod n. Note that the equivalence classes are simply congruence classes mod n: every element of the class $[a]$ is congruent to a mod n. Thus all elements of $[a] = n\mathbb{Z} + a$ yield the same element of $\mathbb{Z}_n$ when reduced mod n. Therefore, the equivalence classes $n\mathbb{Z} + a$ can be identified with elements of $\mathbb{Z}_n$. We will see soon that this is really the correct way to view $\mathbb{Z}_n$. The elements of $\mathbb{Z}_n$ are actually the *equivalence classes* mod n, and we define modular addition and multiplication by

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

and

$$[a][b] = [ab].$$

Of course, we would need to know that these operations are well-defined. We'll take care of that when we revisit $\mathbb{Z}_n$ later on. We will discuss these ideas in more detail when we encounter quotient groups, of which $\mathbb{Z}_n$ will be the foremost example.

Now let's return to general equivalence relations. We said that we were introducing them to allow us to partition sets, so we should check that they actually do this. We've already seen that a partition of a set S actually imposes an equivalence relation on S, so we are really checking that equivalence relations and partitions go hand in hand.

Theorem 2.6.10. *Let $\sim$ be an equivalence relation on a set S. The equivalence classes of $\sim$ **partition** S, in the sense that:*

1. *Given $a, b \in S$, either $[a] \cap [b] = \varnothing$, or $[a] = [b]$.*
2. *S is the union of all the equivalence classes of $\sim$. That is, every element of S belongs to some equivalence class (and only one, by condition 1).*

Proof. Let $a, b \in S$. Then either $[a] \cap [b] = \varnothing$, or $[a] \cap [b]$ is nonempty. In the second case, there is at least one element $c \in [a] \cap [b]$. Then $c \in [a]$, so $a \sim c$, and $c \in [b]$, so $c \sim b$. By transitivity, $a \sim b$. Thus $a \in [b]$, and if $x \in [a]$, then $x \sim b$ by transitivity, so $x \in [b]$ as well. Thus $[a] \subset [b]$. By symmetry, we also have $b \in [a]$, and a similar transitivity argument shows that $[b] \subset [a]$. Thus $[a] = [b]$.

To see that S is the union of the equivalence classes, we just need to notice that every $a \in S$ belongs to one of the equivalence classes. Specifically, $a \in [a]$ since $\sim$ is reflexive. Thus S is contained in the union of the equivalence classes, so in fact S equals the union of the equivalence classes. Therefore, the equivalence classes of $\sim$ partition S. □

2.6.2 Cosets

Let's step back to reality now. (Or, as close to reality as abstract algebra can be.) We are especially interested in the case where the set in question is actually a group, and the equivalence relation has something to do with a given subgroup. That is, we want to partition a group G into subsets, each of which is determined by some fixed subgroup H. Once we have done this, we will able to write down a proof of Lagrange's theorem in a nice way. Our present goal then is to find an equivalence relation on a group G which is somehow related to a subgroup H. There are two very similar ones that we can define, and either one will work.

Let's return for a moment to our key example. The relation $\sim$ that we defined on $\mathbb{Z}$ by

$$a \sim b \iff a \equiv b \bmod n$$

is really equivalent to specifying that $a \sim b$ if and only if $n \mid (a - b)$. This in turn is equivalent to saying that $a - b \in n\mathbb{Z}$. Let's try to generalize this.

Example 2.6.11. Let G be a group and $H \leq G$. Define a relation $\sim_H$ on G by: $a \sim_H b$ if and only if $ab^{-1} \in H$. Is this an equivalence relation?

- **Reflexive:** If $a \in G$, then $aa^{-1} = e \in H$, so $a \sim_H a$.

- **Symmetric:** Suppose that $a, b \in G$ and $a \sim_H b$, so $ab^{-1} \in H$. To show that $b \sim_H a$, we need to know that $ba^{-1} \in H$. But

$$ba^{-1} = (ab^{-1})^{-1} \in H,$$

since H is a subgroup of G. Thus $b \sim_H a$, and the relation is symmetric.

- **Transitive:** Suppose that $a \sim_H b$ and $b \sim_H c$. Is $a \sim_H c$? We need to know that $ac^{-1} \in H$. Well,

$$ac^{-1} = a(b^{-1}b)c^{-1} = (ab^{-1})(bc^{-1}),$$

and $ab^{-1} \in H$ (since $a \sim_H b$) and $bc^{-1} \in H$ (since $b \sim_H c$), so $ac^{-1} \in H$. Thus $a \sim_H c$.

Here's the next logical question: what are the equivalence classes of $\sim_H$? Well,

$$[a]_H = \{b \in G : b \sim_H a\} = \left\{b \in G : ba^{-1} \in H\right\}.$$

If $a \sim_H b$, then $ab^{-1} \in H$, i.e., there exists $h \in H$ such that

$$ba^{-1} = h,$$

or $b = ha$. If we define

$$Ha = \{ha : h \in H\},$$

then we have just shown that $[a]_H \subseteq Ha$. On the other hand, given $b = ha \in Ha$,

$$ba^{-1} = (ha)a^{-1} = h(aa^{-1}) = h \in H,$$

so $b \sim_H a$. Thus $Ha \subseteq [a]_H$. Therefore, we have shown that the equivalence classes are exactly

$$[a]_H = Ha,$$

and we'll generally use Ha instead of the $[\cdot]_H$ notation when working in the group case. These equivalence classes actually have a special name, which we will use constantly from now on.

Definition 2.6.12. Sets of the form Ha for $a \in G$ are called **(right) cosets** of H in G.

Let's summarize what we proved above.

Theorem 2.6.13. *Let G be a group, H a subgroup of G, and let $\sim$ be the relation on G given by*

$$a \sim_H b \iff ab^{-1} \in H.$$

Then $\sim$ is an equivalence relation, and the equivalence classes are precisely the right cosets of H:

$$[a]_H = Ha.$$

Furthermore, if $a, b \in G$, then either $Ha \cap Hb = \varnothing$ or $Ha = Hb$, and

$$Ha = Hb \iff ab^{-1} \in H.$$

Finally, G is the union of all the right cosets of H, so the cosets partition G.

Two of the examples of equivalence relations that we mentioned last time are actually the relation $\sim_H$ in disguise. In fact, this relation is meant to be a generalization of congruence mod n.

Example 2.6.14. 1. Consider the group $\mathbb{Z}$ and the subgroup $H = n\mathbb{Z}$. Given $a, b \in \mathbb{Z}$, we have seen that $a \sim_H b$ if and only if $a - b \in H$, which is just the additive way of writing $ab^{-1} \in H$. Thus $\sim_H$ is really just congruence mod n, and the right cosets of H are

$$[a]_H = n\mathbb{Z} + a = \{nc + a : c \in \mathbb{Z}\},$$

and there are n of them, namely $n\mathbb{Z} + 0, n\mathbb{Z} + 1, \ldots, n\mathbb{Z} + (n-1)$.

2. Consider $H = \mathrm{SL_n}(\mathbb{R}) \leq \mathrm{GL_n}(\mathbb{R})$. If $A, B \in \mathrm{GL_n}(\mathbb{R})$, $A \sim_H B$ precisely when $AB^{-1} \in \mathrm{SL_n}(\mathbb{R})$, or

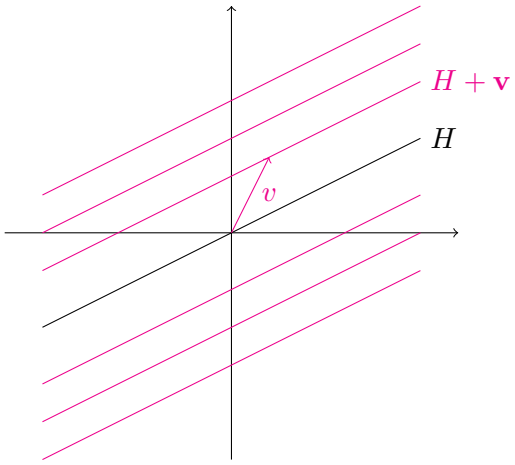
$$\det(AB^{-1}) = 1.$$

But

$$\det(AB^{-1}) = \det(A)\det(B^{-1}) = \frac{\det(A)}{\det(B)},$$

so $A \sim_H B$ if and only if $\det(A) = \det(B)$.

3. Recall that any vector space is an abelian group under addition, and any vector subspace is a subgroup. In particular, let $G = \mathbb{R}^2$ and let H be a 1-dimensional subspace, i.e., a line through the origin. If $\mathbf{v} \in \mathbb{R}^2$ is any vector, the coset $H + \mathbf{v}$ is just a "parallel translate" of the line H.



4. Since Lagrange's theorem deals with finite groups, it would probably be helpful to look at one of those. Let's look at the dihedral group D_3, and let

$$H = \{i, r_1, r_2\},$$

the rotation subgroup. What are the cosets? There are only 2: H itself, and

$$Hm_1 = \{m_1, m_2, m_3\}.$$

(Note that we could also take m_2 or m_3 as a representative of the coset.)

Now that we've built up the appropriate machinery, let's go ahead and use it to formally prove Lagrange's theorem.

Theorem 2.6.15 (Lagrange)**.** *Let G be a finite group and let $H \leq G$. Then $|H|$ divides $|G|$.*

Proof. Let $Ha_1, \ldots, Ha_k$ denote the distinct cosets of H in G. That is, $a_1, \ldots, a_k$ all represent different cosets of H, and these are all the cosets. We know that the cosets of H partition G, so

$$|G| = \#(Ha_1) + \cdots + \#(Ha_k). \tag{2.1}$$

(Here $\#$ means the *cardinality* of the set, or simply the number of elements in that set.) Therefore, it will be enough to show that each coset has the same number of elements as H.

We need to exhibit a bijection between H and Ha_i for each i. For each $i = 1, \ldots, k$, define a function $f_i : H \to Ha_i$ by

$$f(h) = ha_i.$$

If we can prove that f is a bijection, then we will have

$$|H| = \#(Ha_i)$$

for all i. This is fairly straightforward: if $h_1, h_2 \in H$ with $f(h_1) = f(h_2)$, then

$$h_1 a_i = h_2 a_i,$$

which implies that $h_1 = h_2$, so f is one-to-one. To see that it is onto, take $h \in H$; then $f(h) = ha_i$.

Thus all the cosets have the same number of elements, namely $|H|$, and (2.1) really says that

$$|G| = \underbrace{|H| + \cdots + |H|}_{k \text{ times}} = k|H|.$$

Therefore, $|H|$ does indeed divide $|G|$. □

The number that we called k in the proof is actually quite useful, and we will therefore give it a special name.

Definition 2.6.16. The number of distinct (right) cosets of H in G is called the **index** of H in G, denoted by

$$[G : H].$$

The set of all right cosets of H in G is denoted by G/H, so

$$\#(G/H) = [G : H].$$

Note that we can actually rephrase Lagrange's theorem in terms of the index: if G is a finite group and $H \leq G$, then

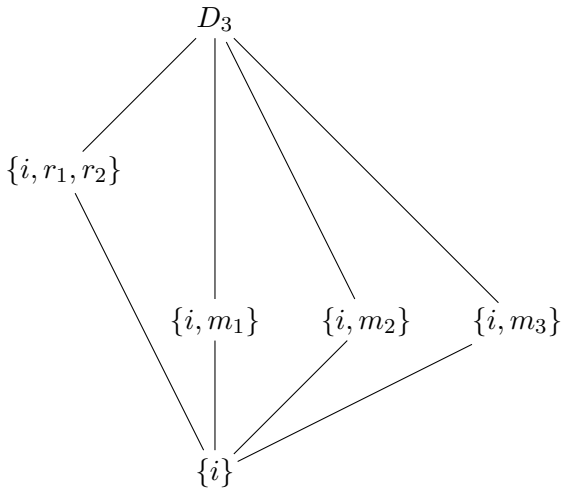
$$|G| = |H|[G : H].$$

We'll now begin to see that Lagrange's theorem has many very useful consequences. For one, it greatly simplifies the search for subgroups of a given finite group.

Example 2.6.17. Let's try to find all the subgroups of D_3. Since $|D_3| = 6$, we know that the only possible orders are $1, 2, 3$, and 6. The subgroups are then

$$\begin{aligned} 1: &\quad \{i\} \\ 2: &\quad \{i, m_1\}, \{i, m_2\}, \{i, m_3\} \\ 3: &\quad \{i, r_1, r_2\} \\ 6: &\quad D_3 \end{aligned}$$

We can even draw a lattice diagram to illustrate the subgroup structure of D_3:



Lagrange's theorem also has a couple of easy yet powerful corollaries. One simply states that the order of any element divides the order of the group, which we already know for cyclic groups. The other tells us that groups of prime order are particularly special, and they behave in a very rigid way.

Corollary 2.6.18. *Let G be a finite group with $|G| = n$, and let $a \in G$. Then $o(a)$ divides $n = |G|$, and*

$$a^n = e.$$

Proof. We defined $o(a)$ to be $|\langle a \rangle|$, and $\langle a \rangle$ is a subgroup of G, so its order divides $|G|$ by Lagrange's theorem. Therefore, $o(a) \mid n$, and we can write $n = o(a)m$ for some $m \in \mathbb{Z}$. Then

$$a^n = a^{o(a)m} = \left(a^{o(a)}\right)^m = e.$$

□

Corollary 2.6.19. *If G is a finite group of prime order p, then G is cyclic.*

Proof. Since p is prime, $p \geq 2$, and G contains at least one element a with $a \neq e$. By the previous corollary, $o(a)$ divides p, and since $a \neq e$, $o(a) \neq 1$. Since p is prime, we must have $o(a) = p$, so a generates G. Thus G is cyclic. □

In this proof, we showed that any nonidentity element of a group G with prime order is actually a generator for the group. This implies the following fact regarding subgroups of such groups.

Corollary 2.6.20. *If G is a finite group of prime order p, then G has no subgroups other than $\{e\}$ and G itself.*

Note that this last part implies that if φ is an isomorphism, then $o(\varphi(a)) = o(a)$ for all $a \in G_1$.

Proposition 2.7.12. *If $\varphi : G_1 \to G_2$ is an isomorphism, then $o(\varphi(a)) = o(a)$ for all $a \in G_1$.*

Proof. Suppose first that $a \in G_1$ has finite order. Then we showed in Proposition 2.7.10 that $o(\varphi(a))$ divides $o(a)$. But φ^{-1} is also a homomorphism, so $o(a) = o(\varphi^{-1}(\varphi(a)))$ must divide $o(\varphi(a))$. Since both integers divide each other, they must be the same, and $o(a) = o(\varphi(a))$.

Now note that Proposition 2.7.10 also implies that a has finite order if and only if $\varphi(a)$ does. It follows that a has infinite order if and only if $\varphi(a)$ does, so we are done. □

2.8 The Symmetric Group Redux

We're now going to take a short detour into the symmetric group, which we introduced quite some time ago. We've put off the details in order to develop some more general tools for working with groups. Never fear; we'll be going back into the abstraction very soon, but right now we'll try to have a little fun with S_n.

The reason that we are doing this is that S_n is really the most fundamental finite nonabelian group, so it would be very useful to understand its structure. We will see via Cayley's theorem that every group is really a permutation group, in a certain sense. In fact, groups were originally thought of only as permutation groups. It was Cayley who first gave the abstract definition of a group and then proved his eponymous theorem. Therefore, if we could somehow understand all of the subgroups of S_n, then we would understand all finite groups. This is an ambitious goal, and it is far too much to ask for. However, understanding the symmetric group will give us information on it and other groups that are easily realized as permutations (like D_n, for example).

2.8.1 Cycle Decomposition

When we originally talked about S_n, we introduced some compact notation for writing down a permutation, called **two-line notation**: if $\sigma \in S_n$, then

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix}.$$

This is obviously a nice, short way to write down an element of S_n, and multiplying two permutations in two-line notation is relatively straightforward. However, because of its simplicity it can obscure the structure of σ, and it's not good enough

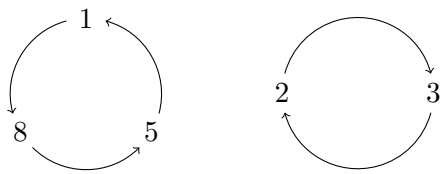
from the standpoint of algebra. For example, it we look at the permutation $\sigma \in S_8$ given by

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

what is really going on?

- σ fixes 4, 6, and 7.
- 2 and 3 only interact with each other.
- 1, 5, and 8 only interact with each other.

Thus we've written down a bunch of extra information that isn't really necessary, and we've obscured some of the structure that σ possesses in the process. The relevant data is the fact that σ "cycles through" 1, 5, and 8 and it "cycles through" 2 and 3.



With this in mind, perhaps there is another way to write down σ that truly captures what σ is doing. There is, and it is called the **cycle decomposition** of σ.

Let's try to be more precise about what we've done above: we have

$$\begin{aligned} \sigma(1) &= 5 \\ \sigma^2(1) &= \sigma(5) = 8 \\ \sigma^3(1) &= \sigma(8) = 1 \\ \sigma^4(1) &= \sigma(1) = 5 \\ &\vdots \end{aligned}$$

and

$$\begin{aligned} \sigma(2) &= 3 \\ \sigma^2(2) &= \sigma(3) = 2 \end{aligned}$$

and so on. This might look familiar from your take-home exam:

Theorem 2.8.1. *Define a relation $\sim$ on the set $\{1, 2, \ldots, n\}$ by*

$$i \sim j \quad \iff \quad \sigma^m(i) = j$$

for some $m \in \mathbb{Z}^+$. Then $\sim$ is an equivalence relation.

The equivalence classes would have the following form, of course:

$$[i] = \{i, \sigma(i), \sigma^2(i), \sigma^3(i), \ldots\}.$$

Example 2.8.2. Let $\sigma \in S_8$ be as above. What are the equivalence classes? We have

$$\begin{aligned}[1] &= \{1, 5, 8\} \\ [2] &= \{2, 3\} \\ [4] &= \{4\} \\ [6] &= \{6\} \\ [7] &= \{7\}\end{aligned}$$

We can use these equivalence relations to break up σ into pieces called **cycles**. In the notation that we are developing, we will write

$$\sigma = (1\ 5\ 8)(2\ 3)(4)(6)(7)$$

Note that we have written the integers in each cycle *in the order in which they are hit by* σ. That is, (1 5 8) and (1 8 5) are not the same permutation. We'll usually suppress the numbers that are fixed by σ, since they information that they encode is extraneous. Therefore, we will simply write

$$\sigma = (1\ 5\ 8)(2\ 3).$$

We've been throwing around the word "cycle" without properly defining it. You probably have the idea already, but we should still be rigorous.

Definition 2.8.3. A permutation of the form

$$\sigma = (i_1\ i_2\ \cdots\ i_k),$$

where $\sigma(i_m) = i_{m+1}$ for $1 \leq m < k$ and $\sigma(i_k) = i_1$ is called a k**-cycle**.

The cycles are important in that they are the building blocks of S_n. When finding the cycle decomposition of a permutation σ, we are really factoring σ into cycles, in much the same way that one factors integers into primes.

Example 2.8.4. Let's consider the permutation $\tau \in S_9$ given by

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 9 & 6 & 3 & 7 & 1 & 8 & 2 & 5 \end{pmatrix}.$$

How do we write τ as a product of cycles? We start with 1, and we find the *cycle determined by it*:

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 9 & 6 & 3 & 7 & 1 & 8 & 2 & 5 \end{pmatrix}$$

Thus we have the cycle

$$(1\ 4\ 3\ 6).$$

Now we go to the next smallest integer, which is 2, and we find its cycle:

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 9 & 6 & 3 & 7 & 1 & 8 & 2 & 5 \end{pmatrix}$$

so the cycle is

$$(2\ 9\ 5\ 7\ 8).$$

Thus

$$\tau = (1\ 4\ 3\ 6)(2\ 9\ 5\ 7\ 8).$$

Example 2.8.5. Let's find the cycle decomposition of $\sigma \in S_5$, where

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

Well, the cycle determined by 1 is

$$(1\ 2\ 3),$$

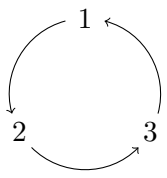
and the smallest integer left is 4, whose cycle is

$$(4\ 5).$$

Therefore,

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

Remark 2.8.6. Note that we have written our permutations in such a way that the cycles always start with the smallest possible integer. This is the standard way to write permutations in cycle notation, but it is not necessary to do so. For example, the cycles (1 2 3) and (3 1 2) are the same. You could think of them both as encodings of the picture



Chapter 3

Ring Theory

We're now done with our study of groups (at least for the sake of groups). The plan now is to move on to more complicated algebraic structures, namely objects called **rings**.

3.1 Rings

You'll probably agree that groups were kind of strange at first. We had a set with a single binary operation that satisfied certain nice properties, and we were able to come up with many examples of such things. Some of these examples were ones we already knew, such as $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{Q}$, all under addition. However, these examples all really have *two* binary operations, namely *addition* and *multiplication*.

Example 3.1.1. We already know that $\mathbb{Z}$ forms an abelian group under addition. What nice properties are satisfied by multiplication on $\mathbb{Z}$?

- associativity
- commutativity
- identity
- distributivity (if $a, b \in \mathbb{Z}$ and $n \in \mathbb{Z}$, then $n(a+b) = na + nb$)

This will mean that $\mathbb{Z}$ is really a model example of a ring.

Definition 3.1.2. A **ring** is a set R equipped with two binary operations, denoted by $+$ and $\cdot$, such that

1. $\langle R, + \rangle$ is an abelian group:
   (a) $+$ is associative and commutative.
   (b) There is an additive identity $0 \in R$ such that $0 + a = a$ for all $a \in R$.

(c) For each $a \in R$, there is an additive inverse $-a \in R$ so that $a + (-a) = 0$.

2. The multiplication operation $\cdot$ is associative.

3. (Distributive law) For all $a, b, r \in R$, we have

$$r \cdot (a + b) = r \cdot a + r \cdot b$$

and

$$(a + b) \cdot r = a \cdot r + b \cdot r.$$

You'll probably notice that we left two things off this list. We did not require that multiplication be commutative, or that there is even a multiplicative identity. Rings that have these properties are special, and thus have special names.

Definition 3.1.3. A **commutative ring** is a ring R for which

$$a \cdot b = b \cdot a$$

for all $a, b \in R$.

Definition 3.1.4. A **ring with identity**[1] (also called a **ring with unity** or a **unital ring**) is a ring R which contains an element $1 \in R$ (with $1 \neq 0$) satisfying

$$1 \cdot a = a \cdot 1 = a$$

for all $a \in R$.

Before we proceed, let's look at some familiar (and less familiar) examples of rings.

Example 3.1.5.

1. We already saw that $\mathbb{Z}$ is a ring, and actually a commutative ring with identity.

2. Similarly, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all commutative rings with identity with respect to their usual operations.

3. For any n, $\mathbb{Z}_n$ is a commutative ring with identity with respect to modular addition and multiplication.

4. Let's try a noncommutative example. Recall that $M_n(\mathbb{R})$ is the set of all $n \times n$ matrices with real coefficients. Then $M_n(\mathbb{R})$ is a ring with respect to matrix addition and multiplication. It is noncommutative, but it has an identity, namely the identity matrix I.

5. Let $2\mathbb{Z}$ be the set of all even integers. Then $2\mathbb{Z}$ is a commutative ring with respect to the usual arithmetic. It does not have an identity, however, since $1 \notin 2\mathbb{Z}$.

[1] Some mathematicians require that a ring contains an identity element. That is, they use "ring" to mean "ring with identity." Some have suggested the term *rng* to mean a ring without identity.

6. Let $C([0,1],\mathbb{R}) = \{f : [0,1] \to \mathbb{R} : f \text{ is continuous}\}$. Then $C([0,1],\mathbb{R})$ is a ring with respect to pointwise addition and multiplication:

$$f + g(x) = f(x) + g(x)$$

and

$$fg(x) = f(x)g(x).$$

It is also commutative, and its multiplicative identity is the function which is identically 1.

7. This will be one of the main examples that we'll study. Let $\mathbb{Q}[x]$ be the set of all polynomials with rational coefficients. Then $\mathbb{Q}[x]$ forms a ring: we add two polynomials by adding their coefficients, say

$$\left[2x^2 + 3x + \frac{1}{2}\right] + \left[5x^3 + \frac{7}{2}x + 2\right] = 5x^3 + 2x^2 + \frac{13}{2}x + \frac{3}{2}.$$

We multiply them by "foiling":

$$(x^3 + 1)(3x^2 + 4x + 2) = 3x^5 + 4x^4 + 2x^3 + 3x^2 + 4x + 2.$$

This turns $\mathbb{Q}[x]$ into a commutative ring with identity. (We could also do this with any commutative ring in place of $\mathbb{Q}$.)

3.2 Basic Facts and Properties of Rings

We'll soon start investigating particular types of rings. Before we can do this, we should prove some relatively simple facts about rings which will be needed in computations.

Proposition 3.2.1. *Let R be a ring. Then for all $a, b \in R$, we have:*

(a) $0 \cdot a = a \cdot 0 = 0$*;*

(b) $(-a)b = a(-b) = -(ab)$*;*

(c) $(-a)(-b) = ab$*; and*

(d) *if* $1 \in R$*, then* $(-1)a = a(-1) = -a$*.*

Proof. (a) If $a \in R$, then we have

$$0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$$

by the right distributive law. But this means that $0 \cdot a$ is an additive idempotent in the abelian group $\langle R, + \rangle$. There is only one such element, so $0 \cdot a = 0$. A similar argument works to show that $a \cdot 0 = 0$.

(b) Let $a, b \in R$. Then by distributivity,

$$ab + (-a)b = (a + (-a))b = 0 \cdot b = 0,$$

so $(-a)b = -(ab)$. The same sort of argument works to show that $a(-b) = -(ab)$.

(c) If we apply part (b), we have

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

(d) Again by part (b), $(-1)a = -(1 \cdot a) = -a$, and $a(-1) = -(a \cdot 1) = -a$. □

Unlike with groups, there are many different types of rings that one can consider. This arises from the fact that there are some strange things that can happen regarding the multiplication in a ring.

Example 3.2.2. Let $R = M_2(\mathbb{R})$, and let $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Then

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

but neither A nor B is the zero matrix. Such anomalies have a special name in ring theory.

Definition 3.2.3. An element $a \neq 0$ of a (commutative) ring R is called a **zero divisor** if there exists $b \in R$ such that $ab = 0$.

Many of the results that we will obtain will involve commutative rings which are free of zero divisors.

Definition 3.2.4. A commutative ring R is called an **integral domain** if R contains no zero divisors. Equivalently, if $a, b \in R$ with $ab = 0$, then either $a = 0$ or $b = 0$.

Example 3.2.5.

1. The rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all integral domains.
2. When n is composite, $\mathbb{Z}_n$ is not an integral domain. For example, in $\mathbb{Z}_6$,
$$2 \cdot 3 = 0.$$
3. The function ring $C([0,1], \mathbb{R})$ is not an integral domain. Let
$$f(x) = \begin{cases} 0 & x < 1/2 \\ x - 1/2 & x \geq 1/2 \end{cases}$$
and
$$g(x) = \begin{cases} x - 1/2 & x < 1/2 \\ 0 & x \geq 1/2. \end{cases}$$
Then $f(x)g(x) = 0$ for all $x \in \mathbb{R}$, but neither function is identically zero.

Another thing you may have noticed is that elements of rings (with identity) do not necessarily possess multiplicative inverses. The ones that do have a special name.

Definition 3.2.6. Let R be a ring with identity. An element $a \in R$ is called a **unit** if there is a $b \in R$ such that

$$ab = ba = 1.$$

(We usually write $b = a^{-1}$.) The set of all units is denoted by $R^\times$, called the **unit group**[2] of R.

Example 3.2.7. Let's compute the unit groups in some rings.

1. What are the units in $\mathbb{Z}$? The only integers which have multiplicative inverses are 1 and -1, so
$$\mathbb{Z}^\times = \{1, -1\}.$$

2. What is the unit group in $\mathbb{Q}$? Every nonzero rational number has a multiplicative inverse, so
$$\mathbb{Q}^\times = \mathbb{Q} - \{0\}.$$
Conveniently enough, this is the notation that we have already used for the group of nonzero rational numbers under multiplication. Similarly, we have $\mathbb{R}^\times = \mathbb{R} - \{0\}$ and $\mathbb{C}^\times = \mathbb{C} - \{0\}$.

3. Let's consider the ring $\mathbb{Z}_n$ for $n \in \mathbb{Z}$. We proved a long time ago that the elements of $\mathbb{Z}_n$ that have multiplicative inverses are precisely those which are relatively prime to n. Therefore,
$$\mathbb{Z}_n^\times = \{a \in \mathbb{Z}_n : \gcd(a, n) = 1\}.$$
Note that if p is prime, then $\mathbb{Z}_p^\times = \{1, 2, \ldots, p-1\} = \mathbb{Z}_p - \{0\}$.

Of course 0 never has a multiplicative inverse, but we've seen that it's possible that everything else might. Such rings are special.

Definition 3.2.8. A ring R with identity is called a **division ring** if every nonzero $a \in R$ is a unit. Equivalently, $R^\times = R - \{0\}$.

Definition 3.2.9. A commutative division ring is called a **field**.

Example 3.2.10.
1. The integral domains $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all fields.

2. Since only 1 and -1 are units, $\mathbb{Z}$ is not a field.

3. If p is prime, $\mathbb{Z}_p$ is a field. It is an example of a **finite field**.

[2] This name should indicate that $R^\times$ forms a group under multiplication. This fact is not too hard to check.

Exercise 3.1. Let R be a commutative ring with identity. Show that if $u \in R$ is a unit, then u cannot be a zero divisor. As a consequence, any field is an integral domain.

3.2.1 The Quaternions

You may have noticed that we produced several examples of fields (and hence of division rings), but no division rings that weren't fields. There is an example of such a thing, and it's an interesting one mathematically and physically.

Example 3.2.11. The quaternions were invented by the Irish mathematician and physicist William Rowan Hamilton in the middle of the 19th century. Motivated by the complex numbers and their geometric interpretations, he had tried to define multiplication on triples of real numbers. (That is, he tried to turn $\mathbb{R}^3$ into a ring.) This didn't quite work (and, in fact, is impossible), but he had a breakthrough one day while walking around campus. It was so exciting to him that he actually carved his initial identity into a stone bridge.

Hamilton's idea was to consider three elements i, j, and k which behave like the imaginary unit in $\mathbb{C}$:

$$i^2 = j^2 = k^2 = ijk = -1.$$

From this single identity, it is possible to deduce that

$$ij = k, jk = i, ki = j,$$

and that these elements *anticommute*:

$$ij = -ji, kj = -kj, ik = -ki.$$

The **quaternions** are then defined to be the set

$$\mathbb{H} = \{a_1 + a_2 i + a_3 j + a_4 k : a_1, a_2, a_3, a_4 \in \mathbb{R}\}.$$

In other words, one can think of $\mathbb{H}$ as a four-dimensional real vector space with basis $\{1, i, j, k\}$. The addition operation works just like addition of vectors in $\mathbb{R}^4$: for example,

$$(1 + 4i + 7j + 2k) + (-2 + 6i - 9j - 4k) = -1 + 10i - 2j - 2k.$$

In general, we would have

$$(a_1+a_2i+a_3j+a_4k)+(b_1+b_2i+b_3j+b_4k) = (a_1+b_1)+(a_2+b_2)i+(a_3+b_3)j+(a_4+b_4)k.$$

Multiplication works like multiplication of complex numbers, but more complicated. We simply multiply everything out, and then use the quaternion identities listed above to simplify the result:

$$\begin{aligned}(3-5k)(1+4i-5j) &= 3+12i-15j-5k-20ki+25kj \\ &= 3+12i-15j-5k+20j-25i \\ &= 3-13i+5j-5k\end{aligned}$$

There is a general formula for the coefficients of the product of two quaternions, but it's usually easier to simply multiply on a case-by-case basis:

$$(a_1+a_2i+a_3j+a_4k)(b_1+b_2i+b_3j+b_4k) = c_1+c_2i+c_3j+c_4k,$$

where

$$\begin{aligned}c_1 &= a_1b_1-a_2b_2-a_3b_3-a_4b_4 \\ c_2 &= a_1b_2+a_2b_1+a_3b_4-a_4b_3 \\ c_3 &= a_1b_3-a_2b_4+a_3b_1+a_4b_2 \\ c_4 &= a_1b_4+a_2b_3-a_3b_2+a_4b_1.\end{aligned}$$

It's possible to check that multiplication is associative and distributive, which we won't do here. Also, note that for example we have

$$\begin{aligned}(1+3i-4j-10k)(1-3i+4j+10k) &= 1-3i+4j+10k+3i-9i^2+12ij+30ik \\ &\quad -4j+12ji-16j^2-40jk-10k+30ki-40kj-100k^2 \\ &= 1+9+12k-30j-12k+16-40i+30j+40i+100 \\ &= 1+9+16+100 \\ &= 126.\end{aligned}$$

In general, one can verify that

$$(a_1+a_2i+a_3j+a_4k)(a_1-a_2i-a_3j-a_4k) = a_1^2+a_2^2+a_3^2+a_4^2.$$

This allows us to define the inverse of any nonzero quaternion: if $a_1+a_2i+a_3j+a_4k \neq 0$, then the quantity

$$\alpha = a_1^2+a_2^2+a_3^2+a_4^2$$

is nonzero, and it can be checked that

$$(a_1+a_2i+a_3j+a_4k)\left(\frac{a_1}{\alpha}-\frac{a_2}{\alpha}i-\frac{a_3}{\alpha}j-\frac{a_4}{\alpha}k\right) = 1.$$

Thus $\mathbb{H}^\times = \mathbb{H}-\{0\}$, and $\mathbb{H}$ is a division ring. It is not a field, however, since we have $ij \neq ji$, for example.

Remark 3.2.12. Hamilton's quaternions may seem somewhat contrived, but they are more than just a method for multiplying vectors in $\mathbb{R}^4$ together. They are actually important in physics, since they can be used to model rotations in $\mathbb{R}^3$.