Notes on Abstract Algebra

John Perry

University of Southern Mississippi
E-mail address: john.perry@usm.edu
URL: http://www.math.usm.edu/perry/

Contents

Reference sheet for notation

$[r]$	the element $r + n\mathbb{Z}$ of $\mathbb{Z}_n$, page 51
$\langle g \rangle$	the group (or ideal) generated by g, page 26
A_3	the alternating group on three elements, page 47
$A := B$	A is defined to be B, page 19
$A \triangleleft G$	A is a normal subgroup of G, page 47
$A \triangleleft I$	for I a ring, A is an ideal of I., page 123
$\mathbb{C}$	the complex numbers $\{a + bi : a, b \in \mathbb{C} \text{ and } i = \sqrt{-1}\}$, page 9
$\mathrm{Conj}_a(G)$	the group of conjugations of G by a, page 67
$\mathrm{conj}_g(x)$	the automorphism of conjugation by g, page 67
D_3	the symmetries of a triangle, page 30
$d \mid n$	d divides n, page 13
D_n	the dihedral group of symmetries of a regular polygon with n sides, page 80
$D_n(\mathbb{R})$	the set of all diagonal matrices whose values along the diagonal is constant, page 39
$d\mathbb{Z}$	the set of integer multiples of d, page 36
$f(S)$	the image of the function f from set S, page 56
G/A	the set of left cosets of A, page 44
$G\backslash A$	the set of right cosets of A, page 44
gA	the left coset of A with g, page 41
$\mathrm{GL}_m(\mathbb{R})$	the general linear group of invertible matrices, page 22
$\prod_{i=1}^{n} G_i$	the ordered n-tuples of G_1, G_2, …, G_n, page 19
$G \times H$	the ordered pairs of elements of G and H, page 19
$H < G$	H is an additive subgroup of H, page 36
I_n	the identity matrix of dimension $n \times n$, page 11
$\ker f$	the kernel of the homomorphism f, page 64
$N_G(H)$	the normalizer of a subgroup H of G, page 50
$\mathbb{N}$	natural numbers $\{0, 1, 2, 3 \ldots\}$, page 9
$\mathbb{N}^+$	the positive integers, page 12
$\mathbb{N}_{>0}$	the positive integers, page 12
$\mathrm{ord}(x)$	the order of x, page 27
P_∞	point at infinity, page 20
Q_8	the group of quaternions, page 25
$\mathbb{Q}$	the rational numbers $\{\frac{a}{b} : a, b \in \mathbb{Z} \text{ and } b \neq 0\}$, page 9
$\mathbb{R}$	the real numbers, those that measure any length along a line, page 9
$\mathbb{R}^{m \times m}$	$m \times m$ matrices with real coefficients, page 10
$\mathbb{R}[x]$	polynomials in one variable with real coefficients, page 9
$\mathbb{R}[x_1, x_2, \ldots, x_n]$	polynomials in n variables with real coefficients, page 9
$\mathrm{swp}\,\alpha$	the sign function of a cycle or permutation, page 89

S_n	the group of all permutations of a list of n elements, page 78
$Z(G)$	centralizer of a group G, page 49
$\mathbb{Z}_n^*$	the set of elements of $\mathbb{Z}_n$ that are *not* zero divisors, page 107
$\mathbb{Z}/n\mathbb{Z}$	quotient group (resp. ring) of $\mathbb{Z}$ modulo the subgroup (resp. ideal) $n\mathbb{Z}$, page 50
$\mathbb{Z}$	the integers $\{\ldots,-1,0,1,2,\ldots\}$, page 9
$\mathbb{Z}_n$	the quotient group $\mathbb{Z}/n\mathbb{Z}$, page 50

A few acknowledgements

These notes are inspired from some of my favorite algebra texts: [**AF05**, **CLO97**, **HA88**, **Lau03**, **LP98**, **Rot06**]. However, boneheaded innovations of mine that looked good at the time but turn out bad in practice should not be blamed on these references.

Thanks to the students who found typos, including Jonathan Yarber, Kyle Fortenberry, Lisa Palchak, Ashley Sanders, Sedrick Jefferson, Shaina Barber, Blake Watkins, and others. I really was thinking about what I wrote at the time, honest!

I have been lucky to have had good algebra professors; in chronological order, they were:

- Vanessa Job at Marymount University;
- Adrian Riskin at Northern Arizona University;
- and at North Carolina State University:
  - Kwangil Koh,
  - Hoon Hong,
  - Erich Kaltofen,
  - Michael Singer, and
  - Agnes Szanto.

The following software helped prepare these notes:

- Sage 3.*x* [**Ste08**],
- Lyx [**Lyx09**] (and therefore LaTeX [**Lam86**, **Grä04**]), along with the packages
  - `hyperref` [**RO08**],
  - $\mathcal{AMS}$-LaTeX[**Soc02**],
  - `mathdesign` [**Pic06**], and
  - `algorithms` (modified slightly from version 2006/06/02) [**Bri**].
- Inkscape [**Bah08**].

I've likely forgotten some other non-trivial resources that I used. Let me know if another citation belongs here.

CHAPTER 1

Introduction

1.1. THREE INTERESTING PROBLEMS

We'd like to motivate this study of algebra with three problems that we hope you will find interesting. We'll also provide some hints at solutions, but at the outset of a course you don't yet know enough to follow the solutions in all their depth. One can compare the situation to walking into the darkened room of a museum; beside you is a guide who tells you what is in the room, but until the guide turns on the lights, the description will be unlikely to meet your expectations.

A CARD TRICK. Take twelve cards. Ask a friend to choose one of them, to replace it in the stack of twelve cards, then to shuffle them thoroughly. Arrange the cards on a table face up, in rows of three. Ask your friend what column the card is in; call that number α.

Now collect the cards, making sure they remain in the same order as they were when you dealt them. Arrange them on a a table face up again, in rows of four. It is essential that you maintain the same order; the first card you placed on the table in rows of three must be the same card you place on the table in rows of four; likewise the last card must remain last. Ask your friend again what column the card is in; call that number β.

In your head, compute $x = 4\alpha - 3\beta$. If x does not lie between 1 and 12 inclusive, add or subtract 12 as many times as necessary until it is. Starting with the first card, and following the order in which you laid the cards on the table, count to the xth card. This will be the card your friend chose.

Mastering this trick isn't hard, and takes only a little practice. To understand *why* it works requires a marvelous result called the *Chinese Remainder Theorem*, so named because many centuries ago the Chinese military used this technique to count the number of casualties of a battle.[1] We cover the Chinese Remainder Theorem later in this course.

INTERNET COMMERCE. You go online to check your bank account. Before you can gain access to your account, your computer must send your login name and password to the bank. Likewise, your bank sends a lot of information, such as your account number and your balance, from its computers to yours. Most likely, you'd rather keep such sensitive information secret from all the other computers through which the information passes on its way to and from the bank. How can you do this?

The solution is called *public-key cryptography*. In public-key cryptography, your computer tells the bank's computer how to send it a message, and the bank's computer tells your computer how to send it a message. One key—a special number used to encrypt the message—is therefore public. Anyone in the world can see it. What's more, anyone in the world can look up the

[1] I asked a Chinese colleague what name the Chinese have for this theorem. He didn't know at first. He went to look in his books, and came back to tell me that they call it Sun Tzu's Theorem, apparently because it appears in Sun Tzu's *Art of War*.

method used to decrypt the message: he just needs to figure out the *hidden key,* the special number required to decrypt the message. It isn't hard to determine what the special properties of this number are; we discuss them later in this course. By solving a"simple" mathematical problem—factoring an integer into two primes: essentially, deciding that $21 = 3 \cdot 7$—an eavesdropper can determine your key and the bank's key, and decrypt the messages.

Why is this method safe for internet commerce? Although the problem to solve is "simple", it takes too long to solve! The prime numbers used contain hundreds of computer bits. Even the world's fastest computers cannot factor these large integers into two primes fast enough to decrypt the message in a reasonable timeframe; in fact, the world's fastest computers won't be fast enough to do this for many decades, if not centuries.

FACTORIZATION. How can we factor polynomials like $p(x) = x^6 + 7x^5 + 19x^4 + 27x^3 + 26x^2 + 20x + 8$? There are a number of ways to do it, but it turns out that one of the most efficient ways involves a trick called *modular arithmetic*. We discuss the theory of modular arithmetic later in the course, but for now the general principle will do: pretend that the only only numbers we can use come from a clock that runs from 0 to 28. Just as with the twelve-hour clock, when we hit the integer 29, we reset to 0; when we hit the integer 30, we reset to 1; and in general for any number that does not lie between 0 and 28, we divide by 29 and take the remainder. For example,

$$25 \cdot 3 + 8 = 83 \rightsquigarrow 25.$$

How does this help us factor? When looking for factors of the polynomial p, we can simplify multiplication by working in this modular arithmetic. This makes it easy for us to reject many possible factorizations before we start. In addition, the set $\{0, 1, \ldots, 28\}$ has many interesting properties under modular arithmetic that we can exploit further.

CONCLUSION. Abstract algebra is, by nature, a theoretical course. We start discussing that somewhat in the next section, and especially in the next chapter. You may be tempted on many occasions to wonder what is the point of all this abstraction and theory; who would ever need it?

The three problems above are examples of how abstract algebra is not only useful, but necessary. It is abstract, but its applications have been profound and broad. Eventually we expect to explain to you how the problems above are solved using the methods of algebra; for now, you can only start to imagine.

1.2. COMMON PATTERNS

Until now, your background in mathematics directed you in the study of a large number of seemingly different sets:

- numbers, of which you have seen
  - the **natural numbers** $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, also written $\mathbb{Z}_{\geq 0}$;
  - the **integers** $\mathbb{Z} = \{\ldots, -1, 0, 1, 2, \ldots\}$;
  - the **rational numbers** $\mathbb{Q} = \left\{\frac{a}{b} : a, b \in \mathbb{Z} \text{ and } b \neq 0\right\}$;
  - the **real numbers** $\mathbb{R}$;
  - the **complex numbers** $\mathbb{C} = \left\{a + bi : a, b \in \mathbb{R} \text{ and } i = \sqrt{-1}\right\}$;
- polynomials, of which you have seen
  - polynomials in one variable $\mathbb{R}[x]$;
  - polynomials in more than one variable $\mathbb{R}[x, y]$, $\mathbb{R}[x, y, z]$, $\mathbb{R}[x_1, x_2, \ldots, x_n]$;

- square matrices $\mathbb{R}^{m\times m}$.

Each set is especially useful for representing certain kinds of objects. Natural numbers can represent objects that we count: two apples, or twenty planks of flooring. Real numbers can represent objects that we measure, such as the length of the hypotenuse of a right triangle.

For each set, you have studied elementary operations with elements in the set, and properties of these operations. The operations in each set represent real-world activities that are fundamentally different.

EXAMPLE 1.1. The natural number 2 can represent a basket of 2 tomatoes; the natural number 10 can represent a basket of ten tomatoes. Adding the two numbers represents counting how many tomatoes are in both baskets: 10+2=12.

Adding two polynomials is somewhat similar, but requires a different method for simplification. The polynomial $f = 2x + 3y$ can represent the amount of money earned when tomatoes (x) and cucumbers (y) are sold on a day where their respective prices are \$2 and \$3 apiece. On another day, the prices may change to \$1 and \$2.50, respectively, so the polynomial $g = x + 2.5y$ represents the amount of money earned on that day. Adding the two polynomials gives $f + g$, which represents the amount of money earned if we sell the same number of tomatoes and cucumbers on both days. We don't count objects in baskets to determine a simplified representation of $f + g$; instead we apply the distributive property. Then

$$f + g = (2x + 3y) + (x + 2.5y) = (2 + 1)\,x + (3 + 2.5)\,y = 3x + 5.5y.$$

Adding two rational numbers is quite a bit different. Let $x, y \in \mathbb{Q}$. Then $x = a/b$ and $y = c/d$ for certain integers a, b, c, d where $b, d \neq 0$. We have

$$x + y = \frac{a}{b} + \frac{c}{d} = \frac{ad}{bd} + \frac{bc}{bd} = \frac{ad + bc}{bd}.$$

Here we had to compute a common denominator in order to add the two rational numbers. This is conceptually different from adding integers or adding polynomials, which is why students who memorize a process instead of a meaning dislike fractions and want instead to add

$$\frac{a}{b} + \frac{c}{d} = \frac{a + c}{b + d}.$$

Someone might decide to define the addition of fractions this way, and *there would be no problem with it!* As long as the operation is defined this way, and corresponds to the real-world phenomena that it models, there is no problem.

Mathematicians do consider this second version of addition wrong, but *only because it does not correspond to the real-world phenomena that fractions usually represent.* Make sure you understand this point: adding fractions in the second way described above is wrong because of what fractions represent. △

Despite the fact that the operations behave differently on objects in each set, we still observe some common properties. These properties have nothing to do with the choice of *how to simplify a given operation*, but are intrinsic to the operation itself, or to the structure of the objects of the set.

For example:

(1) Addition is always **commutative**. That is, $x + y = y + x$ no matter which set contains x, y.

(2) Multiplication is *not* always commutative. Matrices misbehave on this point:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 2 & 0 \end{pmatrix} \neq \begin{pmatrix} 2 & 2 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

(3) Both operations are always **associative**. That is, $x+(y+z) = (x+y)+z$ and $x(yz) = (xy)z$ no matter which set contains x, y, z.

(4) Multiplication is **distributive** over addition. That is, $x(y+z) = xy+xz$ no matter which set contains x, y, z.

(5) Each set has additive and multiplicative **identities**. That is, each set has two special elements, written 0 and 1, such that $0+x = x$ and $1 \cdot x = x \cdot 1 = x$, for every value of x.[2]

(6) Each set has **additive inverses**. That is, for any set and for any x in that set, we can identify an element y such that $x+y = 0$. Usually we write $-x$ instead of y, so that $x+(-x) = 0$.

(7) Not every set contains **multiplicative inverses** for all its elements. To begin with, the zero element never has a multiplicative inverse. For polynomials, only non-zero constant polynomials like 4 or -8 have multiplicative inverses. Polynomials such as x^2 do not have inverses *that are also polynomial*. With matrices the situation is even worse; many matrices have no inverse at all.

(8) In every set specified, $-1 \times x = -x$ and $0 \times x = 0$. But not every set obeys the **zero product property**,

$$\text{if } xy = 0 \text{ then } x = 0 \text{ or } y = 0.$$

Here again it is the matrices that misbehave:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = 0.$$

(9) In some sets you have learned how to **divide** with remainder. Even though non-constant polynomials don't have multiplicative inverses, you can still divide by them. In other sets you have never seen a division, as with matrices, for example.

There are many others, but these will do.

Not very long ago,[3] mathematicians set about asking themselves how they could organize the common principles of these sets and operations as abstractly, simply, and generically as possible.

You might wonder why someone would want to do that. It's a good question, because it isn't very easy to think abstractly. Most people think algebra is abstract enough with polynomials and matrices; why make it any more abstract?

As you will learn from this book, mathematicians have found this abstraction to be enormously useful. If I prove some property about the integers, then I only have a result about integers. If someone comes and asks whether that property is also true about matrices, the only answer I have without further effort is, "I don't know." It might be easy to show that the property is true, but it might be hard to show that the property is true.

If, on the other hand, I prove something about any set that shares certain properties with the integers, and if the matrices also share those properties, then I can answer without any further effort, "Yes!"

[2] For square matrices we actually write I_n, where n is the dimension of the matrix.

[3] Depending on one's point of view. Certainly more than five minutes ago. In fact, more than one hundred years ago. But, less than one thousand years ago.

This is the beauty of abstract algebra.

EXERCISES.

EXERCISE 1.2. For each set S listed below, find a real-world phenomenon that the elements of S represent.

(a) $\mathbb{Z}$
(b) $\mathbb{Q}$
(c) $\mathbb{R}[x,y]$

EXERCISE 1.3. Give a detailed example of a real-world phenomenon where

$$\frac{a}{b}+\frac{c}{d}\neq\frac{a+c}{b+d}.$$

By "real-world phenomenon", I mean that you should not merely add two fractions in the ordinary manner, but describe the problem using objects that you use every day (or at least once a month).

1.3. THEORY OF ARITHMETIC

You will need to use a number of topics that were once covered in high-school algebra, but that you have likely forgotten. Even if you haven't forgotten them, you may not have seen them in a context where the emphasis was on trying to understand *why* something is true.

We start with the absolute basics. We already mentioned $\mathbb{N}$; we will also make use of

$$\mathbb{N}_{>0}=\mathbb{N}^{+}=\mathbb{N}\setminus\{0\}=\{1,2,3,\ldots\}.$$

One of the nice aspects of $\mathbb{R}$ is that **we can order its elements linearly**: for any $a,b\in\mathbb{R}$ exactly one of the following holds: $a<b$, $a=b$, or $a>b$. This ordering extends to all subsets of $\mathbb{R}$, in particular the elements of $\mathbb{N}^{+}$, $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ as well. This is different from $\mathbb{C}$ and $\mathbb{R}^{m\times m}$, where it is impossible to order the elements linearly.

Something that distinguishes $\mathbb{N}$ from the sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{R}^{m\times m}$ is the **well-ordering property**:

Every nonempty subset of $\mathbb{N}$ has a smallest element.

This looks obvious, but it is very important, and what is remarkable is that *no one can prove it.* It is an assumption about the natural numbers. This is why we state it as an axiom.

A consequence of the well-ordering property is the principle of mathematical induction:

THEOREM 1.4 (Mathematical Induction). *Let P be a subset of $\mathbb{N}_{>0}$. If P satisfies (IB) and (IS) where*

(IB) $1\in P$;
(IS) for every $i\in P$, we know that $i+1$ is also in P;
then $P=\mathbb{N}$.

PROOF. Let $S=\mathbb{N}\setminus P$. We will prove the contrapositive, so assume that $P\neq\mathbb{N}$. Thus $S\neq\emptyset$. Note that $S\subseteq\mathbb{N}$. By the well-ordering principle, S has a smallest element; call it n. If $n=1$, then P does not satisfy (IB). If $n>1$, then $n-1\notin S$, so $n-1\in P$. Let $i=n-1$; then $i\in P$ and $i+1=n\notin P$. Then P does not satisfy (IS), since $i=n-1\in P$ and $i+1=n\notin P$.

We have shown that if $P \neq \mathbb{N}$, then P fails to satisfy at least one of (IB) or (IS). This is the contrapositive of the theorem. □

Induction is an enormously useful tool, and we will make use of it from time to time.

We will also need the following important fact.

THEOREM 1.5 (The Division Theorem). *Let $n, d \in \mathbb{Z}$. There exist unique $q, r \in \mathbb{Z}$ satisfying (D1) and (D2) where*

(D1) $n = qd + r$;
(D2) $0 \leq r < d$.

We call n the **dividend**, d the **divisor**, q the **quotient**, and r the **remainder**.

PROOF. We must show two things: first, that q and r exist; second, that r is unique.

Existence of q and r: First we show the existence of q and r that satisfy (D1). Let $S = \{n - qd : q \in \mathbb{Z}\}$ and $M = S \cap \mathbb{N}$. By the well-ordering of $\mathbb{N}$, M has a smallest element; call it r. By definition of S, there exists $q \in \mathbb{Z}$ such that $n - qd = r$. It follows that $n = qd + r$.

Does r satisfy (D2)? By way of contradiction, assume that it does not; then $0 \leq r - d < r$. We can rewrite

$$\begin{aligned} n &= qd + r \\ &= qd + d + (r - d) \\ &= (q+1)d + (r - d). \end{aligned}$$

Hence $r - d = n - (q+1)d$. This form of $r - d$ shows that $r - d \in S$. Since $r - d \geq 0$, $r - d \in M$. This contradicts the choice of r as the *smallest* element of M.

Hence $n = qd + r$ and $0 \leq r < d$; q and r satisfy (D1) and (D2).

Uniqueness of q and r: Suppose that there exist $q', r' \in \mathbb{Z}$ such that $n = q'd + r'$ and $0 \leq r' \leq r < d$. Then $r' = n - q'd \in S$. Since r is the smallest element of S, $r' = r$. Setting $n - q'd = n - qd$ yields $q = q'$. □

DEFINITION 1.6. Let $n, d \in \mathbb{Z}$ and suppose that the Division Theorem gives us $n = qd + r$. If $r = 0$, then $n = qd$. We say in this case that d **divides** n, and write $d \mid n$. We also say that n is **divisible by** d. If on the other hand $r \neq 0$, then we say that d **does not divide** n, and write $d \nmid n$.

EXERCISES.

EXERCISE 1.7. Identify the quotient and remainder when dividing:
(a) 10 by -5;
(b) -5 by 10;
(c) -10 by -5.

EXERCISE 1.8. Let $a, b \in \mathbb{Z}$, and assume that both $a \leq b$ and $b \leq a$. Prove that $a = b$. *Hint:* There are several ways to prove this; one way is to look for a contradiction.

EXERCISE 1.9. Let $S \subset \mathbb{N}$. We know from the well-ordering property that S has a smallest element. Prove that this smallest element is unique. *Hint:* Let m, n be two smallest elements of S. Since m is a smallest element of S, what do you know about m and n? Likewise, since n is a smallest element of S, what do you know about m and n? Then…

EXERCISE 1.10. Let $x, d \in \mathbb{Z}$, where $d > 0$. Define $M = \{x - qd : q \in \mathbb{Z}\}$. Prove that $M \cap \mathbb{N} \neq \emptyset$. *Hint:* If $x > 0$ then you can pretty easily identify a q that finds an element of the intersection $M \cap \mathbb{N}$. Otherwise $x \leq 0$. You know that $d > 0$. What happens when you divide x by d?

Part 1

Elementary group theory

CHAPTER 2

Additive, multiplicative, and general groups

2.1. Common structures for addition

Many sets in mathematics, such as those listed in Chapter 1.2, allow addition of their elements; others allow multiplication. Some allow both. We saw that while addition was commutative for all the examples listed, multiplication was not.

We will use some of the properties we associate with addition to define a common structure for addition. We consider multiplication in the following section.

Definition 2.1. Let G be a set, and $+$ an addition. The pair $(G,+)$ is an **additive group** if $(G,+)$ satisfies the following properties.

(AG1) Addition is closed; that is, $x+y \in G$ for all $x,y \in G$.
(AG2) Addition is associative; that is, $x+(y+z)=(x+y)+z$ for all $x,y,z \in G$.
(AG3) There exists an element $z \in G$ such that $x+z=x$ for all $x \in G$. We call this element the **zero element** or the **additive identity**, and generally write 0 to represent it.
(AG4) For every $x \in G$ there exists an element $y \in G$ such that $x+y=0$. Normally we write $-x$ for this element and call it the **additive inverse**.
(AG5) Addition is commutative; that is, $x+y=y+x$ for all $x,y \in G$. △

We may also refer to an additive group as a **group under addition**. The operation is usually understood from context, so we usually write G rather than $(G,+)$. We will sometimes write $(G,+)$ when we want to emphasize the operation, especially if the operation does not fit the normal intuition of addition (see Exercises 2.10 and 2.11 below) and when we discuss groups under multiplication later.

Example 2.2. Certainly $\mathbb{Z}$ is an additive group. Why?

(AG1) Adding two integers gives another integer.
(AG2) Addition of integers is associative.
(AG3) The additive identity is the number 0.
(AG4) Every integer has an additive inverse.
(AG5) Addition of integers is commutative. △

The same holds true for many of the sets we identified in Chapter 1.2, using the ordinary definition of addition in that set. However, $\mathbb{N}$ is not an additive group. Why not? Although $\mathbb{N}$ is closed, and addition of natural numbers is associative and commutative, no positive natural number has an additive inverse in $\mathbb{N}$.

Our definition of additive groups now allows us to investigate arbitrary additive groups, and to formulate conclusions based on these arbitrary groups. Mathematicians of the 20th century invested substantial effort in an attempt to classify all finite simple groups. (You will learn later what makes a group "simple".) We won't replicate their achievement in this book, but we do want to to take a few steps in this area. First we need a definition.

DEFINITION 2.3. Let S be any set. If there is a finite number of elements in S, then $|S|$ denotes that number, and we say that S has **size** . If there is an infinite number of elements in S, then we write $|S| = \infty$. We also write $|S| < \infty$ to indicate that $|S|$ is finite, without stating a precise number.

For any additive group $(G,+)$ the **order of** G is the size of G. A group has finite order if $|G| < \infty$ and infinite order if $|G| = \infty$. △

We now have enough information to completely classify all finite groups of order two. We will show this by building the addition table for an arbitrary group of order two, and show in the process that it can only have one possible addition table.[1]

EXAMPLE 2.4. Every additive group of order two has the same addition table. To see this, let G be an arbitrary additive group of order two. Then $G = \{0, a\}$ where 0 represents the zero element. As we build the addition table, we observe that we *have* to assign $a + a = 0$. *Why?*

- To satisfy (AG3), we must have $0 + 0 = 0$, $0 + a = a$, and $a + 0 = a$.
- To satisfy (AG4), a must have an additive inverse. The inverse isn't 0, so it must be a itself! That is, $-a = a$. (Read that as, "the additive inverse of a is a.") So $a + (-a) = a + a = 0$.

This leads to the following table.

$+$	0	a
0	0	a
a	a	0

The only assumption we made about G is that it was a group of order two. That means that we have completely determined the addition table of all groups of order two! △

NOTATION. Because it is tiresome to write $x + (-y)$ all the time, we write $x - y$ instead.

The following fact looks obvious–but remember, we're talking about elements of *any* additive group, not only the numbers you have always used.

LEMMA 2.5. *Let G be an additive group and $x \in G$. Then $-(-x) = x$.*

Lemma 2.5 is saying that the additive inverse of the additive inverse of x is x itself; that is, if y is the additive inverse of x, then x is the additive inverse of y.

PROOF. You prove it! See Exercise 2.7. □

We observed in Example 2.4 that the structure of a group compels certain assignments for addition. We can distill this into an important conclusion for additive groups of finite order.

THEOREM 2.6. *Let G be an additive group of finite order, and let $a, b \in G$. Then a appears exactly once in any row or column of the addition table that is headed by b.*

PROOF. The element a appears in a row of the addition table headed by b any time there exists $c \in G$ such that $b + c = a$. Let $c, d \in G$ such that $b + c = a$ and $b + d = a$. Substitution gives us $b + c = b + d$. Properties (AG1), (AG4), and (AG3) give us

$$-b + (b + d) = (-b + b) + d = 0 + d = d.$$

Along with substitution, they also give us

$$-b + (b + d) = -b + (b + c) = (-b + b) + c = 0 + c = c.$$

[1] Later in this chapter we refer to this phenomenon as an isomorphism of groups.

By the transitive property of equality, $c = d$. This shows that if a appears in one row of the addition table, then it does not appear in a different row.

We still have to show that a appears in at least one row of the addition table headed by b. This follows from the fact that the row headed by b has $|G|$ elements. Each element of G can appear at most once. The pigeonhole principle requires that each element appear exactly once. □

EXERCISES.

EXERCISE 2.7. Explain why $-(-x) = x$. *Hint:* Remember that $-$ means the additive inverse. So, you have to show that the additive inverse of $-x$ is x.

EXERCISE 2.8. Let G be an additive group, and $x, y, z \in G$. Show that if $x = y$, then $x + a = y + a$. *Hint:* Use substitution.

EXERCISE 2.9. Show in detail that $\mathbb{R}^{2\times 2}$ is an additive group. *Hint:* Work with arbitrary elements of $\mathbb{R}^{2\times 2}$. The structure of such elements is

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad \text{where } a_{11}, a_{12}, a_{21}, a_{22} \in \mathbb{R}.$$

EXERCISE 2.10. Consider the set $B = \{F, T\}$ with the operation $\vee$ where

$$\begin{aligned} F \vee F &= F \\ F \vee T &= T \\ T \vee F &= T \\ T \vee T &= T. \end{aligned}$$

This operation is called **Boolean or**.

Is $(B, \vee)$ an additive group? If it is, identify the zero element, and for each non-zero element identify its additive inverse. If it is not, explain why not.

EXERCISE 2.11. Consider the set B from Exercise 2.10 with the operation $\oplus$ where

$$\begin{aligned} F \oplus F &= F \\ F \oplus T &= T \\ T \oplus F &= T \\ T \oplus T &= F. \end{aligned}$$

This operation is called **Boolean exclusive or**, or **xor** for short.

Is $(B, \oplus)$ an additive group? If it is, identify the zero element, and for each non-zero element identify its additive inverse. If it is not, explain why not.

EXERCISE 2.12. Define $\mathbb{Z} \times \mathbb{Z}$ to be the set of all ordered pairs whose elements are integers; that is,

$$\mathbb{Z} \times \mathbb{Z} := \{(a,b) : a,b \in \mathbb{Z}\}.$$

Addition in $\mathbb{Z} \times \mathbb{Z}$ works in the following way. For any $x,y \in \mathbb{Z} \times \mathbb{Z}$, write $x = (a,b)$ and $y = (c,d)$ and then

$$x + y = (a+c, b+d).$$

Show that $\mathbb{Z} \times \mathbb{Z}$ is an additive group.

EXERCISE 2.13. Let G and H be additive groups, and define

$$G \times H = \{(a,b) : a \in G,\ b \in H\}.$$

Addition in $G \times H$ works in the following way. For any $x,y \in G \times H$, write $x = (a,b)$ and $y = (c,d)$ and then

$$x + y = (a+c, b+d).$$

Show that $G \times H$ is an additive group.

Note: The symbol $+$ may have different meanings for G and H. For example, the first group might be $\mathbb{Z}$ while the second group might be $\mathbb{R}^{m \times m}$.

EXERCISE 2.14. Let $n \in \mathbb{N}^+$. Let $G_1, G_2, \ldots, G_n$ be additive groups, and define

$$\prod_{i=1}^{n} G_i := G_1 \times G_2 \times \cdots \times G_n = \{(a_1, a_2, \ldots, a_n) : a_i \in G_i \ \forall i = 1,2,\ldots,n\}.$$

Addition in this set works in the following way. For any $x,y \in \prod_{i=1}^{n} G_i$, write $x = (a_1, a_2, \ldots, a_n)$ and $y = (b_1, b_2, \ldots, b_n)$ and then

$$x + y = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n).$$

Show that $\prod_{i=1}^{n} G_i$ is an additive group.

EXERCISE 2.15. Let $m \in \mathbb{N}^+$. Show in detail that $\mathbb{R}^{m \times m}$ is a group under addition. *Hint:* You probably did this in linear algebra, or saw it done. Work with arbitrary elements of $\mathbb{R}^{m \times m}$, which have the structure

$$A = \left(a_{i,j}\right)_{i=1\ldots m, j=1\ldots m}.$$

EXERCISE 2.16. Show that every additive group of order 3 has the same structure.

EXERCISE 2.17. *Not* every additive group of order 4 has the same structure, because there are two addition tables with different structures. One of these groups is the **Klein four-group**, where each element is its own inverse; the other is called a **cyclic group** of order 4, where not every element is its own inverse. Determine addition tables for each group.

2.2. Elliptic Curves

An excellent example of how additive groups can appear in places that you might not expect is in *elliptic curves*. These functions have many applications, partly due to an elegant group structure.

Definition 2.18. Let $a, b \in \mathbb{R}$ such that $16a^3 \neq 27b^2$. We say that $E \subset \mathbb{R}^2$ is an **elliptic curve** if

$$E = \left\{(x,y) \in \mathbb{R}^2 : y^2 = x^3 + ax + b\right\} \cup \{P_\infty\},$$

where P_∞ denotes a **point at infinity**.

What is meant by a point at infinity? If different branches of a curve extend toward infinity, we imagine that they meet at a point, called the point at infinity.

There are different ways of visualizing a point at infinity. One is to imagine the real plane as if it were wrapped onto a sphere. The scale on the axes changes at a rate inversely proportional to one's distance from the origin; in this way, no finite number of steps bring one to the point on the sphere that lies opposite to the origin. On the other hand, this point would be a limit as x or y approaches $\pm\infty$. Think of the line $y = x$. If you start at the origin, you can travel either northeast or southwest on the line. Any finite distance in either direction takes you short of the point opposite the origin, but the limit of both directions meets at the point opposite the origin. This point is the point at infinity.

Example 2.19. Let

$$E = \left\{(x,y) \in \mathbb{R}^2 : y^2 = x^3 - x\right\} \cup \{P_\infty\}.$$

Here $a = -1$ and $b = 0$. Figure 2.1 gives a diagram of E.

It turns out that E is an additive group. Given $P, Q \in E$, we can define addition by:

- If $Q = P_\infty$, then define $P + Q = P$.
- If $P = (p_1, p_2)$ and $Q = (p_1, -p_2)$, then define $P + Q = P_\infty$.
- If $P = Q$, then construct the tangent line ℓ at P. It turns out that ℓ intersects E at another point $S = (s_1, s_2)$ in $\mathbb{R}^2$. Define $P + Q = (s_1, -s_2)$
- Otherwise, construct the line ℓ determined by P and Q. It turns out that ℓ intersects E at another point $S = (s_1, s_2)$ in $\mathbb{R}^2$. Define $P + Q = (s_1, -s_2)$.

The last two statements require us to ensure that, given two distinct and finite points $P, Q \in E$, a line connecting them intersects E at a third point S. Figure 2.2 shows the addition of $P = \left(2, -\sqrt{6}\right)$ and $Q = (0,0)$; the line intersects E at $S = \left(-1/2, \sqrt{6}/4\right)$, so $P + Q = \left(-1/2, -\sqrt{6}/4\right)$. △

Exercises.

Exercise 2.20. Let E be an arbitrary elliptic curve. Show that $\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \neq (0,0)$ for any point on E. *Hint:* You will need the condition that $16a^3 = 27c^2$.

This shows that E is "smooth", and that tangent lines exist at each point in $\mathbb{R}^2$. (This includes vertical lines, where $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} \neq 0$.)

Exercise 2.21. Show that E is an additive group under the addition defined above, with

FIGURE 2.1. A plot of the elliptic curve $y^2 = x^3 - x$.

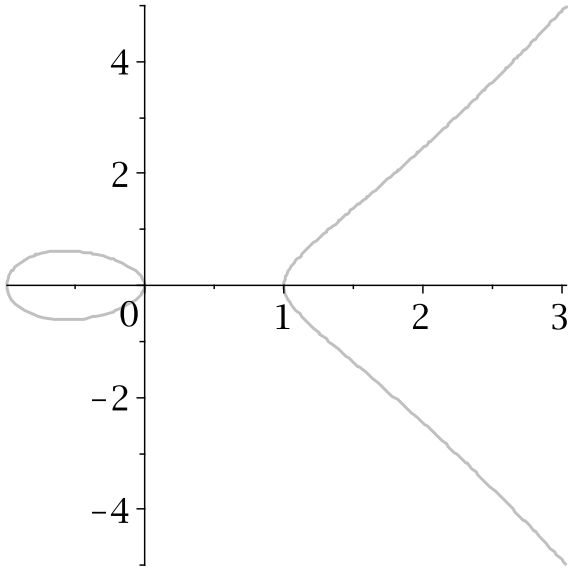
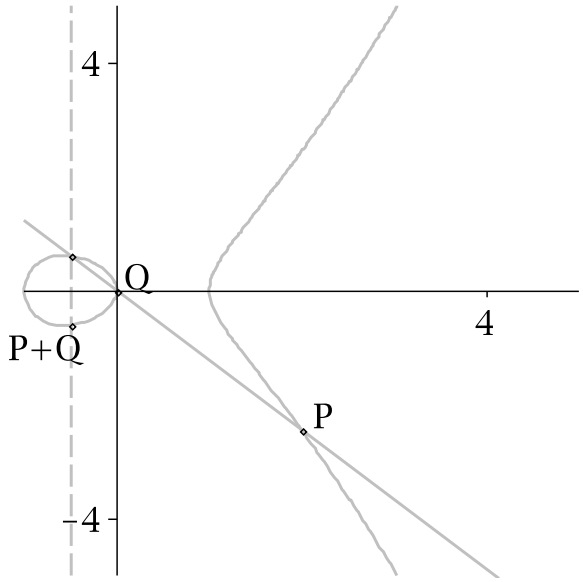


FIGURE 2.2. Addition on an elliptic curve



- P_∞ as the zero element; *and*
- for any $P = (p_1, p_2) \in E$, then $-P = (p_1, -p_2) \in E$.

Hint: For closure, it suffices to show that each line intersects the curve three times, possibly at P_∞.

EXERCISE 2.22. Choose different values for a and b to generate another elliptic curve. Graph it, and illustrate each kind of addition.

2.3. COMMON STRUCTURES FOR MULTIPLICATION

Unlike addition, multiplication is not always commutative. In order to work with that, we will define a structure similar to additive groups, but lacking the requirement that the operation be commutative.

DEFINITION 2.23. Let G be a set, and $\times$ a multiplication. The pair $(G, \times)$ is a **multiplicative group** if $(G, \times)$ satisfies the following properties.

(MG1) Multiplication is closed; that is, $xy \in G$ for all $x, y \in G$.
(MG2) Multiplication is associative; that is, $x(yz) = (xy)z$ for all $x, y, z \in G$.
(MG3) There exists an element $1 \in G$ such that $x \cdot 1 = x$ for all $x \in G$. We call this element the **identity**.
(MG4) For every $x \in G$ there exists an element $y \in G$ such that $xy = e$. Normally we write x^{-1} for this element, and call it the **multiplicative inverse**. △

We may also refer to a multiplicative group as a **group under multiplication**.

Even with this more restricted idea of multiplication, $\mathbb{R}^{m\times m}$ is not a group. However, we can now construct a group using a large subset of $\mathbb{R}^{m\times m}$.

DEFINITION 2.24. Define $\mathrm{GL}_m(\mathbb{R})$ to be the set of invertible matrices, also called the **general linear group**.

EXAMPLE 2.25. $\mathrm{GL}_m(\mathbb{R})$ is a multiplicative group. We leave much of the proof to the exercises, but point out that (MG1) and (MG2) are known from linear algebra.

EXAMPLE 2.26. Every multiplicative group of order 2 has the same multiplication table. To see this, let G be an arbitrary multiplicative group of order two. Then $G = \{1, a\}$. We build the multiplication table, being careful to follow properties (MG3) and (MG4):

$\times$	1	a
1	1	a
a	a	1

Notice that the properties compelled us to assign $a \times a = 1$, for the same reason that they compelled us to assign $a + a = 0$ in the multiplication table of Example 2.4 on page 17.

The only assumption we made about G is that it was a multiplicative group of order two. That means that we have completely determined the addition table of all multiplicative groups of order two! △

The reader may notice from the example that the structure of the multiplication table is *identical* to the structure of the addition table in Example 2.4. This suggests that there is no meaningful difference between additive and multiplicative groups of size two. Likewise, you will find in Exercises 2.36 and 2.37 that the multiplication tables for groups of order 3 and 4 are identical to the structure of the addition tables in Exercises 2.16 and 2.17: even the multiplication is commutative.

At this point a question arises:

Although multiplication was not commutative in $\mathrm{GL}_m(\mathbb{R})$,
could it be commutative in every ***finite*** *multiplicative group?*

The answer is no. In Exercise 2.38, you will study a group of order 8 whose multiplication is not commutative.

Since multiplicative groups of orders 2, 3, and 4 must be commutative, but multiplicative groups of order 8 need not be commutative, a new question arises:

Is multiplication necessarily commutative
in multiplicative groups of order 5, 6, or 7?

The answer is, "it depends on the order". We delay the details until later.

You have now encountered additive and multiplicative groups. The only difference we have seen between them so far is that multiplication need not be commutative. Both Lemma 2.5 and Theorem 2.6 have parallels for multiplicative groups, and we could state them, but it is time to exploit the power of abstraction a little further.

Until now, we have defined groups using arbitrary sets, but specific operations. We now generalize the notion of a group to both arbitrary sets and arbitrary operations. The new definition will incorporate the principles of additive and multiplicative groups without forcing either into the mold of the other.

DEFINITION 2.27. Let G be a set, and $\circ$ an operation. For convenience denote $x \circ y$ as xy. The pair $(G, \circ)$ is a **group under the operation** $\circ$ if $(G, \circ)$ satisfies the following properties.

(G1) The operation is closed; that is, $xy \in G$ for all $x, y \in G$.
(G2) The operation is associative; that is, $x(yz) = (xy)z$ for all $x, y, z \in G$.
(G3) There exists an element $e \in G$ such $xe = ex = x$ for all $x \in G$. We call this element the **identity**.
(G4) For every $x \in G$ there exists an element $y \in G$ such that $xy = yx = e$. Normally we write x^{-1} for this element.

We say that $(G, \circ)$ is an **abelian group**[2] if in addition

(G5) The operation is commutative; that is, $xy = yx$ for all $x, y \in G$. △

NOTATION. In Definition 2.27, the symbol $\circ$ is a placeholder for any operation. It can stand for addition, for multiplication, or for other operations that we have not yet considered. We adopt the following conventions:

- If all we know is that G is a group under some operation, we write $\times$ for the operation and proceed as if the group were multiplicative, writing xy.
- If we know that G is a group and a symbol is provided for its operation, we *usually* (but not always) proceed as if the group were multiplicative, writing xy. In the definition, for example, the symbol $\circ$ is provided for the operation, but we wrote xy instead of $x \circ y$.
- We reserve the symbol $+$ for those cases where G is an abelian group.
- However, in some abelian groups we use multiplicative notation, and write xy.

You can see that the conventions are somewhat loose. As with any language, it takes some time to grow accustomed to the usage.

Definition 2.27 allows us to classify both additive and multiplicative groups as generic groups. Additive groups are guaranteed to be abelian, while multiplicative groups are sometimes abelian, but sometimes not. For this reason, from now on we generally abandon the designation "additive" group, preferring instead the terminology "abelian" group.

[2]Named after Niels Abel, a Norwegian high school mathematics teacher who helped found group theory.

We can now generalize Lemma 2.5 and Theorem 2.6 as promised. The proofs are very easy—one needs merely rewrite them using the notation for a general group—so we leave that to the exercises.

Notice that we change the name of the operation from "addition" in Theorem 2.6 to the generic term "operation" in Theorem 2.29.

LEMMA 2.28. *Let G be a group and $x \in G$. Then $\left(x^{-1}\right)^{-1} = x$.*

THEOREM 2.29. *Let G be a group of finite order, and let $a, b \in G$. Then a appears exactly once in any row or column of the operation table that is headed by b.*

The following lemma may look obviously true, but its proof isn't, and the result is important. It's better to make sure "obvious" things are true than to assume that they are, so we'll make sure of that now.

THEOREM 2.30. *The identity of a group is unique; that is, every group has exactly one identity. Also, the inverse of an element is unique; that is, every element has exactly one inverse element.*

PROOF. Let G be a group, and suppose that e and i are both identity elements. Since i is an identity, we know that

$$e = ei.$$

Since e is an identity, we know that

$$ei = e.$$

Combining the two equations, we conclude that

$$e = i.$$

We chose two arbitrary identity elements of G and showed that they were the same element. Hence there is only one identity element.

A similar strategy shows that the inverse of an element is unique. Let $x \in G$ and suppose that $y, z \in G$ are both inverses of x. Since y is an inverse of x,

$$xy = e.$$

Since z is an inverse of x,

$$xz = e.$$

By substitution,

$$xy = xz.$$

Multiply both sides of this equation on the left by y to obtain

$$y(xy) = y(xz).$$

Apply the associative property of G to obtain

$$(yx)y = (yx)z.$$

Since y is an inverse of x,

$$ey = ez.$$

Since e is the identity of G,

$$y = z.$$

We chose two arbitrary inverses of of x, and showed that they were the same element. Hence the inverse of x is unique. □

EXERCISES.

EXERCISE 2.31. Explain why $\mathrm{GL}_m(\mathbb{R})$ satisfies properties (MG3) and (MG4) of the definition of a multiplicative group.

EXERCISE 2.32.

(a) Show that if $G = \mathrm{GL}_m(\mathbb{R})$, there exist $a, b \in G$ such that $(ab)^{-1} \neq a^{-1}b^{-1}$. *Hint:* Try $m = 2$, and find two invertible matrices A, B such that $(AB)\left(A^{-1}B^{-1}\right) \neq I_2$.
(b) Show that for any $a, b \in \mathrm{GL}_m(\mathbb{R})$, $(ab)^{-1} = b^{-1}a^{-1}$. *Hint:* Use the associative property to help simplify the expression $(ab)\left(b^{-1}a^{-1}\right)$.
(c) Now suppose that G is an arbitrary group. Explain why we cannot assume that $(ab)^{-1} = a^{-1}b^{-1}$, but we can assume that $(ab)^{-1} = b^{-1}a^{-1}$.

EXERCISE 2.33. Let $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$, and $\times$ the ordinary multiplication of real numbers. Show that $\mathbb{R}^+$ is a multiplicative group by explaining why $\left(\mathbb{R}^+, \times\right)$ satisfies properties (MG1)–(MG4).

EXERCISE 2.34. Define $\mathbb{Q}^*$ to be the set of non-zero rational numbers; that is,

$$\mathbb{Q}^* = \left\{ \frac{a}{b} : a, b \in \mathbb{Z} \text{ where } a \neq 0 \text{ and } b \neq 0 \right\}.$$

Show that $\mathbb{Q}^*$ is a multiplicative group.

EXERCISE 2.35. Explain why $\mathbb{Z}$ is not a multiplicative group.

EXERCISE 2.36. Show that every multiplicative group of order 3 has the same multiplication table, and that this structure is in fact identical to that of an additive group of order 3.

EXERCISE 2.37. Show that there are only two possible multiplication tables for multiplicative groups of order 4, and that these correspond to the groups found in Exercise 2.17.

EXERCISE 2.38. Let Q_8 be the set of quaternions, defined by the matrices $\{\pm\mathbf{1}, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k}\}$ where

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{i} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \mathbf{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \mathbf{k} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

(a) Show that $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -\mathbf{1}$.
(b) Show that $\mathbf{ij} = \mathbf{k}$, $\mathbf{jk} = \mathbf{i}$, and $\mathbf{ik} = -\mathbf{j}$
(c) Show that Q_8 is a group under matrix multiplication.
(d) Explain why Q_8 is not an abelian group.

EXERCISE 2.39. Prove Lemma 2.28.

EXERCISE 2.40. Prove Theorem 2.29.

EXERCISE 2.41. Show that $\mathbb{Q}^* \times \mathbb{Q}^*$ is a multiplicative group, where for all $x, y \in \mathbb{Q}^* \times \mathbb{Q}^*$ we have

$$xy = (x_1 y_1, x_2 y_2).$$

EXERCISE 2.42. Let $G_1, G_2, \ldots, G_n$ be groups. Show that $G = G_1 \times G_2 \times \cdots \times G_n$ is also a group, where for all $x, y \in G$ we have

$$xy = (x_1 y_1, x_2 y_2, \ldots, x_n y_n).$$

2.4. CYCLIC GROUPS

At this point you can make an acquaintance with an important class of groups. Groups in this class have a nice, appealing structure.

NOTATION. Let G be a group, and $x \in G$. If we want to perform the operation on x ten times, we could write

$$x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x$$

but this grows tiresome. Instead we will adapt notation from high-school algebra and write

$$x^{10}$$

instead. We likewise define x^{-10} to represent

$$x^{-1} \cdot x^{-1} \cdot x^{-1} \cdot x^{-1} \cdot x^{-1} \cdot x^{-1} \cdot x^{-1} \cdot x^{-1} \cdot x^{-1} \cdot x^{-1}.$$

For consistency we need

$$x^0 = x^{10} x^{-10} = e.$$

For any $n \in \mathbb{N}^+$ and any $g \in G$ we adopt the following convention:

- x^n means to perform the operation on n copies of x;
- x^{-n} means to perform the operation on n copies of x^{-1};
- $x^0 = e$.

In abelian groups we write nx, $(-n)x$, and $0x$ for the same.

DEFINITION 2.43. Let G be a group. If there exists $g \in G$ such that every element $x \in G$ has the form $x = g^n$ for some $n \in \mathbb{Z}$, then G is a **cyclic group** and we write $G = \langle g \rangle$. We call g a **generator** of G. △

In other words, a cyclic group has the form $\{\ldots, g^{-2}, g^{-1}, e, g^1, g^2, \ldots\}$ where $g^0 = e$.

NOTATION. An abelian group is cyclic if for every $x \in G$ there exists $n \in \mathbb{Z}$ such that $x = ng$.

EXAMPLE 2.44. $\mathbb{Z}$ is cyclic, since any $n \in \mathbb{Z}$ has the form $n \cdot 1$. Thus $\mathbb{Z} = \langle 1 \rangle$. In addition, n has the form $(-n) \cdot (-1)$, so $\mathbb{Z} = \langle -1 \rangle$ as well. You will show in the exercises that $\mathbb{Q}$ is not cyclic. △

Notice that in Definition 2.43 we referred to g as *a* generator of G, not as *the* generator. There could in fact be more than one generator; we see this in Example 2.44 from $\mathbb{Z} = \langle 1 \rangle = \langle -1 \rangle$. Here is another example from $\mathrm{GL}_m(\mathbb{R})$.

EXAMPLE 2.45. Let

$$G = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \right\}.$$

It turns out that G is a group, and that the second and third matrices both generate the group. For example,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$
$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

△

The notion of the size of a cyclic group generated by an element is sufficiently important that we describe it with its own terminology.

DEFINITION 2.46. Let G be a group, and $x \in G$. We say that the **order** of x is $\mathrm{ord}(x) = |\langle x \rangle|$. If $\mathrm{ord}(x) = \infty$, we say that x has **infinite order**. △

If the order of a group is finite, then we have many different ways to represent the same element. Taking the matrix we examined in Example 2.45, we can write

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^0 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^4 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^8 = \cdots$$

and

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^{-4} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^{-8} = \cdots.$$

In addition,

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^3 \quad \text{and} \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^{-1}.$$

So it would seem that if the order of an element G is $n \in \mathbb{N}$, then we can write

$$G = \left\{ e, g, g^2, \ldots, g^{n-1} \right\}.$$

The examples we have looked at so far suggest this. To prove it in general, we have to show that for a generic cyclic group $\langle g \rangle$ with $\mathrm{ord}(g) = n$,

- $g^n = e$, and
- if $a, b \in \mathbb{Z}$ and $n \mid (a - b)$, then $g^a = g^b$.

We prove this in the following Theorem.

THEOREM 2.47. *Suppose that G is a group, $g \in G$, and $\mathrm{ord}(g) = d$, where $d \neq \infty$. Then*

- $g^n = e$, *and*

- *if* $a, b \in \mathbb{Z}$ *and* $n \mid (a-b)$, *then* $g^a = g^b$.

The following lemma will prove useful when attacking the theorem, and is a good rule in general .

LEMMA 2.48. *Let* G *be a group,* $g \in G$, *and* $\operatorname{ord}(g) = n$. *If* $0 \leq a < b < n$, *then* $g^a \neq g^b$.

PROOF. Let $H = \langle g \rangle$. By hypothesis, $\operatorname{ord}(g) = n$, so $|H| = n$.

By way of contradiction, suppose that there exist a, b such that $0 \leq a < b < n$ and $g^a = g^b$; then $e = (g^a)^{-1} g^b$. By Exercise 2.51, we can write

$$e = g^{-a} g^b = g^{-a+b}.$$

Write $d = -a + b$. Recall that $b > a$; this implies that $d > 0$. Use the well-ordering property of $\mathbb{N}$ to choose the smallest d such that $g^d = e$ and $n > b > d$. We can now identify the following distinct elements of H:

$$g, g^2, g^3, \ldots, g^d = e.$$

Now $d < n$ implies that this list omits $d - n$ elements of H. Let x be one such element. Since $H = \langle g \rangle$, we can express $x = g^c$ for some $c \in \mathbb{Z}$. We already listed $c = 0$ with $g^d = e$, so $c \neq 0$. If $c < 0$ then either g^{-c} is already listed, or it is not.

- In the first case, Exercise 2.51 tells us that $g^c = \left(g^{-c}\right)^{-1}$. This implies that $g^c = \left(g^{-c}\right)^{-1}$ is already listed as $g^{d-(-c)}$! (Note $0 < d - (-c) < d$.) Hence the first case is impossible.
- In the second case, let q, r be the result from applying the Division Theorem to division of $-c$ by d. Then $g^{-c} = g^{qd+r}$. By Exercise 2.51,

$$g^{-c} = \left(g^d\right)^q \cdot g^r = e^q \cdot g^r = e \cdot g^r = g^r.$$

  Since $0 \leq r < d$, we have already listed g^r. This contradicts the assumption that $g^{-c} = g^r$ was not listed, so the second case is impossible.

Neither of the two cases is possible, but they are the logical conclusion of assuming the existence of a, b such that $0 \leq a < b < n$ and $g^a = g^b$. Hence if $0 \leq a < b < n$, then $g^a \neq g^b$. □

Now we prove Theorem 2.47.

PROOF OF THEOREM 2.47. Let $H = \langle g \rangle$. By hypothesis, $\operatorname{ord}(g) = n$, so $|H| = n$.

If $n = 1$ then $H = \{e\} = \langle e \rangle$, and the theorem is trivial. Assume therefore that $n > 1$.

Since H is a group, $e \in H$; since $H = \langle g \rangle$, some power of g generates e. Let $d \in \mathbb{N}_{>0}$ such that $g^d = e$; since H only contains n elements, $1 < d \leq n$. We cannot have $d < n$: this would contradict Lemma 2.48 (with $a = 0$ and $b = d$). Hence $d = n$, and $g^n = e$.

Let $a, b \in \mathbb{Z}$. Assume that $n \mid (a-b)$. Let $q \in \mathbb{Z}$ such that $nq = a - b$. Then

$$g^b = g^b \cdot e = g^b \cdot e^q = g^b \cdot \left(g^d\right)^q = g^b \cdot g^{dq} = g^b \cdot g^{a-b} = g^{b+(a-b)} = g^a,$$

as desired. □

EXERCISES.

EXERCISE 2.49. In Exercise 2.38 you showed that the quaternions form a group under matrix multiplication. Verify that $H = \{\mathbf{1}, -\mathbf{1}, \mathbf{i}, -\mathbf{i}\}$ is a cyclic group. What elements generate H?

EXERCISE 2.50. Recall from Section 2.2 the elliptic curve E determined by the equation $y^2 = x^3 - x$.

(a) Compute the cyclic group generated by $(0,0)$ in E.

(b) Verify that $\left(\sqrt{2}+1, \sqrt{2}+2\right)$ is a point on E.

(c) Compute the cyclic group generated by $\left(\sqrt{2}+1, \sqrt{2}+2\right)$ in E. *Hint:* This goes a lot faster if you work with approximate numbers.

EXERCISE 2.51. Let G be a group, and $x \in G$. Explain why for all $m, n \in \mathbb{Z}$,

(a) $x^{mn} = (x^m)^n$;

(b) $x^{-m} = (x^m)^{-1}$;

(c) $x^{-(mn)} = (x^m)^{-n}$;

(d) $x^{-(mn)} = \left(x^{-m}\right)^n$.

Hint: Once you show (a), you can use it to explain the rest.

EXERCISE 2.52. Let G be a group, $g \in G$, and $d \in \mathbb{N}^+$. Show that $\operatorname{ord}(g) = d$ if and only if d is the smallest positive integer such that $g^d = e$. *Hint:* Use the Division Theorem.

EXERCISE 2.53. Let G be a group, and $g \in G$. Assume $\operatorname{ord}(g) = d$. Show that $g^n = e$ for all integer multiples n of d. *Hint:* Use Exercise 2.52.

EXERCISE 2.54. Show that any group of 3 elements is cyclic. *Hint:* Look back at Exercise 2.16 on page 19.

EXERCISE 2.55. Is the Klein 4-group (Exercise 2.17 on page 19) cyclic? What about the other group of order 4?

EXERCISE 2.56. Show that Q_8 is not cyclic.

EXERCISE 2.57. Show that $\mathbb{Q}$ is not cyclic. *Hint:* Show that no matter what you choose for $x \in \mathbb{Q}$, there is some $y \in \mathbb{Q}$ such that $y \notin \langle x \rangle$. Use denominators to do this.

EXERCISE 2.58. Use a fact from linear algebra to explain why $\mathbb{R}^{m \times m}$ is not cyclic.

EXERCISE 2.59. Explain why every cyclic group is abelian.

2.5. THE SYMMETRIES OF A TRIANGLE

Here we introduce a very important group, called D_3. It derives from the symmetries of a triangle. What is interesting about this group is that *it is not abelian.* You already know that groups of order 2, 3, and 4 are abelian; in Section 3.3 you will see that a group of order 5 is also abelian. Thus D_3 is the smallest non-abelian group.

Draw an equilateral triangle in $\mathbb{R}^2$, with its center at the origin. What distance-preserving functions map $\mathbb{R}^2$ to itself, while mapping points on the triangle back onto the triangle? To answer this question, we divide it into two parts.

(1) What are the distance-preserving functions that map $\mathbb{R}^2$ to itself, *without moving the origin?*
(2) Which of these functions map points on the triangle back onto the triangle?

Lemma 2.60 answers the first question. The assumption that we not move the origin makes sense in the context of the triangle, because if we preserve distances, the origin will have to stay fixed as well.

LEMMA 2.60. *Let* $\alpha : \mathbb{R}^2 \to \mathbb{R}^2$. *If*

- $\alpha(0,0) = (0,0)$, *and*
- *the distance between* $\alpha(P)$ *and* $\alpha(R)$ *is the same as the distance between* P *and* R *for every* $P, R \in \mathbb{R}^2$,

then α *has one of the following two forms:*

$$\rho = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \quad \exists t \in \mathbb{R}$$

or

$$\varphi = \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix} \quad \exists t \in \mathbb{R}.$$

The values of t might not be the same for ρ and φ; we will see this later.

PROOF. Assume that $\alpha(0,0) = (0,0)$ and for every $P, R \in \mathbb{R}^2$ the distance between $\alpha(P)$ and $\alpha(R)$ is the same as the Euclidean distance between P and R. It turns out that we can determine α merely by considering how it acts on two points in the plane!

First, let $P = (1,0)$. Write $\alpha(P) = Q = (q_1, q_2)$; this is the point where α moves Q. The distance between P and the origin is 1. Since $\alpha(0,0) = (0,0)$, the distance between Q and the origin is $\sqrt{q_1^2 + q_2^2}$. Because α preserves distance,

$$1 = \sqrt{q_1^2 + q_2^2},$$

or

$$q_1^2 + q_2^2 = 1.$$

The only values for Q that satisfy this equation are those points that lie on the circle whose center is the origin. Any point on this circle can be parametrized as

$$(\cos t, \sin t)$$

where $t \in \mathbb{R}$ represents an angle. Hence, $\alpha(P) = (\cos t, \sin t)$.

Let $R = (0,1)$. Write $\alpha(R) = S = (s_1, s_2)$. An argument similar to the one above shows that S also lies on the circle whose center is the origin. Moreover, the distance between P and R is $\sqrt{2}$, so the distance between Q and S is also $\sqrt{2}$. That is,

$$\sqrt{(\cos t - s_1)^2 + (\sin t - s_2)^2} = \sqrt{2},$$

or

$$(\cos t - s_1)^2 + (\sin t - s_2)^2 = 2. \tag{2.5.1}$$

We can simplify (2.5.1) to obtain

$$-2(s_1 \cos t + s_2 \sin t) + \left(s_1^2 + s_2^2\right) = 1. \tag{2.5.2}$$

To solve this, recall that the distance from S to the origin must be the same as the distance from R to the origin, which is 1. Hence

$$\begin{aligned} \sqrt{s_1^2 + s_2^2} &= 1 \\ s_1^2 + s_2^2 &= 1. \end{aligned}$$

Substituting this into (2.5.2), we find that

$$\begin{aligned} -2(s_1 \cos t + s_2 \sin t) + s_1^2 + s_2^2 &= 1 \\ -2(s_1 \cos t + s_2 \sin t) + 1 &= 1 \\ -2(s_1 \cos t + s_2 \sin t) &= 0 \\ s_1 \cos t &= -s_2 \sin t. \end{aligned} \tag{2.5.3}$$

At this point we can see that $s_1 = \sin t$ and $s_2 = \cos t$ would solve the problem. Are there any other solutions?

Recall that $s_1^2 + s_2^2 = 1$, so $s_2 = \pm\sqrt{1 - s_1^2}$. Likewise $\sin t = \pm\sqrt{1 - \cos^2 t}$. Substituting into equation (2.5.3), we find that

$$\begin{aligned} s_1 \cos t &= -\sqrt{1 - s_1^2} \cdot \sqrt{1 - \cos^2 t} \\ s_1^2 \cos^2 t &= \left(1 - s_1^2\right)\left(1 - \cos^2 t\right) \\ s_1^2 \cos^2 t &= 1 - \cos^2 t - s_1^2 + s_1^2 \cos^2 t \\ s_1^2 &= 1 - \cos^2 t \\ s_1^2 &= \sin^2 t \\ \therefore s_1 &= \pm \sin t. \end{aligned}$$

Along with equation (2.5.3), this implies that $s_2 = \mp \cos t$. Thus there are *two* possible values of s_1 and s_2.

It can be shown (see Exercise 2.65) that α is a linear transformation on the vector space $\mathbb{R}^2$ with the basis $\left\{\vec{P}, \vec{R}\right\} = \{(0,1), (1,0)\}$. We can thus describe it by a matrix. If $s = (\sin t, -\cos t)$ then

$$\alpha = \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix};$$

otherwise

$$\alpha = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

The lemma names the first of these forms φ and the second ρ. □

Before answering the second question, let's consider an example of what the two basic forms of α do to the points in the plane.

EXAMPLE 2.61. Consider the set of points $\mathscr{S} = \{(0,2), (\pm 2,1), (\pm 1,-2)\}$; these form a (non-regular) pentagon in the plane. See Figure . Let $t = \pi/4$; then

$$\rho = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \quad \text{and} \quad \varphi = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

If we apply ρ to every point in the plane, then the points of $\mathscr{S}$ move to

$$\begin{aligned}
\rho(\mathscr{S}) &= \{\rho(0,2), \rho(-2,1), \rho(2,1), \rho(-1,-2), \rho(1,-2)\} \\
&= \left\{ \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \right. \\
&\qquad \left. \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\} \\
&= \left\{ \left(-\sqrt{2}, \sqrt{2}\right), \left(-\sqrt{2} - \frac{\sqrt{2}}{2}, -\sqrt{2} + \frac{\sqrt{2}}{2}\right), \right. \\
&\qquad \left. \left(\sqrt{2} - \frac{\sqrt{2}}{2}, \sqrt{2} + \frac{\sqrt{2}}{2}\right), \left(-\frac{\sqrt{2}}{2} + \sqrt{2}, -\frac{\sqrt{2}}{2} - \sqrt{2}\right), \left(\frac{\sqrt{2}}{2} + \sqrt{2}, \frac{\sqrt{2}}{2} - \sqrt{2}\right) \right\} \\
&\approx \{(-1.4, 1.4), (-2.1, -0.7), (0.7, 2.1), (0.7, -2.1), (2.1, -0.7)\}.
\end{aligned}$$

If we apply φ to every point in the plane, then the points of $\mathscr{S}$ move to

$$\begin{aligned}
\varphi(\mathscr{S}) &= \{\varphi(0,2), \varphi(-2,1), \varphi(2,1), \varphi(-1,-2), \varphi(1,-2)\} \\
&\approx \{(1.4, -1.4), (-0.7, -2.1), (2.1, 0.7), (-2.1, 0.7), (-0.7, 2.1)\}.
\end{aligned}$$

This is shown in Figure 2.1 . The line of reflection for φ has slope $\left(1 - \cos\frac{\pi}{4}\right) / \sin\frac{\pi}{4}$. (You will show this in Exercise 2.67) △

The second questions asks which of the matrices described by Lemma 2.60 also preserve the triangle. To answer this, let's draw a diagram of an equilateral triangle at the origin.

- The first solution (ρ) corresponds to a rotation of degree t of the plane. To preserve the triangle, we can only have $t = 0, 2\pi/3, 4\pi/3$ (0°, 120°, 240°). (See Figure 2.3(a).) Let ι correspond to $t = 0$, the identity rotation; notice that

$$\iota = \begin{pmatrix} \cos 0 & -\sin 0 \\ \sin 0 & \cos 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

FIGURE 2.1. Actions of ρ and φ on a pentagon, with $t = \pi/4$

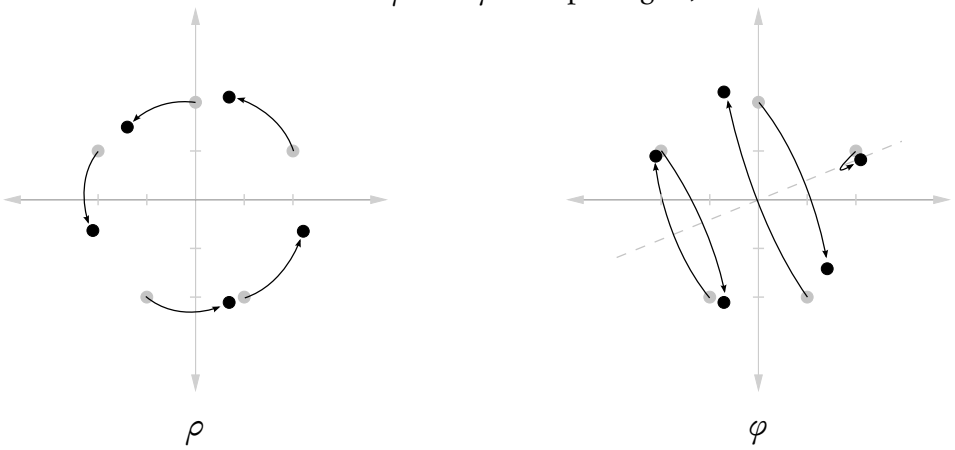


FIGURE 2.2. An equilateral triangle at the origin

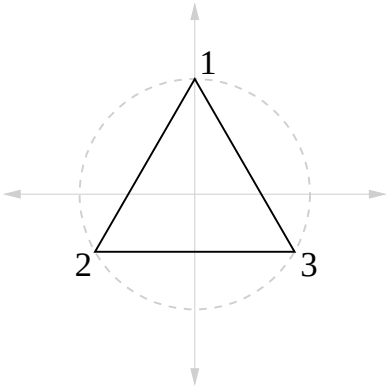
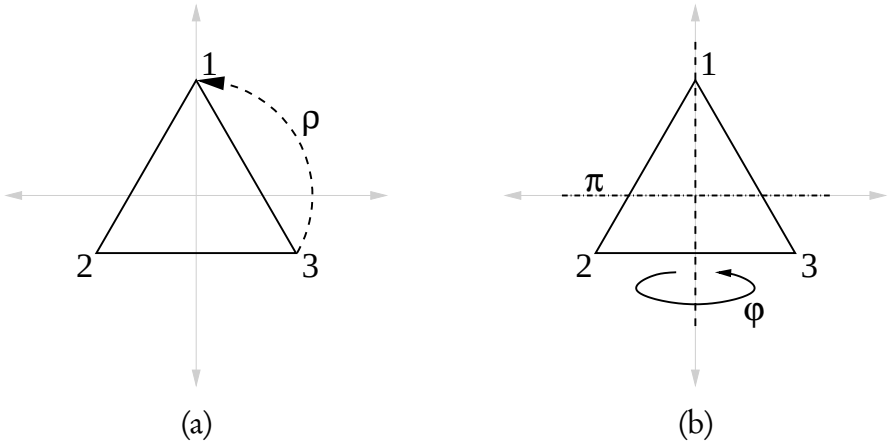


FIGURE 2.3. Rotation and reflection of the triangle



which is what we would expect for the identity. We can let ρ correspond to a counterclockwise rotation of 120°, so

$$\rho = \begin{pmatrix} \cos\frac{2\pi}{3} & -\sin\frac{2\pi}{3} \\ \sin\frac{2\pi}{3} & \cos\frac{2\pi}{3} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}.$$

A rotation of 240° is the same as rotating 120° twice. We can write that as $\rho \circ \rho$ or ρ^2; matrix multiplication gives us

$$\rho^2 = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}.$$

- The second solution (φ) corresponds to a flip along the line whose slope is

$$m = (1 - \cos t) / \sin t.$$

One way to do this would be to flip across the y-axis (see Figure 2.3(b)). For this we need the slope to be undefined, so the denominator needs to be zero and the numerator needs to be non-zero. One possibility for t is $t = \pi$ (but not $t = 0$). Any flip is its own inverse; that is $\varphi^2 = \varphi \circ \varphi = \iota$. To preserve the triangle, we can only have $t = \pi/3, \pi, 5\pi/3$ (60°, 180°, 300°). We can let φ correspond to a flip $t = \pi$ so

$$\varphi = \begin{pmatrix} \cos \pi & \sin \pi \\ \sin \pi & -\cos \pi \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

There are two other flips, but we can actually ignore them, because they are combinations of φ and ρ. (Why? See Exercise 2.64.)

We have the following interesting consequence.

COROLLARY 2.62. *In D_3, $\varphi\rho = \rho^2\varphi$.*

PROOF. Compare

$$\varphi\rho = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$$

and

$$\begin{aligned} \rho^2\varphi &= \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}. \end{aligned}$$

□

Did you notice something interesting about Corollary 2.62? It implies that multiplication in D_3 is non-commutative! We have $\varphi\rho = \rho^2\varphi$, and a little logic (or an explicit computation) shows that $\rho^2\varphi \neq \rho\varphi$: thus $\varphi\rho \neq \rho\varphi$.

Let $D_3 = \{\iota, \varphi, \rho, \rho^2, \rho\varphi, \rho^2\varphi\}$. These are matrices, and we denote the multiplication of these matrices by $\circ$. We can start to fill in a multiplication table for D_3 using everything that we have studied so far:

$\circ$	ι	φ	ρ	ρ^2	$\rho\varphi$	$\rho^2\varphi$
ι	ι	φ	ρ	ρ^2	$\rho\varphi$	$\rho^2\varphi$
φ	φ	ι	$\rho^2\varphi$			
ρ	ρ	$\rho\varphi$		ι		
ρ^2	ρ^2					
$\rho\varphi$	$\rho\varphi$					
$\rho^2\varphi$	$\rho^2\varphi$					

You will complete the table in the exercises, and explain why D_3 is a group.

EXERCISES.

EXERCISE 2.63. Complete the multiplication table for D_3. Explain how D_3 satisfies the properties of a group. *Hint:* To show that D_3 satisfies the properties of a group, you may use the fact that D_3 is a subset of $\mathrm{GL}(2)$, the multiplicative group of 2×2 invertiable matrices. Thus D_3 "inherits" certain properties of $\mathrm{GL}(2)$, but which ones? For the others, simple inspection of the multiplication table should suffice.

EXERCISE 2.64. There are two other values of t that allow us to define flips. Find these values of t, and explain why their matrices are equivalent to the matrices $\rho\varphi$ and $\rho^2\varphi$.

EXERCISE 2.65. Show that any function α satisfying the requirements of Theorem 2.60 is a linear transformation; that is, for all $P, Q \in \mathbb{R}^2$ and for all $a, b \in \mathbb{R}$, $\alpha(aP + bQ) = a\alpha(P) + b\alpha(Q)$. Use the following steps.

(a) Prove that $\alpha(P)\cdot\alpha(Q) = P\cdot Q$, where $\cdot$ denotes the usual dot product (or inner product) on $\mathbb{R}^2$. *Hint:* You may use the property that $|P-Q|^2 = |P|^2 + |Q|^2 - 2P\cdot Q$, where $|X|$ indicates the distance of X from the origin, and $|X-Y|$ indicates the distance between X and Y.
(b) Show that $\alpha(1,0)\cdot\alpha(0,1) = 0$.
(c) Show that $\alpha((a,0)+(0,b)) = a\alpha(1,0) + b\alpha(0,1)$. *Hint:* Use the hint from part (a), along with the result in part (a), to show that the distance between the vectors is zero. Also use the property of dot products that for any vector X, $X\cdot X = |X|^2$. Don't use part (b).
(d) Show that $\alpha(aP) = a\alpha(P)$.
(e) Show that $\alpha(P+Q) = \alpha(P) + \alpha(Q)$.

EXERCISE 2.66. Show that the only point in $\mathbb{R}^2$ left stationary by ρ is the origin. That is, if $\rho(P) = P$, then $P = (0,0)$. *Hint:* Let $P = (p_1, p_2)$ be an arbitrary point in $\mathbb{R}^2$, and assume that ρ leaves it stationary. You can represent P by a vector. The equation $\rho\cdot\vec{P} = \vec{P}$ gives you a system of two linear equations in two variables; you can solve this system for p_1 and p_2.

EXERCISE 2.67. Show that the only points in $\mathbb{R}^2$ left stationary by φ lie along the line whose slope is $(1-\cos t)/\sin t$. *Hint:* Repeat what you did in Exercise 2.66. This time the system of linear equations will have infinitely many solutions. You know from linear algebra that in $\mathbb{R}^2$ this describes a line. Solve one of the equations for p_2 to obtain the equation of this line.

CHAPTER 3

Subgroups

3.1. SUBGROUPS

Subgroups play an important role in group theory and its applications.

DEFINITION 3.1. Let G be a group and $H \subseteq G$ a nonempty subset. If H is also a group under the same operation as G, then H is a **subgroup** of G. If $\{e\} \subsetneq H \subsetneq G$ then H is a **proper subgroup** of G. △

NOTATION. If H is a subgroup of G then we write $H < G$.

EXAMPLE 3.2. Check that the following statements are true by verifying that properties (G1)–(G4) are satisfied.

(a) $\mathbb{Z}$ is an abelian subgroup of $\mathbb{Q}$.
(b) $4\mathbb{Z} := \{4m : m \in \mathbb{Z}\} = \{\ldots, -4, 0, 4, 8, \ldots\}$ is an abelian subgroup of $\mathbb{Z}$.
(c) Let $d \in \mathbb{Z}$. Then $d\mathbb{Z} := \{dm : m \in \mathbb{Z}\}$ is an abelian subgroup of $\mathbb{Z}$.
(d) $\langle \mathbf{i} \rangle$ is a subgroup of Q_8. △

Checking all of properties (G1)–(G5) is burdensome. If we can verify that a set is a subgroup by checking fewer properties, that would be lovely. From the start we can eliminate (G2) and (G5) from consideration: if $H \subseteq G$, then the operation remains associative and commutative even when H is not a subgroup.

LEMMA 3.3. *Let G be a group and $H \subseteq G$. Then H satisfies the associative property (G2) of a group. In addition, if G is abelian, then H satisfies the commutative property (G5) of an abelian group.*

We are *not* saying that H is a group. Any one of the other three properties may not be satisfied: it may not be closed; it may lack an identity; or some element may lack an inverse. We are merely pointing out that H satisfies two important properties of a group.

PROOF. If $H = \emptyset$ then the lemma is true trivially.
Otherwise $H \neq \emptyset$. Let $a, b, c \in H$. Since $H \subseteq G$, we have $a, b, c \in G$. Since the operation is associative in G, $a(bc) = (ab)c$. If G is abelian, then $ab = ba$. □

The upshot of Lemma 3.3 is that whenever we want to prove that a subset of a group is also a subgroup, we do not have to prove the associative and commutative properties, (G2) and (G5). We need to prove only that the subsets have an identity, have inverses, and are closed under the operation.

LEMMA 3.4. *Let $H \subseteq G$ be nonempty. The following are equivalent:*

(A) $H < G$;
(B) H satisfies (G1), (G3), and (G4).

PROOF. By definition of a group, (A) implies (B). It remains to show that (B) implies (A).

Assume (B). Then H satisfies (G1), (G3), and (G4). Lemma 3.3 shows us that H also satisfies (G2). Hence H is a group, from which we have (A). □

Lemma 3.4 has reduced the number of requirements for a subgroup. However, three is still too many; it turns out that we can simplify the process by checking *only one criterion*.

THEOREM 3.5 (The Subgroup Theorem). *Let $H \subseteq G$ be nonempty. The following are equivalent:*

(A) $H < G$;
(B) for every $x, y \in H$, we have $xy^{-1} \in H$.

PROOF. Assume (A). Let $x, y \in H$. By (A), H is a group; from (G4) we have $y^{-1} \in H$, and from (G1) we have $xy^{-1} \in H$. Thus (A) implies (B).

Conversely, assume (B). By Lemma 3.4, we need to show only that H satisfies (G1), (G3), and (G4). We do this slightly out of order:

(G3): Let $x \in H$. By (B), $e = x \cdot x^{-1} \in H$.[1]
(G4): Let $x \in H$. Since H satisfies (G3), $e \in H$. By (B), $x^{-1} = e \cdot x^{-1} \in H$.
(G1): Let $x, y \in H$. Since H satisfies (G4), $y^{-1} \in H$. By (B), $xy = x \cdot \left(y^{-1}\right)^{-1} \in H$.

Since H satisfies (G1), (G3), and (G4), $H < G$. □

The Subgroup Theorem makes it much easier to decide whether a subset of a group is a subgroup, because we need to consider only the one criterion given. Our first example is from $\mathbb{Z}$. Remember that if G is abelian, we generally write $x - y$ instead of xy^{-1} in (B).

EXAMPLE 3.6. Let $d \in \mathbb{Z}$. We claim that $d\mathbb{Z} < \mathbb{Z}$. *Why?* Let's use the Subgroup Theorem.

Let $x, y \in d\mathbb{Z}$. By definition, $x = dm$ and $y = dn$ for some $m, n \in \mathbb{Z}$. Note that $-y = -(dn) = d(-n)$. Then

$$x - y = x + (-y) = dm + d(-n) = d(m + (-n)) = d(m - n).$$

Now $m - n \in \mathbb{Z}$, so $x - y = d(m - n) \in d\mathbb{Z}$. By the Subgroup Theorem, $d\mathbb{Z} < \mathbb{Z}$. ____△

The following geometric example gives a visual image of what a subgroup "looks" like.

EXAMPLE 3.7. Let G be the set of points in the x-y plane. Define an addition for elements of G in the following way. For $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$, define

$$P_1 + P_2 = (x_1 + y_1, x_2 + y_2).$$

You showed in Exercise 2.13 that this makes G a group. (Actually you proved it for $G \times H$ where G and H were groups. Here $G = H = \mathbb{R}$.)

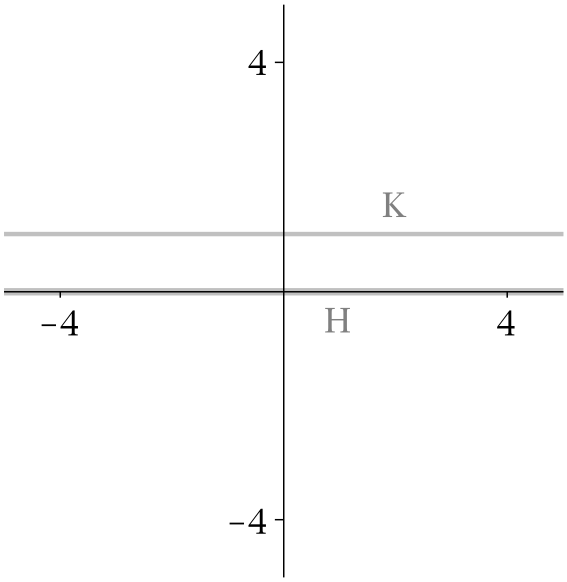
Let $H = \{x \in G : x = (a, 0)\ \exists a \in \mathbb{R}\}$. We claim that $H < G$. *Why?* We use the subgroup theorem. Let $P, Q \in H$. Write $P = (p, 0)$ and $Q = (q, 0)$ where $p, q \in \mathbb{R}$. Then

$$P - Q = P + (-Q) = (p, 0) + (-q, 0) = (p - q, 0).$$

Membership in H requires the second ordinate to be zero. The second ordinate of $P - Q$ is in fact zero, so $P - Q \in H$. The Subgroup Theorem implies that $H < G$.

[1] Notice that here we are replacing the y in (B) with x. This is fine, since nothing in (B) requires x and y to be distinct.

FIGURE 3.1. H and K from Example 3.7



Let $K = \{x \in G : x = (a,1)\ \exists a \in \mathbb{R}\}$. We claim that $K \not< G$. *Why not?* Again we use the Subgroup Theorem. Let $P, Q \in K$. Write $P = (p,1)$ and $Q = (q,1)$ where $p, q \in \mathbb{R}$. Then

$$P - Q = P + (-Q) = (p,1) + (-q,-1) = (p-q,0).$$

Membership in K requires the second ordinate to be one, but the second ordinate of $P - Q$ is zero, not one. Since $P - Q \notin K$, the Subgroup Theorem tells us that K is not a subgroup of G.

We have diagrammed H and K in Figure 3.1. You will diagram another subgroup of G in Exercise 3.13. △

Examples 3.6 and 3.7 give us examples of how the Subgroup Theorem verifies subgroups of *abelian* groups. Two interesting examples of nonabelian subgroups appear in D_3.

EXAMPLE 3.8. Recall D_3 from Section 2.5. Both $H = \{\iota, \varphi\}$ and $K = \{\iota, \rho, \rho^2\}$ are subgroups of D_3. *Why?* We verify each using the Subgroup Theorem, exploiting the fact that both H and K are cyclic groups: $H = \langle\varphi\rangle$ and $K = \langle\rho\rangle$.

For H: Let $x, y \in H$. Since $H = \langle\varphi\rangle$, $x = \varphi^m$ and $y = \varphi^n$ for some $m, n \in \mathbb{Z}$. Applying Exercise 2.51 on page 29, $xy^{-1} = \varphi^m \varphi^{-n} = \varphi^{m-n} \in \langle\varphi\rangle = H$. By the Subgroups Theorem, $H < D_3$.

For K: Repeat the argument for H, using ρ instead of φ. △

If a group satisfies a given property, a natural question to ask is whether its subgroups also satisfy this property. Cyclic groups are a good example: is every subgroup of a cyclic group also cyclic? The answer relies on the Division Theorem (Theorem 1.5 on page 13).

THEOREM 3.9. *Subgroups of cyclic groups are also cyclic.*

PROOF. Let G be a cyclic group, and $H < G$. From the fact that G is cyclic, choose $g \in G$ such that $G = \langle g\rangle$.

First we must find a candidate generator of H. Because $H \subseteq G$, every element $x \in H$ can be written in the form $x = g^i$ for some $i \in \mathbb{Z}$. A good candidate would be the smallest positive

power of g in H, if one exists. Let S be the set of all positive integers i such that $g^i \in H$. From the well-ordering of the integers, there exists a smallest element of S; call it d, and assign $h = g^d$.

We have found a candidate; we claim that $H = \langle h \rangle$. Let $x \in H$; then $x \in G$. By hypothesis G is cyclic, so $x = g^a$ for some $a \in \mathbb{Z}$. By the Division Theorem we know that there exist unique $q, r \in \mathbb{Z}$ such that

- $a = qd + r$, and
- $0 \leq r < d$.

Let $y = g^r$; by Exercise 2.51 we can rewrite this as

$$y = g^r = g^{a-qd} = g^a g^{-(qd)} = x \cdot \left(g^d\right)^{-q} = x \cdot h^{-q}.$$

Now $x \in H$ by definition, and $h^{-q} \in H$ by closure (G1) and the existence of inverses (G4), so by closure $y = x \cdot h^{-q} \in H$ as well. We chose d as the smallest positive power of g in H, and we just showed that $g^r \in H$. Recall that $0 \leq r < d$. If $0 < r$; then $g^r \in H$, contradicting the choice of d as the smallest power of g in H. Hence r cannot be positive; instead, $r = 0$ and $x = g^a = g^{qd} = h^q \in \langle h \rangle$.

Since x was arbitrary in H, every element of H is in $\langle h \rangle$; that is, $H \subseteq \langle h \rangle$. Since $h \in H$ and H is a group, closure implies that $H \supseteq \langle h \rangle$, so $H = \langle h \rangle$. In other words, H is cyclic. □

We again look to $\mathbb{Z}$ for an example.

EXAMPLE 3.10. Recall from Example 2.44 on page 26 that $\mathbb{Z}$ is cyclic; in fact $\mathbb{Z} = \langle 1 \rangle$. By Theorem 3.9, $d\mathbb{Z}$ is cyclic. In fact, $d\mathbb{Z} = \langle d \rangle$. Can you find another generator of $d\mathbb{Z}$? ___△

EXERCISES.

EXERCISE 3.11. Show that even though the Klein 4-group is not cyclic, each of its proper subgroups is cyclic (see Exercises 2.17 on page 19 and 2.55 on page 29). *Hint:* Start with the smallest possible subgroup, then add elements one at a time. Don't forget the adjective "proper" subgroup.

EXERCISE 3.12.

(a) Let $D_n(\mathbb{R}) = \{aI_n : a \in \mathbb{R}\} \subseteq \mathbb{R}^{n \times n}$; that is, $D_n(\mathbb{R})$ is the set of all diagonal matrices whose values along the diagonal is constant. Show that $D_n(\mathbb{R}) < \mathbb{R}^{n \times n}$.

(b) Let $D_n^*(\mathbb{R}) = \{aI_n : a \in \mathbb{R} \setminus \{0\}\} \subseteq \mathrm{GL}_n(\mathbb{R})$; that is, $D_n^*(\mathbb{R})$ is the set of all non-zero diagonal matrices whose values along the diagonal is constant. Show that $D_n^*(\mathbb{R}) < \mathrm{GL}_n(\mathbb{R})$.

EXERCISE 3.13. Let $G = \mathbb{R}^2 := \mathbb{R} \times \mathbb{R}$, with addition defined as in Exercise 2.13 and Example 3.7.

(a) Let $L = \{x \in G : x = (a,a)\ \exists a \in \mathbb{R}\}$. Show that $L < G$.
(b) Describe L geometrically.
(c) Suppose $\ell \subseteq G$ is any line. Identify as general a criterion as possible that decides whether $\ell < G$. Justify your answer. *Hint:* Look at what H from Example 3.7 and L have in common.

EXERCISE 3.14. Let G be an abelian group. Let H, K be abelian subgroups of G. Let $H+K = \{x+y : x \in H, y \in K\}$. Show that $H+K < G$.

EXERCISE 3.15. Let $H = \{\iota, \varphi\}$; recall that $H < D_3$.

(a) Find a different subgroup K of D_3 with only two elements.
(b) Let $HK = \{xy : x \in H, y \in K\}$. Show that $HK \not< D_3$.
(c) Why does the result of (b) not contradict the result of Exercise 3.14?

EXERCISE 3.16. Explain why $\mathbb{R}$ cannot be cyclic. *Hint:* Use Exercise 2.57 on page 29.

3.2. COSETS

Recall the Division Theorem (Theorem 1.5 on page 13). Normally, we think of division of n by d as dividing n into q parts, each containing d elements, with r elements left over. For example, $n = 23$ apples divided among $d = 6$ bags gives $q = 3$ apples per bag and $r = 5$ apples left over.

Another way to look at division by d is that it divides $\mathbb{Z}$ into d sets of integers. Each integer falls into a set according to its remainder after division. An illustration using $n = 4$:

$\mathbb{Z}$:	…	-2	-1	0	1	2	3	4	5	6	7	8	…
		↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
division by 4:	…	2	3	0	1	2	3	0	1	2	3	0	…

Here $\mathbb{Z}$ is divided into four sets

$$\begin{aligned} A &= \{\ldots, -4, 0, 4, 8, \ldots\} \\ B &= \{\ldots, -3, 1, 5, 9, \ldots\} \\ C &= \{\ldots, -2, 2, 6, 10, \ldots\} \\ D &= \{\ldots, -1, 3, 7, 11, \ldots\}. \end{aligned} \tag{3.2.1}$$

Observe two important facts:

- the sets A, B, C, and D *cover* $\mathbb{Z}$; that is,

$$\mathbb{Z} = A \cup B \cup C \cup D;$$

and

- the sets A, B, C, and D are *disjoint;* that is,

$$A \cap B = A \cap C = A \cap D = B \cap C = B \cap D = C \cap D = \emptyset.$$

We can diagram this:

$\mathbb{Z}$

A
B
C
D

This phenomenon, where a set is the union of smaller, disjoint sets, is important enough to highlight with a definition.

DEFINITION 3.17. Suppose that A is a set and $\mathscr{B} = \{B_\lambda\}$ a family of subsets of A, called **classes**. We say that $\mathscr{B}$ is a **partition** of A if

- the classes **cover** A: that is, $A = \bigcup B_\lambda$; and
- the classes are disjoint: that is, if $B_1, B_2 \in \mathscr{B}$ are unequal, then $B_1 \cap B_2 = \emptyset$. △

EXAMPLE 3.18. Let $\mathscr{B} = \{A, B, C, D\}$ where A, B, C, and D are defined as in (3.18). Then $\mathscr{B}$ is a partition of $\mathbb{Z}$. △

Two aspects of division allow us to use it to partition $\mathbb{Z}$ into sets:

- *existence of a remainder,* which implies that every integer belongs to at least one class, which in turn implies that the union of the classes covers $\mathbb{Z}$; and
- *uniqueness of the remainder,* which implies that every integer ends up in only one set, so that the classes are disjoint.

Re-examine this phenomenon using the vocabulary of groups. In the example above, you might have noticed that $A = 4\mathbb{Z}$. (If not, look back at the definition of $4\mathbb{Z}$ on page 3.2.) So, $A < \mathbb{Z}$. Meanwhile, all the elements of B have the form $1 + x$ for some $x \in A$. For example, $-3 = 1 + (-4)$. Likewise, all the elements of C have the form $2 + x$ for some $x \in A$, and all the elements of D have the form $3 + x$ for some $x \in A$. Define

$$1 + A := \{1 + x : x \in A\},$$

then set

$$B = 1 + A.$$

Likewise, set $C = 2 + A$ and $D = 3 + A$.

What about $0 + A$? Clearly $0 + A = A$; in fact $x + A = A$ for every $x \in A$. Pursuing this further,

$$\cdots = -4 + A = A = 0 + A = 4 + A = 8 + A = \cdots$$

and for that matter

$$\cdots = -3 + A = B = 1 + A = 5 + A = 9 + A = \cdots$$

and so forth. Interestingly, $B = 1 + A$, and $B = 5 + A$. Notice that $1 - 5 = -4 \in A$. We could do the same with C: $C = 2 + A$ and $C = 10 + A$, and $2 - 10 = -8 \in A$. This relationship will prove important at the end of the section.

So the partition by remainders is related to the subgroup A. This will become very important in Chapter 6, and it is important in general.

Mathematicians love to generalize any phenomena they observe, and this is no exception. How can we generalize this to arbitrary subgroups?

DEFINITION 3.19. Let G be a group and $A < G$. Let $g \in G$. We define the **left coset of** A **with** g as

$$gA = \{ga : a \in A\}$$

and the **right coset of** A **with** g as

$$Ag = \{ag : a \in A\}.$$

If A is an abelian subgroup, we write the coset of A with g as

$$g + A := \{g + a : a \in A\}.$$

△

In general, left cosets and right cosets are not equal, partly because the operation might not commute. Example 3.20 illustrates this.

EXAMPLE 3.20. Recall the group D_3 from Section (2.5) and the subgroup $H = \{\iota, \varphi\}$ from Example 3.8. In this case,

$$\rho H = \{\rho, \rho\varphi\} \text{ and } H\rho = \{\rho, \varphi\rho\}.$$

Since $\varphi\rho = \rho^2\varphi \neq \rho\varphi$, we see that $\rho H \neq H\rho$. △

Sometimes, the left coset and the right coset *are* equal. This is always true in abelian groups, as illustrated by Example 3.21.

EXAMPLE 3.21. Consider the subgroup $H = \{(a,0) : a \in \mathbb{R}\}$ of $\mathbb{R}^2$ from Exercise 3.13 on page 39. Let $p = (3,-1) \in \mathbb{R}^2$. The coset of H with p is

$$\begin{aligned} p + H &= \{(3,-1) + q : q \in H\} \\ &= \{(3,-1) + (a,0) : a \in \mathbb{R}\} \\ &= \{(3+a,-1) : a \in \mathbb{R}\}. \end{aligned}$$

Sketch some of the points in $p + H$, and compare them to your sketch of H in Exercise 3.13. How does the coset compare to the subgroup?

Generalizing this further, every coset of H has the form $p + H$ where $p \in \mathbb{R}^2$. Elements of $\mathbb{R}^2$ are points, so $p = (x,y)$ for some $x, y \in \mathbb{R}$. The coset of H with p is

$$p + H = \{(x+a, y) : a \in \mathbb{R}\}.$$

Sketch several more cosets. How would you describe the set of *all* cosets of H in $\mathbb{R}^2$? △

The group does not *have* to be abelian in order to have the left and right cosets equal. When deciding if $gA = Ag$, we are not deciding *whether elements commute*, but *whether sets are equal*. Returning to D_3, we can find a subgroup whose left and right cosets are equal even though the group is not abelian and the operation is not commutative.

EXAMPLE 3.22. Let $K = \{\iota, \rho, \rho^2\}$; certainly $K < D_3$. In this case, $\alpha K = K\alpha$ for all $\alpha \in D_3$:

α	αK	$K\alpha$
ι	K	K
φ	$\{\varphi, \varphi\rho, \varphi\rho^2\} = \{\varphi, \rho\varphi, \rho^2\varphi\}$	$\{\varphi, \rho\varphi, \rho^2\varphi\}$
ρ	K	K
ρ^2	K	K
$\rho\varphi$	$\{\rho\varphi, (\rho\varphi)\rho, (\rho\varphi)\rho^2\} = \{\rho\varphi, \varphi, \rho^2\varphi\}$	$\{\rho\varphi, \varphi, \rho^2\varphi\}$
$\rho^2\varphi$	$\{\rho^2\varphi, (\rho^2\varphi)\rho, (\rho^2\varphi)\rho^2\} = \{\rho^2\varphi, \rho\varphi, \varphi\}$	$\{\rho^2\varphi, \rho\varphi, \varphi\}$

In each case, the sets φK and $K\varphi$ are equal, even though φ does not commute with ρ. (You should verify these computations by hand.) △

We can now explain why cosets of a subgroup partition a group.

THEOREM 3.23. *The cosets of a subgroup partition the group.*

PROOF. Let G be a group, and $A < G$. We have to show two things:

(CP1) distinct cosets of A are disjoint, and
(CP2) their union is G.

We show (CP2) first. Let $g \in G$. The definition of a group tells us that $g = ge$. Since $e \in A$ by definition of subgroup, $g = ge \in gA$. Since g was arbitrary, every element of G is in some coset of A. Hence the union of all the cosets is G (CP2).

For (CP1), let $x, y \in G$. We proceed by the contrapositive: suppose that $(xA) \cap (yA) \neq \emptyset$. We want to show that $xA = yA$. This requires us to show that two sets are equal, so we will show that $xA \subseteq yA$ and then $xA \supseteq yA$.

Let $g \in xA$ and $h \in (xA) \cap (yA)$. Then $g = xa_1$, and $h = xa_2 = ya_3$ for some $a_1, a_2, a_3 \in A$. From the latter equations we obtain $x = y\left(a_3 a_2^{-1}\right)$. Thus

$$g = xa_1 = \left(y\left(a_3 a_2^{-1}\right)\right) a_1 = y\left(\left(a_3 a_2^{-1}\right) a_1\right) \in yA.$$

Since g was arbitrary in xA, we have shown $xA \subseteq yA$.

A similar argument shows that $xA \supseteq yA$. Thus $xA = yA$. We have shown that if xA and yA have an intersection, then they are equal. The contrapositive of this statement is precisely (CP1). Having shown (CP1) and (CP2), we have shown that the cosets of A partition G. □

Before we finish, we should observe two facts about cosets that parallel facts about A in the example at the beginning of the section. These facts allow us to decide when two cosets are equal. They have enormous consequences later on.

LEMMA 3.24 (Equality of cosets). *Let G be a group and $H < G$. Then (CE1), (CE2), and (CE2) are always true, where:*

(CE1) $eH = H$.
(CE2) For all $a \in G$, $a \in H$ iff $aH = H$.
(CE3) For all $a, b \in G$, $aH = bH$ if and only if $a^{-1}b \in H$.

As usual, you should keep in mind that in additive groups these conditions translate to

(CE1) $0 + H = H$.
(CE2) For all $a \in G$, if $a \in H$ then $a + H = H$.
(CE3) For all $a, b \in H$, $a + H = b + G$ if and only if $a - b \in H$.

PROOF. You do it! See Exercise 3.30, which gives a substantial hint. □

Cosets can seem like an odd thing to study, but they lie at the foundation of many applications of algebra. You will see this in later chapters.

EXERCISES.

EXERCISE 3.25. Let $\{e, a, b, a+b\}$ be the Klein 4-group. (See Exercises 2.17 on page 19, 2.55 on page 29, and 3.11 on page 39.) Compute the cosets of $\langle a \rangle$.

EXERCISE 3.26. In Exercise 3.15 on page 40, you found another subgroup K of order 2 in D_3. Does K satisfy the property $\alpha K = K\alpha$ for all $\alpha \in D_3$?

EXERCISE 3.27. Recall the subgroup L of $\mathbb{R}^2$ from Exercise 3.13 on page 39.

(a) Give a geometric interpretation of the coset $(3, -1) + L$.
(b) Give an algebraic expression that describes $p + L$, for arbitrary $p \in \mathbb{R}^2$.
(c) Give a geometric interpretation of the cosets of L in $\mathbb{R}^2$.

(d) Use your geometric interpretation of the cosets of L in $\mathbb{R}^2$ to explain why the cosets of L partition $\mathbb{R}^2$.

EXERCISE 3.28. Recall $D_n(\mathbb{R})$ from Exercise 3.12 on page 39. Give a description in set notation for

$$\begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix} + D_2(\mathbb{R}).$$

List some elements of the coset.

EXERCISE 3.29. In the proof of Theorem 3.23 on page 42, we stated that "A similar argument shows that $xA \supseteq yA$." Give this argument.

EXERCISE 3.30. Prove Lemma 3.24 on the preceding page. *Hint:* For (CE1), you have to show that two sets are equal. Follow the structure of the proof for Theorem 3.23 on page 42. Take an arbitrary element of eH, and show that it also an element of H; that gives $eH \subseteq H$. Then take an arbitrary element of H, and show that it is an element of eH; that gives $eH \supseteq H$. The two inclusions give $eH = H$.

As for (CE2) and (CE3), you can prove them in a manner similar to that of (CE1), or you can explain how they are actually consequences of (CE1).

3.3. LAGRANGE'S THEOREM AND THE ORDER OF AN ELEMENT OF A GROUP

NOTATION. Let G be a group, and $A < G$. We write G/A for the set of all left cosets of A. That is,

$$G/A = \{gA : g \in G\}.$$

We also write $G\backslash A$ for the set of all right cosets of A (but not as often):

$$G\backslash A = \{Ag : g \in G\}.$$

EXAMPLE 3.31. Let $G = \mathbb{Z}$ and $A = 4\mathbb{Z}$. We saw in Example 3.18 that

$$G/A = \mathbb{Z}/4\mathbb{Z} = \{A, 1+A, 2+A, 3+A\}.$$

We actually "waved our hands" in Example 3.18, without providing a very detailed argument, so let's show the details here. Recall that $4\mathbb{Z}$ is the set of multiples of $\mathbb{Z}$, so $x \in A$ iff x is a multiple of 4. What about the remaining elements of $\mathbb{Z}$?

Let $x \in \mathbb{Z}$; then

$$x + A = \{x + z : z \in A\} = \{x + 4n : n \in \mathbb{Z}\}.$$

Use the Division Theorem to write

$$x = 4q + r$$

for unique $q, r \in \mathbb{Z}$, where $0 \leq r < 4$. Then

$$x + A = \{(4q + r) + 4n : n \in \mathbb{Z}\} = \{r + 4(q + n) : n \in \mathbb{Z}\}.$$

Given any integer n, let $m = q + n$; by closure, $m \in \mathbb{Z}$. Thus

$$x + A = \{r + 4m : m \in \mathbb{Z}\} = \{r + z : z \in 4\mathbb{Z}\} = r + 4\mathbb{Z}.$$

The distinct cosets of A are thus determined by the distinct remainders from division by 4. Since the remainders from division by 4 are 0, 1, 2, and 3, we conclude that

$$\mathbb{Z}/A = \{A, 1+A, 2+A, 3+A\}$$

as claimed above. △

EXAMPLE 3.32. Let $G = D_3$ and $H = \{\iota, \varphi\}$ as in Example 3.22. Then

$$G/K = D_3/\langle\varphi\rangle = \left\{K, \rho A, \rho^2 A\right\}.$$

Likewise, if $K = \left\{\iota, \rho, \rho^2\right\}$ as in Example 3.22, then

$$G/K = D_3/\langle\rho\rangle = \{K, \varphi A\}.$$

△

EXAMPLE 3.33. Let $H < \mathbb{R}^2$ be as in Example 3.7 on page 37; that is,

$$H = \left\{(a, 0) \in \mathbb{R}^2 : \ a \in \mathbb{R}\right\}.$$

Then

$$\mathbb{R}^2/H = \left\{r + H : \ r \in \mathbb{R}^2\right\}.$$

It is not possible to list all the elements of G/A, but some examples would be

$$(1,1) + H, \ (4,-2) + H.$$

Speaking *geometrically*, what do the elements of G/A look like? △

It is important to keep in mind that G/A is a set whose elements are also sets. As a result, showing equality of two elements of G/A requires one to show that two sets are equal.

When G is finite, a simple formula gives us the size of G/A.

THEOREM 3.34 (Lagrange's Theorem). *Let G be a group of finite order, and $A < G$. Then*

$$|G/A| = \frac{|G|}{|A|}.$$

It is important to note that Lagrange's Theorem is *not* obvious, regardless of what the notation suggests. The formula is saying that the size of the set of cosets is the same as the quotient of the order of G by the order of A. Since G/A is not a number, we cannot move the absolute value bars "inside" the fraction without some sort of explanation.

PROOF. From Theorem 3.23 we know that the cosets of A partition G. There are $|G/A|$ cosets of A. Each of them has the same size, $|A|$. The number of elements of G is thus the product of the number of elements in each coset and the number of cosets. That is, $|G/A| \cdot |A| = |G|$. This implies the theorem. □

The next-to-last sentence of the proof contains the statement $|G/A| \cdot |A| = |G|$. Since $|A|$ is the order of the group A, and $|G/A|$ is an integer, we conclude that:

COROLLARY 3.35. *The order of a subgroup divides the order of a group.*

EXAMPLE 3.36. Let G be the Klein 4-group (see Exercises 2.17 on page 19, 2.55 on page 29, and 3.11 on page 39). Every subgroup of the Klein 4-group is cyclic, and has order 1, 2, or 4. Notice that the orders of the subgroups divide the order of the group, as predicted by Corollary 3.35 on the preceding page.

Likewise, the order of $\{\iota, \varphi\}$ divides the order of D_3.

By contrast, the subset HK of D_3 that you computed in Exercise 3.15 on page 40 has four elements. Since $4 \nmid 6$, the contrapositive of Lagrange's Theorem implies that HK *cannot* be a subgroup of D_3. △

Using the fact that every element g generates a cyclic subgroup $\langle g \rangle < G$, Lagrange's Theorem also implies an important consequence about the order of any element of any finite group.

COROLLARY 3.37. *In a finite group G, the order of any element divides the order of a group.*

PROOF. You do it! See Exercise 3.38. □

EXERCISES.

EXERCISE 3.38. Prove Corollary 3.37.

EXERCISE 3.39. Suppose that a group G has order 8, but is not cyclic. Show that $g^4 = e$ for all $g \in G$.

EXERCISE 3.40. Suppose that a group has five elements. Will it be cyclic? *Hint:* Use Corollary 3.37.

EXERCISE 3.41. Find a sufficient (but not necessary) condition on the order of a group that guarantees that the group is cyclic. *Hint:* See Exercises 2.54 on page 29 and 3.40.

3.4. QUOTIENT GROUPS

Let $A < G$ and suppose that we try to define an operation on the left cosets of A by

$$(gA)(hA) = (gh)A.$$

Will this give us a group?

Even before thinking about the question, we have to ensure that the operation is well-defined. *What does that mean?* A coset can have different representations. The procedure defined above would not be an operation if two different representations of gA gave us two different answers.

EXAMPLE 3.42. Recall the subgroup $A = 4\mathbb{Z}$ of $\mathbb{Z}$. Let $B, C, D \in \mathbb{Z}/A$, so $B = b + \mathbb{Z}$, $C = c + \mathbb{Z}$, and $D = d + \mathbb{Z}$ for some $b, c, d \in \mathbb{Z}$.

Our concern is rooted in the possibility that $B = D$ but $B + C \neq D + C$. From Lemma 3.24, we know that $B = D$ iff $b - d \in A = 4\mathbb{Z}$. That is, $b - d = 4m$ for some $m \in \mathbb{Z}$. Let $x \in B + C$; then $x = (b + c) + 4n$ for some $n \in \mathbb{Z}$; we have $x = ((d + 4m) + c) + 4n = (d + c) + 4(m + n) \in D + C$. Since x was arbitrary in $B + C$, we have $B + C \subseteq D + C$. A similar argument shows that $B + C \supseteq D + C$, so $B + C = D + C$. △

So the operation was well-defined here. This procedure looks promising, doesn't it? However, when we rewrote

$$((d+4m)+c)+4n=(d+c)+4(m+n)$$

we relied on the fact that addition commutes in an abelian group. Without that fact, we could not have swapped c and $4m$. Example 3.43 shows how it can go wrong.

EXAMPLE 3.43. Recall $A=\langle\varphi\rangle$ from Example 3.32; again, $A<D_3$. By the definition of the operation, we have

$$(\rho A)\left(\rho^2 A\right)=\left(\rho\circ\rho^2\right)A=\rho^3 A=\iota A=A.$$

Another representation of $\rho A=\{\rho\varphi,\rho\varphi^2\}$ is $(\rho\varphi)A$. If the operation is well-defined, then we should have $((\rho\varphi)A)\left(\rho^2A\right)=(\rho A)\left(\rho^2A\right)=A$. That is *not* the case:

$$((\rho\varphi)A)\left(\rho^2A\right)=\left((\rho\varphi)\rho^2\right)A=\left(\rho\left(\varphi\rho^2\right)\right)A=(\rho(\rho\varphi))A=\left(\rho^2\varphi\right)A\neq A.$$

△

The procedure described at the beginning of this section does *not* always result in an operation on cosets of non-abelian groups. Can we identify a condition on a subgroup that would guarantee that the procedure results in an operation?

The key in Example 3.42 was not really that $\mathbb{Z}$ is abelian. Rather, the key was that we could swap $4m$ and c in the expression $((d+4m)+c)+4m$. In a general group setting where $A<G$, for every $c\in G$ and for every $a\in A$ we would need to find $a'\in A$ to replace ac with ca'. The abelian property makes it easy to do that, but we don't *need* G to be abelian; we need A to satisfy this property.

Think about this again: for every $c\in G$ and for every $a\in A$, we want $a'\in A$ such that $ac=ca'$. That makes $cA\subseteq Ac$. The other direction must also be true, so $cA\supseteq Ac$. In other words,

The operation defined above is well-defined
iff $cA=Ac$ *for all* $c\in G$.

This property merits a definition.

DEFINITION 3.44. Let $A<G$. If

$$gA=Ag$$

for every $g\in G$, then A is a **normal subgroup** of G.

NOTATION. We write $A\triangleleft G$ to indicate that A is a normal subgroup of G.

An easy generalization of the argument of Example 3.42 shows the following Theorem.

THEOREM 3.45. *Let G be an abelian group, and $H<G$. Then $H\triangleleft G$.*

PROOF. You do it! See Exercise 3.52. □

We now present our first non-abelian normal subgroup.

EXAMPLE 3.46. Let

$$A_3=\left\{\iota,\rho,\rho^2\right\}<D_3.$$

We call A_3 the **alternating group** on three elements. We claim that $A_3\triangleleft D_3$. Indeed,

σ	σA_3	$A_3\sigma$
ι	A_3	A_3
ρ	A_3	A_3
ρ^2	A_3	A_3
φ	$\varphi A_3 = \{\varphi, \varphi\rho, \varphi\rho^2\} = \{\varphi, \rho^2\varphi, \rho\varphi\} = A_3\varphi$	$A_3\varphi = \varphi A_3$
$\rho\varphi$	$\{\rho\varphi, (\rho\varphi)\rho, (\rho\varphi)\rho^2\} = \{\rho\varphi, \varphi, \rho^2\varphi\} = \varphi A_3$	φA_3
$\rho^2\varphi$	$\{\rho^2\varphi, (\rho^2\varphi)\rho, (\rho^2\varphi)\rho^2\} = \{\rho^2\varphi, \rho\varphi, \varphi\} = \varphi A_3$	φA_3

(We have left out some details of the computation. You should check the computations very carefully, using extensively the fact that $\varphi\rho = \rho^2\varphi$.) Since A_3 is a normal subgroup of D_3, D_3/A_3 is a group. By Lagrange's Theorem, it has $6/3 = 2$ elements. The composition table is

$\circ$	A_3	φA_3
A_3	A_3	φA_3
φA_3	φA_3	A_3

Compare the operation table of D_3/A_3 to those of Examples 2.4 on page 17 and 2.26 on page 22.) $\triangle$

Normal subgroups allow us to turn the set of cosets into a group G/A.

THEOREM 3.47. *Let G be a group. If $A \triangleleft G$, then G/A is a group.*

PROOF. We show that G/A satisfies properties (G1)–(G4) of a group.

(G1): Closure follows from the fact that multiplication of cosets is well-defined when $A \triangleleft G$, as discussed earlier in this section: Let $X, Y \in G/A$, and choose $g_1, g_2 \in G$ such that $X = g_1A$ and $Y = g_2A$. Then $XY = (g_1A)(g_2A) = (g_1g_2)A \in G/A$.

(G2): The associative property follows from the associative property of the elements of the group. Let $X, Y, Z \in G/A$; choose $g_1, g_2, g_3 \in G$ such that $X = g_1A$, $Y = g_2A$, and $Z = g_3A$. Then

$$\begin{aligned}(XY)Z &= [(g_1A)(g_2A)](g_3A)\\ &= ((g_1g_2)A)(g_3A)\\ &= ((g_1g_2)g_3)A\\ &= (g_1(g_2g_3))A\\ &= (g_1A)((g_2g_3)A)\\ &= (g_1A)[(g_2A)(g_3A)]\\ &= X(YZ).\end{aligned}$$

(G3): The identity element is A itself. For any $X \in G/A$, choose $g \in G$ such that $X = gA$. Since $e \in A$, Lemma 3.24 on page 43 implies that $A = eA$, so

$$XA = (gA)(eA) = (ge)A = gA = X.$$

(G4): Let $X \in G/A$. Choose $g \in G$ such that $X = gA$; then

$$X \cdot \left(g^{-1}A\right) = (gA)\left(g^{-1}A\right) = \left(gg^{-1}\right)A = eA = A.$$

Hence X has an inverse in G/A.

$\square$

Theorem 3.47 tells us that the set of cosets of a normal subgroup is itself a group. This leads to a definition for a new kind of group.

DEFINITION 3.48. Let G be a group, and $A \triangleleft G$. Then G/A is **the quotient group of G with respect to A**, also called G **mod** A.

Normally we simply say "the quotient group" rather than "the quotient group of G with respect to A." We meet a very interesting and important quotient group in Section 3.5.

EXERCISES.

EXERCISE 3.49. Let $H = \langle \mathbf{i} \rangle < Q_8$.
(a) Show that $H \triangleleft Q_8$ by computing all the cosets of H.
(b) Compute the multiplication table of Q_8/H.

EXERCISE 3.50. Let $H = \langle -1 \rangle < Q_8$.
(a) Show that $H \triangleleft Q_8$ by computing all the cosets of H.
(b) Compute the multiplication table of Q_8/H.
(c) With which well-known group does Q_8/H have the same structure?

EXERCISE 3.51. Recall the subgroup L of $\mathbb{R}^2$ from Exercises 3.13 on page 39 and 3.27 on page 43.
(a) Explain why $L \triangleleft \mathbb{R}^2$.
(b) Sketch two elements of $\mathbb{R}^2/L$ and show their addition.

EXERCISE 3.52. Let G be an abelian subgroup. Explain why for any $H < G$ we know that $H \triangleleft G$.

EXERCISE 3.53. Explain why every subgroup of $D_m(\mathbb{R})$ is normal. (*Hint:* Theorem 3.45 tells us that the subgroup of an abelian group is normal. If you can show that $D_m(\mathbb{R})$ is abelian, then you are finished.)

EXERCISE 3.54. Show that Q_8 is not a normal subgroup of $\mathrm{GL}_m(\mathbb{R})$.

EXERCISE 3.55. Let G be a group. Define the **centralizer** of G as

$$Z(G) = \{g \in G : \ xg = gx \ \forall x \in G\}.$$

Show that $Z(G) \triangleleft G$. *Hint:* It is self-evident that $Z(G) \subseteq G$. You must show first that $Z(G) < G$. Then you must show that $Z(G) \triangleleft G$. Make sure that you separate these steps and justify each one carefully!

EXERCISE 3.56. Let G be a group, and $H < G$. Define the **normalizer** of H as

$$N_G(H) = \{g \in G : \ gH = Hg\}.$$

Show that $H \triangleleft N_G(H)$. *Hint:* First you must show that $H \subseteq N_G(H)$. Then you must show that $H < N_G(H)$. Finally you must show that $H \triangleleft N_G(H)$. Make sure that you separate these steps and justify each one carefully!

EXERCISE 3.57. Let G be a group, and $A < G$. Suppose that $|G/A| = 2$; that is, the subgroup A partitions G into precisely two left cosets. Show that $A \triangleleft G$. *Hint:* List the two left cosets, then the two right cosets. What does a partition mean? Given that, what sets must be equal?

3.5. A NEW GROUP

By Theorem 3.45, every subgroiup H of $\mathbb{Z}$ is normal. Let $n \in \mathbb{Z}$; since $n\mathbb{Z} < \mathbb{Z}$, it follows that $n\mathbb{Z} \triangleleft \mathbb{Z}$. Thus $\mathbb{Z}/n\mathbb{Z}$ is a quotient group.

We have made a lot of use of the subgroup $n\mathbb{Z}$ of $\mathbb{Z}$. One reason is that you are accustomed to working with $\mathbb{Z}$, so it is conceptually easy for you. Another part of it is that the quotient group $\mathbb{Z}/n\mathbb{Z}$ has important applications in number theory and computer science. You will see some of these applications in Chapters 6 and **??**. Because this group is so important, we give it several special names.

DEFINITION 3.58. Let $n \in \mathbb{Z}$. We call the quotient group $\mathbb{Z}/n\mathbb{Z}$

- $\mathbb{Z}$ mod $n\mathbb{Z}$, or
- $\mathbb{Z}$ **mod** n, or
- the linear residues **modulo** n.

NOTATION. It is common to write $\mathbb{Z}_n$ instead of $\mathbb{Z}/n\mathbb{Z}$.

This group has several different properties that are both interesting and powerful.

THEOREM 3.59. *$\mathbb{Z}_n$ is a finite group for every $n \in \mathbb{Z}$. In fact $\mathbb{Z}_n$ has n elements corresponding to the remainders from division by n: 0, 1, 2, …, $n-1$.*

Theorem 3.59 tells us not only show that $\mathbb{Z}_n$ is finite; it tells us how many elements are in $\mathbb{Z}_n$. It should not surprise you that the proof relies on the Division Theorem, since we said that the elements of $\mathbb{Z}_n$ correspond to the remainders from division by n. The structure of the proof is similar to the discussion in Example 3.31 on page 44, so you might want to go back and reread it.

PROOF. Let $n \in \mathbb{Z}$. To show that $\mathbb{Z}_n$ is finite, we will list its elements. Since $\mathbb{Z}_n$ is the set of cosets of $n\mathbb{Z}$, any element of $\mathbb{Z}_n$ has the form $a + n\mathbb{Z}$ for some $a \in \mathbb{Z}$.

Let $A \in n\mathbb{Z}$ and choose a such that $A = a + n\mathbb{Z}$. Use the Division Theorem to find $q, r \in \mathbb{Z}$ such that $a = qn + r$ and $0 \leq r < n$. Then

$$\begin{aligned} A &= a + n\mathbb{Z} \\ &= \{a + nz : z \in \mathbb{Z}\} \\ &= \{(qn + r) + nz : z \in \mathbb{Z}\} \\ &= \{r + n(q + z) : z \in \mathbb{Z}\} \\ &= \{r + nm : m \in \mathbb{Z}\} \\ &= r + n\mathbb{Z}. \end{aligned}$$

Thus A corresponds to a coset $r + n\mathbb{Z}$, where r is a remainder from division by n. Since A was arbitrary, *every* element of $\mathbb{Z}_n$ corresponds to a coset $r + n\mathbb{Z}$, where r is a remainder from division by n. How many remainders are there? The possible values are $0, 1, \ldots, n-1$, so all the elements of $\mathbb{Z}_n$ are $n\mathbb{Z}$, $1 + n\mathbb{Z}$, $2 + n\mathbb{Z}$, $\ldots$, $(n-1) + n\mathbb{Z}$. It follows that $\mathbb{Z}_n$ is finite. □

It is burdensome to write $a + n\mathbb{Z}$ whenever we want to discuss an element of $\mathbb{Z}_n$, so mathematicians usually adopt the following convention.

NOTATION. Let $A \in \mathbb{Z}_n$ and choose r such that $A = r + n\mathbb{Z}$ and $0 \leq r < n$.

- If it is clear from context that A is an element of $\mathbb{Z}_n$, then we simply write r instead of $r + n\mathbb{Z}$.
- If we want to emphasize that A is an element of $\mathbb{Z}_n$ (perhaps there are a lot of integers hanging about) then we write $[r]$ instead of $r + n\mathbb{Z}$.

To help you grow accustomed to the second notation $[r]$, we use it for the rest of this chapter, even when it is mind-bogglingly clear that we are talking about elements of $\mathbb{Z}_n$.

Since $\mathbb{Z}_n$ is finite, we can create the addition table for every $n \in \mathbb{Z}$. Since the representation of elements of $\mathbb{Z}_n$ is the remainder on division by n, we want $[a] + [b] = [r]$ where $0 \leq r < n$. For small numbers this isn't too hard. In $\mathbb{Z}_3$ for example,

$$[1] + [1] = (1 + 3\mathbb{Z}) + (1 + 3\mathbb{Z}) = (1 + 1) + 3\mathbb{Z} = [2].$$

But what should we do with larger sums, such as $[1] + [2]$? We don't want to write $[3]$, because 3 is not a valid remainder when we divide by 3.

LEMMA 3.60. *Let $n \in \mathbb{Z}$ and $[a], [b] \in \mathbb{Z}_n$. Use the Division Theorem to find $q, r \in \mathbb{Z}$ such that $a + b = qn + r$ and $0 \leq r < n$. Then*

$$[a] + [b] = [r].$$

PROOF.

$$\begin{aligned}
[a]+[b] &= (a+n\mathbb{Z})+(b+n\mathbb{Z}) \\
&= (a+b)+n\mathbb{Z} \\
&= \{(a+b)+nz:\ z\in\mathbb{Z}\} \\
&= \{(qn+r)+nz:\ z\in\mathbb{Z}\} \\
&= \{r+n(q+z):\ z\in\mathbb{Z}\} \\
&= \{r+nm:\ m\in\mathbb{Z}\} \\
&= r+n\mathbb{Z} \\
&= [r].
\end{aligned}$$

□

We can use Lemma 3.60 to build addition tables for $\mathbb{Z}_n$ easily.

It should be clear from Examples 2.4 on page 17 and 2.26 on page 22 as well as Exercises 2.16 on page 19 and 2.36 on page 25 that we learn nothing particularly insightful from the addition tables from $\mathbb{Z}_2$ and $\mathbb{Z}_3$, since the groups of order 2 and 3 are completely determined.

On the other hand, we saw in Exercises 2.17 on page 19 and 2.37 on page 25 that there are two possible structures for a group of order 4: the Klein 4-group, and a cyclic group. Which one is $\mathbb{Z}_4$?

EXAMPLE 3.61. Before building the table for $\mathbb{Z}_4$, we recall that it is abelian. Use Lemma 3.60 on the preceding page to observe that

$$\begin{aligned}
[1]+[3] &= [0] \\
[2]+[2] &= [0] \\
[2]+[3] &= [1] \\
[3]+[1] &= [0] \\
[3]+[2] &= [1] \\
[3]+[3] &= [2].
\end{aligned}$$

The addition table is thus

$+$	$[0]$	$[1]$	$[2]$	$[3]$
$[0]$	$[0]$	$[1]$	$[2]$	$[3]$
$[1]$	$[1]$	$[2]$	$[3]$	$[0]$
$[2]$	$[2]$	$[3]$	$[0]$	$[1]$
$[3]$	$[3]$	$[0]$	$[1]$	$[2]$

This is certainly not the Klein 4-group, since not every element is its own inverse. It must be the cyclic group of four elements, and in fact

$$\begin{aligned}
\langle[1]\rangle &= \{[1],[2],[3],[0]\} = \mathbb{Z}_4 \\
\langle[3]\rangle &= \{[3],[2],[1],[0]\} = \mathbb{Z}_4.
\end{aligned}$$

Not every non-zero element generates $\mathbb{Z}_4$, however, since

$$\langle[2]\rangle = \{[2],[0]\}.$$

△

The fact that $\mathbb{Z}_4$ was cyclic makes one wonder: is $\mathbb{Z}_n$ always cyclic? Yes!

THEOREM 3.62. *$\mathbb{Z}_n$ is cyclic for every $n \in \mathbb{Z}$.*

PROOF. Let $n \in \mathbb{Z}$. We claim that $\mathbb{Z}_n = \langle [1] \rangle$. *Why?* Let $x \in \mathbb{Z}_n$. Looking at Definition 2.43 on page 26, we need to show that $x = m\,[1]$ for some $m \in \mathbb{Z}$.

We can write $x = [r]$ for some $0 \leq r < n$. We proceed by induction on r.

Inductive base: If $r = 0$, then $x = [0] = 0 \cdot [1]$, so $x \in \langle [1] \rangle$.

Inductive hypothesis: Assume that for every $i = 0, 1, 2, \ldots, r-1$ we know that if $x = [i]$ then $x = i\,[1] \in \langle [1] \rangle$.

Inductive step: Since $r < n$, it follows from Lemma 3.60 that

$$[r] = [r-1] + [1] = (r-1)\,[1] + [1] = r\,[1] \in \langle [1] \rangle .$$

By induction, $x = [r] \in \langle [1] \rangle$. □

We saw in Example 3.61 that not *every* non-zero element necessarily generates $\mathbb{Z}_n$. A natural and interesting followup question to ask is, which non-zero elements *do* generate $\mathbb{Z}_n$? You need a bit more background in number theory before you can answer that question, but in the exercises you will build some more addition tables and use them to formulate a hypothesis.

EXERCISES.

EXERCISE 3.63. As discussed in the text, we know already that $\mathbb{Z}_2$ and $\mathbb{Z}_3$ are not very interesting, because their addition tables are predetermined. Since their addition tables should be easy to determine, go ahead and write out the addition tables for these groups.

EXERCISE 3.64. Write down the addition table for $\mathbb{Z}_5$. Which elements generate $\mathbb{Z}_5$?

EXERCISE 3.65. Write down the addition table for $\mathbb{Z}_6$. Which elements generate $\mathbb{Z}_6$?

EXERCISE 3.66. Compare the results of Example 3.61 and Exercises 3.63, 3.64, and 3.65. Formulate a conjecture as to which elements generate $\mathbb{Z}_n$. Do not try to prove your example.

CHAPTER 4

Isomorphisms

4.1. From functions to isomorphisms

We have on occasion observed that different groups have the same addition or multiplication table. We have also talked about different groups having the same structure: regardless of whether a group of order two is additive or multiplicative, its elements behave in exactly the same fashion. The groups may look superficially different because of their elements and operations, but the "group behavior" is identical.

Group theorists describe such a relationship between two groups as *isomorphic*. We aren't ready to give a precise definition of the term, but we can provide an intuitive definition:

If two groups G and H have identical group structure,

we say that G and H are **isomorphic**.

To define isomorphism precisely, we need to reconsider another topic that you studied in the past, functions.

Let G and H be groups. A mapping $f : G \to H$ is a **function** if for every input $x \in G$ the output $f(x)$ has precisely one value. In high school algera, you learned that this means that f passes the "vertical line test." The reader might suspect at this point—one could hardly blame you—that we are going to generalize the notion of function to something more general, just as we generalized $\mathbb{Z}$, $\mathrm{GL}_m(\mathbb{R})$, etc. to groups. To the contrary; we will *specialize* the notion of a function in a way that tells us important information about the group.

We want a function *that preserves the action of the operation* between the domain G and the range H. What does that mean? Let $x, y \in G$ and suppose that $f(x) = a$ and $f(y) = b$. Consider $z = xy$ and suppose that $f(z) = c$. If we are to preserve the operation:

- since $xy = z$,
- we want $ab = c$, or $f(x) f(y) = f(z)$.

Substituting $xy = z$ suggests that we should be interested in the property

$$f(x) f(y) = f(xy).$$

DEFINITION 4.1. Let G, H be groups and $f : G \to H$ a function. We say that f is a **group homomorphism**[1] from G to H if it satisfies the property that $f(x) f(y) = f(xy)$ for every $x, y \in G$.

NOTATION. You have to be careful with the fact that different groups have different operations. Depending on the context, the proper way to describe the homomorphism property may be

- $f(xy) = f(x) + f(y)$;
- $f(x + y) = f(x) f(y)$;

[1] The word comes Greek words that mean *common change*. Here the *change* that stays *common* is the effect of the operation on the elements of the group. The function shows that the group operation behaves the same way on elements of the range as on elements of the domain.

- $f(x \circ y) = f(x) \odot f(y)$;
- etc.

EXAMPLE 4.2. Let $f : \mathbb{Z} \to 2\mathbb{Z}$ by $f(x) = 4x$. Then f is a group homomorphism, since for any $x \in \mathbb{Z}$ we have

$$f(x) + f(y) = 4x + 4y = 4(x+y) = f(x+y).$$

△

The homomorphism property should remind you of certain special functions and operations that you have studied in Linear Algebra or Calculus. Recall from Exercise 2.33 that $\mathbb{R}^+$, the set of all positive real numbers, is a multiplicative group.

EXAMPLE 4.3. Let $f : (\mathrm{GL}_m(\mathbb{R}), \times) \to (\mathbb{R}^+, \times)$ by $f(A) = |\det A|$. An important fact from Linear Algebra tells us that for any two square matrices A and B, $\det A \det B = \det AB$. Thus

$$f(A) \cdot f(B) = |\det A| \cdot |\det B| = |\det A \cdot \det B| = |\det AB| = f(AB),$$

implying that f is a homomorphism of groups. △

Let's look at the new group that we studied in the previous section.

EXAMPLE 4.4. Let $n \in \mathbb{Z}$ such that $n \neq 0$, and let $f : (\mathbb{Z}, +) \to (\mathbb{Z}_n, +)$ by the assignment $f(x) = [r]$, where r is the remainder of the division of x by n. We claim that f is a homomorphism.

Why? Before giving a detailed, general explanation, let's look at an example. Suppse $n = 6$; then $f(-3) = [3]$ and $f(22) = [4]$. The operation in both the domain and the range is *addition,* so if f is a homomorphism, then we should observe the homomorphism property $f(x) f(y) = f(xy)$. In the context of these *additive* sets, that becomes $f(-3+22) = f(-3) + f(22)$. In fact,

$$f(-3+22) = f(19) = [1]$$
$$f(-3) + f(22) = [3] + [4] = [7] = [1].$$

This doesn't prove that f is a homomorphism, but it does give a good sign. It also gives us a hint at the general case: we will have to argue that congruence classes such as [7] and [1] are equal.

In general, let $x, y \in \mathbb{Z}$. Write $[a] = f(x)$ and $[b] = f(y)$. By definition of a and b, there exist $q_x, q_y \in \mathbb{Z}$ such that $x = q_x n + a$, $y = q_y n + b$, and $0 \leq a, b < n$. We need to show that $f(x+y) = [a+b] = [a] + [b] = f(x) + f(y)$.

Let $[r] \in \mathbb{Z}_n$ such that $[a] + [b] = [r]$. By notation, $(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = r + n\mathbb{Z}$. By definition of the quotient group, $(a+b) + n\mathbb{Z} = r + n\mathbb{Z}$. By Lemma 3.24 on page 43, $(a+b) - r \in n\mathbb{Z}$. By definition of $n\mathbb{Z}$, n divides $(a+b) - r$. Let $d \in \mathbb{Z}$ such that $nd = (a+b) - r$.

Again, we want to show that $f(x+y) = [a] + [b] = [r]$. That is, we want to show that two cosets are equal. We will try to apply Lemma 3.24 using the fact that $nd = (a+b) - r$, $x = q_x n + a$, and $y = q_y n + b$. Observe that

$$\begin{aligned} nd &= (a+b) - r \\ &= \big((x - q_x n) + \big(y - q_y n\big)\big) - r \\ &= ((x+y) - r) - \big(q_x + q_y\big) n. \end{aligned}$$

Thus

$$n\left(d+q_x+q_y\right)=(x+y)-r,$$

and by Lemma 3.24 $[x+y]=[r]$. Thus $r=[x+y]$ is the remainder after division of $x+y$ by n; hence $[r]=f(x+y)$, giving us $f(x+y)=[r]=f(x)+f(y)$.

We conclude that f is a homomorphism. △

The homomorphism property of preserving the operation guarantees that a homomorphism tells us an enormous amount of information about a group. The fact that they preserve the behavior of the operation shows that elements of the image of a group G act the same way as their preimages act in G.

On the other hand, it doesn't mean that the *group structure* is the same. In Example 4.4, for example, f is a homomorphism from an infinite group to a finite group; even if the group operations behave in a similar way, the groups themselves are inherently different. If we can show that the groups have the same "size" in addition to a similar operation, then the groups are, for all intents and purposes, identical. How do we decide that two groups have the same size? For finite groups, it is easy: count the elements. We can't do that for infinite groups, so we need something a little more general.

DEFINITION 4.5. Let $f: G \to H$ be a homomorphism of groups. If f is one-to-one and onto, then f is an **isomorphism**[2] and the groups G and H are **isomorphic**.[3] △

NOTATION. If the groups G and H are isomorphic, we write $G \cong H$.

You may not remember the definitions of one-to-one and onto, or you may not understand how to prove them, so we provide them here as a reference, along with two examples.

DEFINITION 4.6. Let $f: S \to U$ be a mapping of sets.

- We say that f is **one-to-one** when for every $a, b \in S$, if $f(a)=f(b)$ then $a=b$.
- We say that f is **onto** when for every $x \in U$, there exists an $a \in S$ such that $f(a)=x$.

Another way of saying that a function $f: S \to U$ is onto is to say that $f(S)=U$. Here, $f(S)$ is the **image** of the function f; that is, the set of all values in U that correspond via f to some element of S:

$$f(S)=\left\{u \in U : \exists s \in S \text{ such that } f(s)=u\right\}.$$

Thus the statement $f(S)=U$ means that *every* element of U corresponds via f to some element of S.

EXAMPLE 4.7. Recall the homomorphism of Example 4.2,

$$f: \mathbb{Z} \to 2\mathbb{Z} \quad \text{by} \quad f(x)=4x.$$

We show that f is one-to-one, but not onto.

That f is one-to-one: Let $a, b \in \mathbb{Z}$. Assume that $f(a)=f(b)$. By definition of f, $4a=4b$. Then $4(a-b)=0$; by the zero product property of the integers, $4=0$ or $a-b=0$. Since $4 \neq 0$, we must have $a-b=0$, or $a=b$.

We assumed $f(a)=f(b)$ and showed that $a=b$. Since a and b were arbitrary, f is one-to-one.

[2] The word comes Greek words that mean *identical change*.

[3] The standard method in set theory of showing that two sets are the same "size" is to show that there exists a one-to-one, onto function between the sets. For example, one can use this definition to show that $\mathbb{Z}$ and $\mathbb{Q}$ are the same size, but $\mathbb{Z}$ and $\mathbb{R}$ are not.

That f is not onto: There is no element $a \in \mathbb{Z}$ such that $f(a) = 2$. If there were, $4a = 2$. The only possible solution to this equation is $a = 1/2 \notin \mathbb{Z}$. △

EXAMPLE 4.8. Recall the homomorphism of Example 4.3,

$$f : \mathrm{GL}_m(\mathbb{R}) \to \mathbb{R}^+ \quad \text{by} \quad f(A) = |\det A|.$$

We claim that f is onto, but not one-to-one.

That f is not one-to-one: Observe that f maps each of the following two diagonal matrices to 0, even though the matrices are unequal:

$$A = 0 = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & \ddots \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots \end{pmatrix}.$$

(Unmarked entries are zeroes.)

That f is onto: Let $x \in \mathbb{R}^+$; then $f(A) = x$ where A is the diagonal matrix

$$A = \begin{pmatrix} x & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots \end{pmatrix}.$$

(Again, unmarked entries are zeroes.) △

We *cannot* conclude from these examples that $\mathbb{Z} \ncong 2\mathbb{Z}$ and that $\mathbb{R}^+ \ncong \mathbb{R}^{m\times n}$. *Why not?* In each case, we were considering only one of (possibly many) homomorphisms. It is quite possible that a different homomorphism would show that $\mathbb{Z} \cong 2\mathbb{Z}$ and that $\mathbb{R}^+ \cong \mathbb{R}^{m\times n}$. You will show in the exercises that the first assertion is in fact true, while the second is not.

We conclude this chapter with three important properties of homomorphisms. This result lays the groundwork for important results in later sections, and is generally useful.

THEOREM 4.9. *Let $f : G \to H$ be a homomorphism of groups. Denote the identity of G by e_G, and the identity of H by e_H. Then f*
preserves identities: $f(e_G) = e_H$; and
preserves inverses: for every $x \in G$, $f\left(x^{-1}\right) = f(x)^{-1}$.

Theorem 4.9 applies of course to isomorphisms as well. It should not surprise you that, if the operation's behavior is preserved, the identity is mapped to the identity, and inverses are mapped to inverses.

PROOF. *That f preserves identities:* Let $x \in G$. By the property of homomorphisms,

$$e_H f(x) = f(x) = f(e_G x) = f(e_G) f(x).$$

Thus

$$e_H f(x) = f(e_G) f(x).$$

Multiply both sides of the equation *on the right* by $f(x)^{-1}$ to obtain

$$e_H = f(e_G).$$

That f preserves inverses: Let $x \in G$. By the property of homomorphisms and by the fact that f preserves identity,

$$e_H = f(e_G) = f\left(x \cdot x^{-1}\right) = f(x) \cdot f\left(x^{-1}\right).$$

Thus

$$e_H = f(x) \cdot f\left(x^{-1}\right).$$

Pay careful attention to what this equation says! Since the product of $f(x)$ and $f\left(x^{-1}\right)$ is the identity, those two elements must be inverses! Hence $f\left(x^{-1}\right)$ is the inverse of $f(x)$, which we write as

$$f\left(x^{-1}\right) = f(x)^{-1}.$$

□

COROLLARY 4.10. *Let $f : G \to H$ be a homomorphism of groups. Then $f\left(x^{-1}\right)^{-1} = f(x)$ for every $x \in G$.*

PROOF. You do it! See Exercise 4.17. □

The following theorem is similar to the previous one, but has a different proof.

THEOREM 4.11. *Let $f : G \to H$ be a homomorphism of groups. Then f preserves powers of elements of G. That is, if $f(g) = h$, then $f(g^n) = f(g)^n = h^n$.*

PROOF. You do it! See Exercise 4.19. □

COROLLARY 4.12. *Let $f : G \to H$ be a homomorphism of groups. If $G = \langle g \rangle$ is a cyclic group, then $f(g)$ determines f completely. In other words, the image $f(G)$ is a cyclic group, and $f(G) = \langle f(g) \rangle$.*

PROOF. Since G is cyclic, for any $g' \in G$ there exists $n \in \mathbb{N}^+$ such that $g' = g^n$, and thus $f(g') = f(g^n) = f(g)^n$. □

EXERCISES.

EXERCISE 4.13.

(a) Show that $f : \mathbb{Z} \to 2\mathbb{Z}$ by $f(x) = 2x$ is an isomorphism. Hence $\mathbb{Z} \cong 2\mathbb{Z}$.
(b) Show that $\mathbb{Z} \cong n\mathbb{Z}$ for every nonzero integer n. *Hint:* Generalize the isomorphism of (a).

EXERCISE 4.14. Show that $\mathbb{Z}_2$ is isomorphic to the group of order two from Example 2.26 on page 22. *Caution!* Notice that the first group is usually written using addition, but the second group is multiplicative. Your proof should observe these distinctions.

EXERCISE 4.15. Show that $\mathbb{Z}_2$ is isomorphic to the Boolean xor group of Exercise 2.11 on page 18. *Caution!* Remember to denote the operation in the Boolean xor group correctly.

EXERCISE 4.16. Recall the subgroup L of $\mathbb{R}^2$ from Exercises 3.13 on page 39, 3.27 on page 43, and 3.51 on page 49. Show that $L \cong \mathbb{R}$. *Hint:* For a homomorphism function, think about the equation that describes the points on L.

EXERCISE 4.17. Prove Corollary 4.10. *Hint:* Since it's a corollary to Theorem 4.9, you should use that theorem.

EXERCISE 4.18. Let $f : G \to H$ be an isomorphism. Isomorphisms are by definition one-to-one functions, so f has an inverse function f^{-1}. Show that $f^{-1} : H \to G$ is also an isomorphism.

EXERCISE 4.19. Prove Theorem 4.11. *Hint:* Use induction on the positive powers of g; use a theorem for the nonpositive powers of g.

EXERCISE 4.20. Let $f : G \to H$ be a homomorphism of groups. Assume that G is abelian.
(a) Show that $f(G)$ is abelian.
(b) Is H abelian? Explain why or why not. *Hint:* Let $G = \mathbb{Z}_2$ and $H = D_3$; find a homomorphism from G to H.

EXERCISE 4.21. Let $f : G \to H$ be a homomorphism of groups. Let $A < G$. Show that $f(A) < H$. *Hint:* Recall that

$$f(A) = \{y \in H : f(x) = y \ \exists x \in A\},$$

and use the Subgroup Theorem.

EXERCISE 4.22. Let $f : G \to H$ be a homomorphism of groups. Let $A \triangleleft G$.
(a) Show that $f(A) \triangleleft f(G)$.
(b) Do you think that $f(A) \triangleleft H$? Justify your answer. *Hint:* See the last part of Exercise 4.20.

4.2. CONSEQUENCES OF ISOMORPHISM

In this section we provide a sequence of theorems that show that if two groups are isomorphic, then they are indistinguishable *as groups*. It may be that the elements of the sets are different, and the operation may be defined differently, but as groups the two are identical.

Our strategy will be to follow the general outline of the chapter to this point. Suppose that two groups G and H are isomorphic. We will show that

- G is abelian iff H is abelian;
- G is cyclic iff H is cyclic;
- every subgroup A of G corresponds to a unique subgroup A' of H (in particular, if A is of order n, so is A');
- every normal subgroup N of G corresponds to a unique normal subgroup N' of H;
- the quotient group G/N corresponds to a quotient group H/N'.

All of these depend on the fact that if $G \cong H$ then there exists an isomorphism $f : G \to H$. In particular, uniqueness is guaranteed only for any one isomorphism; if two different isomorphisms f, f' exist between G and H, then a subgroup A of G may very well correspond to two different subgroups B and B' of H.

THEOREM 4.23. *Suppose that $G \cong H$ as groups. Then G is abelian iff H is abelian.*

PROOF. Let $f : G \to H$ be an isomorphism. Assume that G is abelian. We must show that H is abelian. By Exercise 4.20, $f(G)$ is abelian. Since f is an isomorphism, and therefore onto, $f(G) = H$. Hence H is abelian.

A similar argument shows that if H is abelian, so is G. Hence G is abelian iff H is. □

THEOREM 4.24. *Suppose $G \cong H$ as groups. Then G is cyclic iff H is cyclic.*

PROOF. Let $f : G \to H$ be an isomorphism. Assume that G is cyclic. We must show that H is cyclic; that is, we must show that every element of H is generated by a fixed element of H.

Since G is cyclic, by definition $G = \langle g \rangle$ for some $g \in G$. Let $h = f(g)$; then $h \in H$. We claim that $H = \langle h \rangle$.

Let $x \in H$. Since f is an isomorphism, it is onto, so there exists $a \in G$ such that $f(a) = x$. Since G is cyclic, there exists $n \in \mathbb{Z}$ such that $a = g^n$. By Theorem 4.11,

$$x = f(a) = f(g^n) = f(g)^n = h^n.$$

Since x was an arbitrary element of H and x is generated by h, all elements of H are generated by h. Hence $H = \langle h \rangle$ is cyclic.

A similar proof shows that if H is cyclic, then so is G. □

THEOREM 4.25. *Suppose $G \cong H$ as groups. Every subgroup A of G corresponds to a subgroup A' of H. This correspondence is unique up to isomorphism. Moreover:*

(A) A is of finite order n iff A' is of finite order n.
(B) A is normal iff A' is normal.

PROOF. Let $f : G \to H$ be an isomorphism. Let A be a subgroup of G. By Exercise 4.21, $f(A) < H$. Let $A' = f(A)$; then $A' < H$. Uniqueness follows from the fact that f is one-to-one.

(A) follows from the fact that f is one-to-one.

For (B), assume $A \triangleleft G$. We want to show that $A' \triangleleft H$; that is, $xA' = A'x$ for every $x \in H$. So let $x \in H$ and $y \in A'$; since f is an isomorphism, it is onto, so $f(g) = x$ and $f(a) = y$ for some $g \in G$ and some $a \in A$. Then

$$xy = f(g) f(a) = f(ga).$$

Since $A \triangleleft G$, $gA = Ag$, so there exists $a' \in A$ such that $ga = a'g$. Let $y' = f(a')$. Thus

$$xy = f(a'g) = f(a') f(g) = y'x.$$

Notice that $y' \in f(A) = A'$, so $xy = y'x \in A'x$.

We have shown that for arbitrary $x \in H$ and arbitrary $y \in A'$, there exists $y' \in A'$ such that $xy = y'x$. Hence $xA' \subseteq A'x$. A similar argument shows that $xA' \supseteq A'x$, so $xA' = A'x$. This is the definition of a normal subgroup, so $A' \triangleleft H$.

A similar argument shows that if $A' \triangleleft H$, then its preimage $A = f^{-1}(A')$ is normal in G, as claimed. □

THEOREM 4.26. *Suppose $G \cong H$ as groups. Every quotient group of G is isomorphic to a quotient group of H.*

We use Lemma 3.24(CE3) on page 43 on coset equality heavily in this proof; you may want to go back and review it.

PROOF. Let $f : G \to H$ be an isomorphism. Let X be a quotient group of G defined by G/A, where $A \triangleleft G$. Let $A' = f(A)$; by Theorem 4.25 $A' \triangleleft H$, so H/A' is a quotient group. We want to show that $G/A \cong H/A'$.

Let $f_A : G/A \to H/A'$ by

$$f_A(X) = f(g)A' \qquad \text{where } gA = X \in G/A.$$

We claim that f_A is a well-defined homomorphism, and is one-to-one and onto.

That f_A is well-defined: Let $X \in G/A$ and consider two representations g_1A and g_2A of X. Then

$$f_A(g_1A) = f(g_1)A' \quad \text{and} \quad f_A(g_2A) = f(g_2)A'.$$

We must show that the cosets $f_A(g_1)A'$ and $f_A(g_2)A'$ are equal in H/A'. By hypothesis, $g_1A = g_2A$. Lemma 3.24(CE3) implies that $g_2^{-1}g_1 \in A$. Recall that $f(A) = A'$; this implies that $f\left(g_2^{-1}g_1\right) \in A'$. The homomorphism property implies that $f(g_2)^{-1}f(g_1) = f\left(g_2^{-1}\right)f(g_1) \in A'$. Lemma 3.24(CE3) again implies that $f(g_1)A' = f(g_2)A'$. In other words,

$$f_A(X) = f(g_1)A' = f(g_2)A'$$

so there is no ambiguity in the definition of f_A as to the image of X in H/A'; the function is well-defined.

That f_A is a homomorphism: Let $X, Y \in G/A$ and consider write $X = g_1A$ and $Y = g_2A$ for appropriate $g_1, g_2 \in G$. Now

$$\begin{aligned}
f_A(XY) &= f_A((g_1A)\cdot(g_2A)) \\
&= f_A(g_1g_2\cdot A) \\
&= f(g_1g_2)A' \\
&= f(g_1)f(g_2)\cdot A' \\
&= f(g_1)A'\cdot f(g_2)A' \\
&= f_A(g_1A)\cdot f_A(g_2A) \\
&= f_A(X)\cdot f_A(Y)
\end{aligned}$$

where each equality is justified by (respectively) the definitions of X and Y; the definition of coset multiplication in G/A; the definition of f_A; the homomorphism property of f; the definition of coset multiplication in H/A'; the definition of f_A; and the definitions of X and Y. The chain of equalities shows clearly that f_A is a homomorphism.

That f_A is one-to-one: Let $X, Y \in G/A$ and assume that $f_A(X) = f_A(Y)$. Let $g_1, g_2 \in G$ such that $X = g_1A$ and $Y = g_2A$. The definition of f_A implies that

$$f(g_1)A' = f_A(X) = f_A(Y) = f(g_2)A',$$

so by Lemma 3.24(CE3) $f(g_2)^{-1}f(g_1) \in A'$. Recall that $A' = f(A)$, so there exists $a \in A$ such that $f(a) = f(g_2)^{-1}f(g_1)$. The homomorphism property implies that

$$f(a) = f\left(g_2^{-1}\right)f(g_1) = f\left(g_2^{-1}g_1\right).$$

Recall that f is an isomorphism, hence one-to-one. The definition of one-to-one implies that

$$g_2^{-1}g_1 = a \in A.$$

Applying Lemma 3.24(CE3) again gives us $g_1A = g_2A$, and

$$X = g_1A = g_2A = Y.$$

We took arbitrary $X, Y \in G/A$ and showed that if $f_A(X) = f_A(Y)$, then $X = Y$. It follows that f_A is one-to-one.

That f_A is onto: You do it! See Exercise 4.27. □

EXERCISES.

EXERCISE 4.27. Show that the function f_A defined in the proof of Theorem 4.26 is onto. *Hint:* It's quite a bit easier than the proof that f_A is one-to-one.

EXERCISE 4.28. Recall from Exercise 2.49 on page 29 that $\langle \mathbf{i} \rangle$ is a cyclic group of Q_8.

(a) Show that $\langle \mathbf{i} \rangle \cong \mathbb{Z}_4$ by giving an explicit isomorphism.
(b) Let A be a proper subgroup of $\langle \mathbf{i} \rangle$. Find the corresponding subgroup of $\mathbb{Z}_4$.
(c) Use the proof of Theorem 4.26 to determine the quotient group of $\mathbb{Z}_4$ to which $\langle \mathbf{i} \rangle / A$ is isomorphic.

EXERCISE 4.29. Recall from Exercise 4.16 on page 58 that the set

$$L = \left\{ x \in \mathbb{R}^2 : \; x = (a, a) \; \exists a \in \mathbb{R} \right\}$$

defined in Exercise 3.13 on page 39 is isomorphic to $\mathbb{R}$.

(a) Show that $\mathbb{Z} \triangleleft \mathbb{R}$.
(b) Give the precise definition of $\mathbb{R}/\mathbb{Z}$.
(c) Explain why we can think of $\mathbb{R}/\mathbb{Z}$ as the set of classes $[a]$ such that $a \in [0, 1)$. Choose one such $[a]$ and describe the elements of this class.
(c) Find the subgroup A of L that corresponds to $\mathbb{Z} < \mathbb{R}$. What do this section's theorems imply that you can conclude about A and L/A? *Hint:* Use the isomorphism you developed in Exercise 4.16 on page 58.
(d) Use the answer to (c) to describe L/A intuitively. Choose an element of L/A and describe the elements of this class.

4.3. THE ISOMORPHISM THEOREM

Because quotient groups provide imporant information about a group, algebraists study them extensively. However, the nature of cosets makes quotient groups difficult to grasp intuitively. To get a better handle on them, algebraists try to identify other groups that are isomorphic to the quotient group of interest.

EXAMPLE 4.30. Recall $A_3 = \{\iota, \rho, \rho^2\} \triangleleft D_3$ from Example 3.46. We saw that D_3/A_3 has only two elements, so it must be isomorphic to the group of two elements. First we show this explicitly: Let $\mu : D_3/A_3 \to \mathbb{Z}_2$ by

$$\mu(X) = \begin{cases} 0, & X = A_3; \\ 1, & \text{otherwise.} \end{cases}$$

Is μ a homomorphism? Recall that A_3 is the identity element of D_3/A_3, so for any $X \in D_3/A_3$

$$\mu(X \cdot A_3) = \mu(X) = \mu(X) + 0 = \mu(X) + \mu(A_3).$$

This verifies the homomorphism property for all products in the operation table of D_3/A_3 except $(\varphi A_3, \varphi A_3)$, which is easy to check:

$$\mu((\varphi A_3)\cdot(\varphi A_3)) = \mu(A_3) = 0 = 1+1 = \mu(\varphi A_3)+\mu(\varphi A_3).$$

Hence μ is a homomorphism. The property of isomorphism follows from the facts that

- $\mu(A_3) \neq \mu(\varphi A_3)$, so μ is one-to-one, and
- both 0 and 1 have preimages, so μ is onto.

Something subtle is at work here. Let $f : D_3 \to \mathbb{Z}_2$ by

$$f(x) = \begin{cases} 0, & x \in A_3; \\ 1, & \text{otherwise.} \end{cases}$$

Is f a homomorphism? The elements of A_3 are ι, ρ, and ρ^2; f maps these elements to zero, and the other three elements of S_3 to 1. Let $x, y \in D_3$ and consider the various cases:

Case 1. $x, y \in A_3$.
Since A_3 is a group, closure implies that $xy \in A_3$. Thus

$$f(xy) = 0 = 0+0 = f(x)+f(y).$$

Case 2. $x \in A_3$ and $y \notin A_3$.
Since A_3 is a group, closure implies that $xy \notin A_3$. (Otherwise $xy = z$ for some $z \in A_3$, and multiplication by the inverse implies that $y = x^{-1}z \in A_3$, a contradiction.) Thus

$$f(xy) = 1 = 0+1 = f(x)+f(y).$$

Case 3. $x \notin A_3$ and $y \in A_3$.
An argument similar to the case above shows that $f(xy) = f(x)+f(y)$.

Case 4. $x, y \notin A_3$.
Inspection of the operation table of D_3 (Exercise 2.63 on page 35) shows that $xy \in A_3$. Hence

$$f(xy) = 0 = 1+1 = f(x)+f(y).$$

We have shown that f is a homomorphism from D_3 to $\mathbb{Z}_2$.

In addition, consider the function $\eta : D_3 \to D_3/A_3$ by

$$\eta(x) = \begin{cases} A_3, & x \in A_3; \\ \varphi + A_3, & \text{otherwise.} \end{cases}$$

It is easy to show that this is a homomorphism; we do so presently.

Now comes the important observation: Look at the composition function $\eta \circ \mu$ whose domain is D_3 and whose range is $\mathbb{Z}_2$:

$$\begin{aligned}
(\mu\circ\eta)(\iota) &= \mu(\eta(\iota)) = \mu(A_3) = 0; \\
(\mu\circ\eta)(\rho) &= \mu(\eta(\rho)) = \mu(A_3) = 0; \\
(\mu\circ\eta)\left(\rho^2\right) &= \mu\left(\eta\left(\rho^2\right)\right) = \mu(A_3) = 0; \\
(\mu\circ\eta)(\varphi) &= \mu(\eta(\varphi)) = \mu(\varphi + A_3) = 1; \\
(\mu\circ\eta)(\rho\varphi) &= \mu(\eta(\rho\varphi)) = \mu(\varphi + A_3) = 1; \\
(\mu\circ\eta)\left(\rho^2\varphi\right) &= \mu\left(\eta\left(\rho^2\varphi\right)\right) = \mu(\varphi + A_3) = 1.
\end{aligned}$$

We have

$$(\mu \circ \eta)(x) = \begin{cases} 0, & x \in A_3; \\ 1, & \text{otherwise,} \end{cases}$$

or in other words

$$\mu \circ \eta = f.$$

△

This remarkable correspondence makes it easier to study quotient groups G/A:

- find a group H that is "easy" to work with; and
- find a homomorphism $f : G \to H$ such that
  - $f(g) = e_H$ for all $g \in A$, and
  - $f(g) \neq e_H$ for all $g \notin A$.

If we can do this, then $H \cong G/A$, and as we saw in Section 4.2 studying G/A is equivalent to studying H.

The reverse is also true: a group G and its quotient groups are relatively easy to study, whereas another group H is difficult. The isomorphism theorem helps us identify a quotient group G/A that is isomorphic to H, making it easier to study.

We need to formalize this observation in a theorem, but first we have to confirm something that we claimed earlier:

LEMMA 4.31. *Let G be a group and $A \triangleleft G$. The function $\eta : G \to A$ by*

$$\eta(g) = gA$$

is a homomorphism.

PROOF. You do it! See Exercise 4.35. □

DEFINITION 4.32. We call the homomorphism η of Lemma 4.31 the **natural homomorphism**.

We need another definition, which you might remember from linear algebra. It will prove important in subsequent sections and chapters.

DEFINITION 4.33. Let G and H be groups, and $f : G \to H$ a homomorphism. Let $Z = \{g \in G : f(g) = e_H\}$; that is, Z is the set of all elements of G that f maps to the identity of H. We call Z the **kernel** of f, written $\ker f$.

We now formalize the observation of Example 4.30.

THEOREM 4.34 (The Isomorphism Theorem). *Let G and H be groups, and $A \triangleleft G$. Let $\eta : G \to A$ be the natural homomorphism. If there exists a homomorphism $f : G \to H$ such that f is onto and $\ker f = A$, then $G/A \cong H$. Moreover, the isomorphism $\mu : G/A \to H$ satisfies $f = \mu \circ \eta$.*

We can illustrate Theorem 4.34 with the following diagram:

$$\begin{array}{ccc} G & \xrightarrow{f} & H \\ & {\scriptstyle\eta}\searrow \quad \nearrow{\scriptstyle\mu} & \\ & G/A & \end{array}$$

The idea is that "the diagram commutes", or $f = \mu \circ \eta$.

PROOF. We are given G, H, A, and η. Assume that there exists a homomorphism $f : G \to H$ such that $\ker f = A$. Define $\mu : G/A \to H$ in the following way:

$$\mu(x) = \begin{cases} e_H, & x = A; \\ f(g), & x = gA \quad \exists g \notin A. \end{cases}$$

We claim that μ is an isomorphism from G/A to H, and moreover that $f = \mu \circ \eta$.

Since the domain of μ consists of cosets which may have different representations, we must show first that μ is well-defined. Suppose that $X \in G/A$ has two different representations $X = gA = g'A$ where $g, g' \in G$ and $g \neq g'$. We need to show that $\mu(gA) = \mu(g'A)$. From Lemma 3.24(CE3), we know that $g^{-1}g' \in A$, so there exists $a \in A$ such that $g^{-1}g' = a$, so $g' = ga$. Applying the homomorphism property,

$$\mu(g'A) = f(g') = f(ga) = f(g) f(a).$$

Recall that $a \in A = \ker f$, so

$$\mu(g'A) = f(g) \cdot e_H = f(g) = \mu(gA).$$

Hence $\mu(g'A) = \mu(gA)$ and $\mu(X)$ is well-defined.

Is μ a homomorphism? Let $X, Y \in G/A$; we can represent $X = gA$ and $Y = g'A$ for some $g, g' \in G$. Applying the homomorphism property of f, we see that

$$\mu(XY) = \mu((gA)(g'A)) = \mu((gg')A) = f(gg') = f(g) f(g') = \mu(gA)\mu(g'A).$$

Thus μ is a homomorphism.

Is μ one-to-one? Let $X, Y \in G/A$ and assume that $\mu(X) = \mu(Y)$. Represent $X = gA$ and $Y = g'A$ for some $g, g' \in G$; by the homomorphism property of f, we see that

$$\begin{aligned} f(g^{-1}g') &= f(g^{-1}) f(g') \\ &= f(g)^{-1} f(g') \\ &= \mu(gA)^{-1} \mu(g'A) \\ &= \mu(X)^{-1} \mu(Y) \\ &= \mu(Y)^{-1} \mu(Y) \\ &= e_H, \end{aligned}$$

so $g^{-1}g' \in \ker f$. It is given that $\ker f = A$, so $g^{-1}g' \in A$. Lemma 3.24(CE3) now tells us that $gA = g'A$, so $X = Y$. Thus μ is one-to-one.

Is μ onto? Let $h \in H$; we need to find an element $X \in G/A$ such that $\mu(X) = h$. It is given that f is onto, so there exists $g \in G$ such that $f(g) = h$. Then

$$\mu(gA) = f(g) = h,$$

so μ is onto.

We have shown that μ is an isomorphism; we still have to show that $f = \mu \circ \eta$, but the definition of μ makes this trivial: for any $g \in G$,

$$(\mu \circ \eta)(g) = \mu(\eta(g)) = \mu(gA) = f(g).$$

□

EXERCISES.

EXERCISE 4.35. Prove Lemma 4.31.

EXERCISE 4.36. Recall the normal subgroup L of $\mathbb{R}^2$ from Exercises 3.13 on page 39, 3.27 on page 43, and 3.51 on page 49. In Exercise 4.16 on page 58 you found an explicit isomorphism $L \cong \mathbb{R}$.

(a) Use the Isomorphism Theorem to find an isomorphism $\mathbb{R}^2/L \cong \mathbb{R}$. *Hint:* Consider $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(a) = b$ where the point $a = (a_1, a_2)$ lies on the line $y = x + b$.
(b) Describe geometrically how the cosets of L are mapped to elements of $\mathbb{R}$.

EXERCISE 4.37. Recall the normal subgroup $\langle -1 \rangle$ of Q_8 from Exercises 2.38 on page 25 and 3.50 on page 49.

(a) Use Lagrange's Theorem to explain why $Q_8/\langle -1 \rangle$ has order 4.
(b) We know from Exercise 2.17 on page 19 that there are only two groups of order 4, the Klein 4-group and the cyclic group of order 4, which we can represent by $\mathbb{Z}_4$. Use the Isomorphism Theorem to determine which of these groups is isomorphic to $Q_8/\langle -1 \rangle$. *Hint:* You already know the answer from Exercise 3.50 on page 49; find a homomorphism f from Q_8 to that group such that $\ker f = \langle -1 \rangle$.

4.4. AUTOMORPHISMS AND GROUPS OF AUTOMORPHISMS

In this final section of Chapter 4, we use a special kind isomorphism to build a new group.

DEFINITION 4.38. Let G be a group. If $f : G \to G$ is an isomorphism, then we call f an **automorphism**.[4]

An automorphism is an isomorphism whose domain and range are the same set. Thus, to show that some function f is an automorphism, you must show first that the domain and the range of f are the same set. Afterwards, you show that f satisfies the homomorphism property, and then that it is both one-to-one and onto.

EXAMPLE 4.39.

(a) An easy automorphism for any group G is the identity isomorphism $\iota(g) = g$:
- its range is by definition G;
- it is a *homomorphism* because $\iota(g \cdot g') = g \cdot g' = \iota(g) \cdot \iota(g')$;
- it is *one-to-one* because $\iota(g) = \iota(g')$ implies (by evaluation of the function) that $g = g'$; and
- it is *onto* because for any $g \in G$ we have $\iota(g) = g$.

(b) An automorphism in $(\mathbb{Z}, +)$ is $f(x) = -x$:
- its range is $\mathbb{Z}$ because of closure;
- it is a *homomorphism* because $f(x+y) = -(x+y) = -x - y = f(x) + f(y)$;
- it is *one-to-one* because $f(x) = f(y)$ implies that $-x = -y$, so $x = y$; and
- it is *onto* because for any $x \in \mathbb{Z}$ we have $f(-x) = x$.

(c) An automorphism in D_3 is $f(x) = \rho^2 x \rho$:

[4]The word comes Greek words that mean *self* and *change*.

- its range is D_3 because of closure;
- it is a *homomorphism* because $f(xy) = \rho^2(xy)\rho = \rho^2(x\cdot\iota\cdot y)\rho = \rho^2(x\cdot\rho^3\cdot y)\rho = (\rho^2x\rho)\cdot(\rho^2y\rho) = f(x)\cdot f(y)$;
- it is *one-to-one* because $f(x) = f(y)$ implies that $\rho^2x\rho = \rho^2y\rho$, and multiplication on the left by ρ and on the right by ρ^2 gives us $x = y$; and
- it is *onto* because for any $y \in D_3$, choose $x = \rho y\rho^2$ and then $f(x) = \rho^2(\rho y\rho^2)\rho = (\rho^2\rho)\cdot y\cdot(\rho^2\rho) = \iota\cdot y\cdot\iota = y$. △

The automorphism of Example 4.39(c) generalizes to an important automorphism.

DEFINITION 4.40. Let G be a group and $a \in G$. Define the function of **conjugation by** a to be $\text{conj}_a(x) = a^{-1}xa$. △

In Example 4.39(c), we had $a = \rho$ and $\text{conj}_a(x) = a^{-1}xa = \rho^2x\rho$.

LEMMA 4.41. *Let G be a group, and $a \in G$. Then* conj_a *is an automorphism. Moreover,*

$$\{\text{conj}_a(g) : g \in G\} < G.$$

PROOF. You do it! See Exercise 4.49. □

The subgroup $\{\text{conj}_a(g) : g \in G\}$ is important enough to identify by a special name.

DEFINITION 4.42. We say that $\{\text{conj}_a(g) : g \in G\}$ is the **group of conjugations of** G **by** a, and denote it by $\text{Conj}_a(G)$. △

Conjugation of subgroups is *not* necessarily an automorphism; it is quite possible that for some $H < G$ and for some $a \in G\backslash H$ we do *not* have $H = \{\text{conj}_a(h) : h \in H\}$. On the other hand, if H is a *normal* subgroup of G then we *do* have $H = \{\text{conj}_a(h) : h \in H\}$. You will explore this in the exercises.

Now it is time to identify the new group that we promised at the beginning of the chapter.

NOTATION. Write $\text{Aut}(G)$ for the set of all automorphisms of G. In addition, we typically denote automorphisms by Greek letters, rather than Latin letters.

EXAMPLE 4.43. We compute $\text{Aut}(\mathbb{Z}_4)$. Let $\alpha \in \text{Aut}(\mathbb{Z}_4)$ be arbitrary; what do we know about α? We know by definition that its range is $\mathbb{Z}_4$, and by Theorem 4.9 on page 57 we know that $\alpha(0) = 0$. Aside from that, we consider all the possibilities that preserve the isomorphism properties.

Recall from Theorem 3.62 on page 53 that $\mathbb{Z}_4$ is a cyclic group; in fact $\mathbb{Z}_4 = \langle 1\rangle$. Corollary 4.12 on page 58 tells us that $\alpha(1)$ will tell us everything we want to know about α. So, what can $\alpha(1)$ be?

Case 1. Can we have $\alpha(1) = 0$? If so, then $\alpha(n) = 0$ for all $n \in \mathbb{Z}_4$. This is not one-to-one, so we cannot have $\alpha(1) = 0$.

Case 2. Can we have $\alpha(1) = 1$? Certainly $\alpha(1) = 1$ if α is the identity homomorphism ι, so we can have $\alpha(1) = 1$.

Case 3. Can we have $\alpha(1) = 2$? If so, then the homomorphism property implies that

$$\alpha(2) = \alpha(1+1) = \alpha(1) + \alpha(1) = 4 = 0.$$

An automorphism must be a homomorphism, but if $\alpha(1) = 2$ then α is not one-to-one: by Theorem 4.9 on page 57, $\alpha(0) = 0 = \alpha(2)$! So we *cannot* have $\alpha(1) = 2$.

Case 4. Can we have $\alpha(1) = 3$? If so, then the homomorphism property implies that

$$\alpha(2) = \alpha(1+1) = \alpha(1) + \alpha(1) = 3+3 = 6 = 2; \text{ and}$$
$$\alpha(3) = \alpha(2+1) = \alpha(2) + \alpha(1) = 2+3 = 5 = 1.$$

In this case, α is both one-to-one *and* onto. We were careful to observe the homomorphism property when determining α, so we know that α is a homomorphism. So we *can* have $\alpha(1) = 2$.

We found only two possible elements of $\mathrm{Aut}(\mathbb{Z}_4)$: the identity automorphism and the automorphism determined by $\alpha(1) = 3$. △

It turns out that $\mathrm{Aut}(G)$ is itself a group!

LEMMA 4.44. *For any group G, $\mathrm{Aut}(G)$ is a group under the operation of composition of functions.*

PROOF. Let G be any group. We show that $\mathrm{Aut}(G)$ satisfies each of the group properties from Definition 2.27.

(G1) Let $\alpha, \theta \in \mathrm{Aut}(G)$. We must show that $\alpha \circ \theta \in \mathrm{Aut}(G)$ as well:
- the domain and range of $\alpha \circ \theta$ are both G because the domain and range of both α and θ are both G;
- $\alpha \circ \theta$ is a *homomorphism* because for any $g, g' \in G$ we can apply the homomorphism property that applies to α and θ to obtain

$$\begin{aligned}(\alpha \circ \theta)(g \cdot g') &= \alpha(\theta(g \cdot g')) \\ &= \alpha(\theta(g) \cdot \theta(g')) \\ &= \alpha(\theta(g)) \cdot \alpha(\theta(g')) \\ &= (\alpha \circ \theta)(g) \cdot (\alpha \circ \theta)(g');\end{aligned}$$

- $\alpha \circ \theta$ is *one-to-one* because $(\alpha \circ \theta)(g) = (\alpha \circ \theta)(g')$ implies that $\alpha(\theta(g)) = \alpha(\theta(g'))$; since α is one-to-one we infer that $\theta(g) = \theta(g')$; since θ is one-to-one we conclude that $g = g'$; and
- $\alpha \circ \theta$ is *onto* because for any $z \in G$,
  - α is onto, so there exists $y \in G$ such that $\alpha(y) = z$, and
  - θ is onto, so there exists $x \in G$ such that $\theta(x) = y$, so
  - $(\alpha \circ \theta)(x) = \alpha(\theta(x)) = \alpha(y) = z$.

We have shown that $\alpha \circ \theta$ satisfies the properties of an automorphism; hence, $\alpha \circ \theta \in \mathrm{Aut}(G)$, and $\mathrm{Aut}(G)$ is closed under the composition of functions.

(G2) The associative property is sastisfied because the operation is composition of functions, which is associative.

(G3) Denote by ι the identity homomorphism; that is, $\iota(g) = g$ for all $g \in G$. We showed in Example 4.39(a) that ι is an automorphism, so $\iota \in \mathrm{Aut}(G)$. Let $f \in \mathrm{Aut}(G)$; we claim that $\iota \circ f = f \circ \iota = f$. Let $x \in G$ and write $f(x) = y$. We have

$$(\iota \circ f)(x) = \iota(f(x)) = \iota(y) = y = f(x),$$

and likewise $(f \circ \iota)(x) = f(x)$. Since x was arbitrary in G, we have $\iota \circ f = f \circ \iota = f$.

(G2) Let $\alpha \in \mathrm{Aut}(G)$. Since α is an automorphism, it is an isomorphism. You showed in Exercise 4.18 that α^{-1} is also an isomorphism. The domain and range of α are both G, so the domain and range of α^{-1} are also both G. Hence $\alpha^{-1} \in \mathrm{Aut}(G)$.

□

Since $\text{Aut}(G)$ is a group, we can compute $\text{Aut}(\text{Aut}(G))$. For finite groups, $|\text{Aut}(G)| < |G|$ (we do not prove this here) so any chain of automorphism groups must eventually stop. In the exercises you will compute $\text{Aut}(G)$ for some other groups.

EXAMPLE 4.45. Recall from Example 4.43 on page 67 that $\text{Aut}(\mathbb{Z}_4)$ has only two elements. We saw early on that there is only one group of two elements, so $\text{Aut}(\mathbb{Z}_4) \cong \mathbb{Z}_2$. _____△

EXERCISES.

EXERCISE 4.46. Show that $f(x) = x^2$ is an automorphism on the group $(\mathbb{R}^+, \times)$.

EXERCISE 4.47.
(a) List the elements of $\text{Conj}_\rho(D_3)$.
(b) List the elements of $\text{Conj}_\varphi(D_3)$.
(c) Will $\text{Conj}_a(G)$ always be a *normal* subgroup of G?

EXERCISE 4.48. List the elements of $\text{Conj}_\mathbf{i}(Q_8)$.

EXERCISE 4.49. Prove Lemma 4.41 on page 67 in two parts:
(a) Show first that conj_g is an automorphism.
(b) Show that $\{\text{conj}_a(g) : g \in G\}$ is a group.
Hint: Use some of the ideas from Example 4.39 on page 66(c).

EXERCISE 4.50. Determine the automorphism group of the Klein 4-group.

EXERCISE 4.51. Determine the automorphism group of D_3. *Hint:* We can think of D_3 as generated by the elements ρ and φ, and each of these generates a non-trivial cyclic subgroup. Any automorphism α is therefore determined by these generators, so you can find all automorphisms α by finding all possible results for $\alpha(\rho)$ and $\alpha(\varphi)$, then examining that carefully.

EXERCISE 4.52. Let G be a group, $g \in G$, and $H < G$. Write $g^{-1}Hg = \{\text{conj}_g(h) : h \in H\}$. Show that $H \triangleleft G$ iff for every $g \in G$ we have $H = g^{-1}Hg$. *Hint:* This problem requires you to show twice that two sets are equal.

CHAPTER 5

Groups of permutations

5.1. PERMUTATIONS; TABULAR NOTATION; CYCLE NOTATION

Certain applications of mathematics involve the rearrangement of a list of n elements. It is common to refer to such rearrangements as *permutations*.

DEFINITION 5.1. Let V be any finite list. A **permutation** is a one-to-one function whose domain and range are both V.

We say that V is a list rather than a set, because the order of the elements matters: $(a,d,k,r) \neq (a,k,d,r)$ even though $\{a,d,k,r\} = \{a,k,d,r\}$. For the sake of convenience we usually write V as a list of integers between 1 and $|V|$, but it can be any finite list.

EXAMPLE 5.2. Let $S = (a,d,k,r)$. To denote the ith element of the list, we write s_i. So $s_1 = a$, $s_2 = d$, etc. Define a permutation on the elements of S by

$$f(x) = \begin{cases} r, & x = a; \\ a, & x = d; \\ k, & x = k; \\ d, & x = r. \end{cases}$$

Notice that f is one-to-one, and $f(S) = (r,a,k,d)$.

We can represent the same permutation on a generic list of four elements $V = (1,2,3,4)$. Define a permutation on the elements of V by

$$\pi(i) = \begin{cases} 2, & i = 1; \\ 4, & i = 2; \\ 3, & i = 3; \\ 1, & i = 4. \end{cases}$$

Here π is one-to-one, and $\pi(i) = j$ is interpreted as "the jth element of the permuted list is the ith element of the original list." You could visualize this as

position in original list i		position in permuted list j
1	$\rightarrow$	2
2	$\rightarrow$	4
3	$\rightarrow$	3
4	$\rightarrow$	1

Thus $\pi(V) = (4,1,3,2)$. If you look back at $f(S)$, you will see that in fact the first element of the permuted list, $f(S)$, is the fourth element of the original list, S. $\triangle$

In Section 5.2 we show that the set of all permutations of a set of n elements is a group. In this section we go over the necessary prelimiaries of this study, considering especially the question of how to write them. We need the following lemma.

LEMMA 5.3. *The composition of two permutations is a permutation.*

PROOF. Let V be a set of n elements, and α, β permutations of V. Let $\gamma = \alpha \circ \beta$. We claim that γ is a permutation. To show this, we must show that γ is a one-to-one function whose domain and range are both V. From the definition of α and β, it follows that the domain and range of γ are both V; it remains to show that γ is one-to-one. Let $x, y \in V$ and assume that $\gamma(x) = \gamma(y)$; by definition of γ,

$$\alpha(\beta(x)) = \alpha(\beta(y)).$$

As permutations, α and β are one-to-one functions, and as such have inverse functions. Thus

$$\beta^{-1}\left(\alpha^{-1}(\alpha(\beta(x)))\right) = \beta^{-1}\left(\alpha^{-1}(\alpha(\beta(y)))\right)$$
$$x = y.$$

Hence γ is a one-to-one function whose domain and range are both V; in other words, a permutation. □

In Example 5.2, we wrote a permutation as a piecewise function. This is burdensome; we would like a more efficient way to denote permutations.

NOTATION. The **tabular notation** for a permutation on a list of n elements is a $2 \times n$ matrix

$$\alpha = \begin{pmatrix} 1 & 2 & \cdots & n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}$$

denoting that $\alpha(1) = \alpha_1$, $\alpha(2) = \alpha_2$, …, $\alpha(n) = \alpha_n$. Again, $\alpha(i) = j$ indicates that the jth element of the permuted list is the ith element of the original list.

EXAMPLE 5.4. Recall V and π from Example 5.2. We can also write

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

because π moves

- the element in the first position to the second position;
- the element in the second position to the fourth position;
- the element in the third position nowhere; and
- the element in the fourth position to the first position.

Then

$$\pi(1,2,3,4) = (4,1,3,2).$$

Notice that the tabular notation for π looks similar to the table in Example 5.2.

We can also use π to permute different lists, so long as the new lists have four elements:

$$\pi(3,2,1,4) = (4,3,1,2);$$
$$\pi(2,4,3,1) = (1,2,3,4);$$
$$\pi(a,b,c,d) = (d,a,c,b).$$

△

Permutations are frequently used to anyalyze problems that involves lists. Indeed they are used so frequently that even the tabular notation is considered burdensome; we need a simpler notation.

DEFINITION 5.5. A **cycle** is a vector

$$\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}$$

that corresponds to the permutation where the entry in position α_1 is moved to position α_2; the entry in position α_2 is moved to position α_3, ... and the element in position α_n is moved to position α_1. If a position is not listed in α, then the entry in that position is not moved. We call such positions **stationary**. For the identity cycle where no entry is moved, we write

$$\iota = (1).$$

The fact that the permutation α moves the entry in position α_n to position α_1 is the reason that this is called a *cycle;* applying it repeatedly cycles the list of elements around, and on the nth application we have returned to the original list.

EXAMPLE 5.6. We can write π from Example 5.4 as a cycle. In tabular notation,

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

To write it as a cycle, we can start with any position we like. However, the convention is to start with the smallest position that changes. Since π moves elements out of position 1, we start with

$$\pi = \begin{pmatrix} 1 & ? \end{pmatrix}.$$

The second entry in cycle notation tells us where π moves the element whose position is that of the first entry. The first entry indicates position 1. From the tabular notation, we see that π moves the element in position 1 to position 2, so

$$\pi = \begin{pmatrix} 1 & 2 & ? \end{pmatrix}.$$

The third entry of cycle notation tells us where π moves the element whose position is that of the second entry. The second entry indicates position 2. From the tabular notation, we see that π moves the element in position 2 to position 4, so

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

The fourth entry of cycle notation tells us where π moves the element whose position is that of the third entry. The third element indicates position 4. From the tabular notation, we see that π moves the element in position 4 to position 1, so you might feel the temptation to write

$$\pi = \begin{pmatrix} 1 & 2 & 4 & 1 & ? \end{pmatrix},$$

but there is no need. At this point we take advantage of the cycle notation to close the cycle:

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

Here the first cycle in π, $\begin{pmatrix} 1 & 2 & 4 \end{pmatrix}$, indicates that

- the element in position 1 of a list moves to the position 2;
- the element in position 2 of a list moves to position 4;
- the element in position 4 of a list moves to position 1.

What about the element in position 3? According to the piecewise and tabular notations for π, it doesn't move anywhere. This is reflected by the fact that 3 does not appear in the cycle notation for π. △

Not all permutations can be written as one cycle.

EXAMPLE 5.7. Consider the permutation in tabular notation

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

We can easily start the cycle with $\alpha = \begin{pmatrix} 1 & 2 \end{pmatrix}$, and this captures the behavior on the elements in the first and second positions of a list, but what about the third and fourth? This presents a temporary difficulty. △

To solve this problem, we need to develop some simple arithmetic of cycles. Cycles represent permutations; permutations are one-to-one functions; functions can be *composed*.

EXAMPLE 5.8. Consider the two cycles

$$\beta = \begin{pmatrix} 2 & 3 & 4 \end{pmatrix} \quad \text{and} \quad \gamma = \begin{pmatrix} 1 & 2 & 4 \end{pmatrix}.$$

What is the cycle notation for

$$\beta \circ \gamma = \begin{pmatrix} 2 & 3 & 4 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 4 \end{pmatrix}?$$

We can answer this by considering an example list; let $V = (1,2,3,4)$ and compute $(\beta \circ \gamma)(V)$. Since $(\beta \circ \gamma)(x) = \beta(\gamma(x))$, first we apply γ:

$$\gamma(V) = (4,1,3,2),$$

followed by β:

$$\beta(\gamma(V)) = (4,2,1,3).$$

Thus

- the element in position 1 eventually moved to position 3;
- the element in position 3 eventually moved to position 4;
- the element in position 4 eventually moved to position 1;
- the element in position 2 did not move.

In cycle notation, we write this as

$$\beta \circ \gamma = \begin{pmatrix} 1 & 3 & 4 \end{pmatrix}.$$

△

Another phenomenon occurs when each permutation moves elements that the other does not.

EXAMPLE 5.9. Consider the two cycles

$$\beta = \begin{pmatrix} 1 & 3 \end{pmatrix} \quad \text{and} \quad \gamma = \begin{pmatrix} 2 & 4 \end{pmatrix}.$$

There is no way to simplify $\beta \circ \gamma$ into a *single* cycle, because β operates only on the first and third elements of a list, and γ operates only on the second and fourth elements of a list. The only way to write them is as the composition of two cycles,

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

△

This motivates the following.

DEFINITION 5.10. We say that two cycles are **disjoint** if none of their entries are common.

Disjoint cycles enjoy an important property.

LEMMA 5.11. *Let α, β be two disjoint cycles. Then $\alpha \circ \beta = \beta \circ \alpha$.*

PROOF. Let $n \in \mathbb{N}^+$ be the largest entry in α or β. Let $V = (1, 2, \ldots, n)$. Let $i \in V$. We consider the following cases:

Case 1. $\alpha(i) \neq i$.

Let $j = \alpha(i)$. The definition of cycle notation implies that j appears immediately after i in the cycle α. Recall that α and β are disjoint. Since i and j are entries of α, they cannot be entries of β. By definition of cycle notation, $\beta(i) = i$ and $\beta(j) = j$. Hence

$$(\alpha \circ \beta)(i) = \alpha(\beta(i)) = \alpha(i) = j = \beta(j) = \beta(\alpha(i)) = (\beta \circ \alpha)(i).$$

Case 2. $\alpha(i) = i$.

Subcase (a): $\beta(i) = i$.

We have $(\alpha \circ \beta)(i) = i = (\beta \circ \alpha)(i)$.

Subcase (b): $\beta(i) \neq i$.

Let $j = \beta(i)$. We have

$$(\beta \circ \alpha)(i) = \beta(\alpha(i)) = \beta(i) = j.$$

The definition of cycle notation implies that j appears immediately after i in the cycle β. Recall that α and β are disjoint. Since j is an entry of β, it cannot be an entry of α. By definition of cycle notation, $\alpha(j) = j$. Hence

$$(\alpha \circ \beta)(i) = \alpha(j) = j = (\beta \circ \alpha)(i).$$

In both cases, we had $(\alpha \circ \beta)(i) = (\beta \circ \alpha)(i)$. Since i was arbitrary, $\alpha \circ \beta = \beta \circ \alpha$. □

NOTATION. Since the composition of two disjoint cycles $\alpha \circ \beta$ cannot be simplified, we normally write them consecutively, without the circle that indicates composition, for example

$$(\ 1\ \ 2\ )(\ 3\ \ 4\ ).$$

By Lemma 5.11, we can also write this as

$$(\ 3\ \ 4\ )(\ 1\ \ 2\ ).$$

That said, the usual convention for cycles is to write the smallest entry first, and to write cycles with smaller first entries before cycles with larger first entries. Thus we prefer

$$(\ 1\ \ 4\ )(\ 2\ \ 3\ )$$

to either of

$$(\ 1\ \ 4\ )(\ 3\ \ 2\ ) \quad \text{or} \quad (\ 2\ \ 3\ )(\ 1\ \ 4\ ).$$

Beyond that, the convention for writing a permutation in cycle form is the following:

(1) Rotate each cycle so that the first entry is the smallest entry in each cycle.
(2) Simplify the permutation by computing the composition of cycles that are not disjoint. Discard all cycles of length 1.
(3) The remaining cycles will be disjoint. From Lemma 5.11, we know that they commute; write them in order from smallest first entry to largest first entry.

EXAMPLE 5.12. We return to Example 5.7, with

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

To write this permutation in cycle notation, we begin again with

$$\alpha = \begin{pmatrix} 1 & 2 \end{pmatrix} \ldots ?$$

Since α also moves entries in positions 3 and 4, we need to add a second cycle. We start with the smallest position whose entry changes position, 3:

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

Since α moves the element in position 3 to position 4, we write

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

Now α moves the element in position 4 to position 3, so we can close the second cycle:

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

Now α moves no more entries, so the cycle notation is complete. △

We now come to the main goal of this section.

THEOREM 5.13. *Every permutation can be written as a composition of cycles.*

The proof is constructive.

PROOF. Let π be a permutation; denote its domain by V. Without loss of generality, we write $V = (1, 2, \ldots, n)$.

Let i_1 be the smallest element of V such that $\pi(i_1) \neq i_1$. Recall that the range of π has at most n elements; since π is one-to-one, eventually $\pi^k(i_1) = i_1$ for some $k \leq n$. Let $\alpha^{(1)}$ be the cycle $\begin{pmatrix} i_1 & \pi(i_1) & \pi(\pi(i_1)) & \cdots & \pi^k(i_1) \end{pmatrix}$ where $\pi^{k+1}(i_1) = i_1$.

At this point, either every element of V that is not stationary with respect to π appears in $\alpha^{(1)}$, or it does not. If there is some $i_2 \in V$ such that i_2 is not stationary with respect to π and $i_2 \notin \alpha^{(1)}$, then generate the cycle $\alpha^{(2)}$ by $\begin{pmatrix} i_2 & \pi(i_2) & \pi(\pi(i_2)) & \cdots & \pi^\ell(i_2) \end{pmatrix}$ where as before $\pi^\ell(i_2) = i_2$.

Repeat this process until every non-stationary element of V corresponds to a cycle, generating $\alpha^{(3)}, \ldots, \alpha^{(m)}$ for non-stationary $i_3 \notin \alpha^{(1)}, \alpha^{(2)}$, $i_4 \notin \alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \ldots, i_m \notin \alpha^{(1)}, \ldots, \alpha^{(m-1)}$.

The remainder of the proof consists of two claims.

Claim 1: $\alpha^{(i)}$ and $\alpha^{(j)}$ are disjoint for any $i \neq j$.

Suppose to the contrary that there exists an integer r such that $r \in \alpha^{(i)}$ and $r \in \alpha^{(j)}$. By definition, the next entry of both $\alpha^{(i)}$ and $\alpha^{(j)}$ is $\pi(r)$. The subsequent entry of both is $\pi(\pi(r))$, and so forth. This cycles through both $\alpha^{(i)}$ and $\alpha^{(j)}$ until we reach $\pi^\lambda(r) = r$ for some $\lambda \in \mathbb{N}$. Hence $\alpha^{(i)} = \alpha^{(j)}$.

Claim 2: $\pi = \alpha^{(1)}\alpha^{(2)} \cdots \alpha^{(m)}$.

Let $i \in V$. If $\pi(i) = i$, then by definition $\alpha^{(j)}(i) = i$ for all $j = 1, 2, \ldots, m$. Otherwise, i appears in $\alpha^{(j)}$ for some $j = 1, 2, \ldots, m$. By definition, $\alpha^{(j)}(i) = \pi(i)$. By claim 1, both i and

$\pi(i)$ appear in *only* one of the α. Hence

$$\begin{aligned}\left(\alpha^{(1)}\alpha^{(2)}\cdots\alpha^{(m)}\right)(i) &= \alpha^{(1)}\left(\alpha^{(2)}\left(\cdots\alpha^{(m-1)}\left(\alpha^{(m)}(i)\right)\right)\right)\\ &= \alpha^{(1)}\left(\alpha^{(2)}\left(\cdots\alpha^{(j-1)}\left(\alpha^{(j)}(i)\right)\right)\right)\\ &= \alpha^{(1)}\left(\alpha^{(2)}\left(\cdots\alpha^{(j-1)}(\pi(i))\right)\right)\\ &= \pi(i).\end{aligned}$$

□

EXAMPLE 5.14. Consider the permutation

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

Using the proof of Theorem 5.13, we define the cycles

$$\begin{aligned}\alpha^{(1)} &= \begin{pmatrix} 1 & 7 \end{pmatrix}\\ \alpha^{(2)} &= \begin{pmatrix} 2 & 5 & 4 \end{pmatrix}\\ \alpha^{(3)} &= \begin{pmatrix} 6 & 8 \end{pmatrix}.\end{aligned}$$

Notice that the α are all disjoint. In addition, the only element of $V = (1,2,\ldots,8)$ that does not appear in an α is 3, because $\pi(3) = 3$. Inspection verifies that

$$\pi = \alpha^{(1)}\alpha^{(2)}\alpha^{(3)}.$$

△

We conclude with some examples where two permutations are composed.

EXAMPLE 5.15. Let $\alpha = \begin{pmatrix} 1 & 3 \end{pmatrix}\begin{pmatrix} 2 & 4 \end{pmatrix}$ and $\beta = \begin{pmatrix} 1 & 3 & 2 & 4 \end{pmatrix}$. Notice that $\alpha \neq \beta$; check this on $V = (1,2,3,4)$ if this isn't clear. In addition, α and β are not disjoint.

(1) We compute the cycle notation for $\gamma = \alpha \circ \beta$. We start with the smallest entry moved by either α or β:

$$\gamma = \begin{pmatrix} 1 & ? \end{pmatrix}.$$

The notation $\alpha \circ \beta$ means to apply β first, *then* α. What does β do with the entry in position 1? It moves it to position 3. Subsequently, α moves the entry in position 3 back to the entry in position 1. The next entry in the first cycle of γ should thus be 1, but that's also the first entry in the cycle, so we close the cycle. So far, we have

$$\gamma = \begin{pmatrix} 1 \end{pmatrix}\ldots?$$

We aren't finished, since α and β also move other entries around. The next smallest entry moved by either α or β is 2, so

$$\gamma = \begin{pmatrix} 1 \end{pmatrix}\begin{pmatrix} 2 & ? \end{pmatrix}.$$

Now β moves the entry in position 2 to the entry in position 4, and α moves the entry in position 4 to the entry in position 2. The next entry in the second cycle of γ should thus be 2, but that's also the first entry in the second cycle, so we close the cycle. So far, we have

$$\gamma = \begin{pmatrix} 1 \end{pmatrix}\begin{pmatrix} 2 \end{pmatrix}\ldots?$$

Next, β moves the entry in position 3, so

$$\gamma = (\ 1\ )(\ 2\ )(\ 3\ \ ?\ ).$$

Where does β move the entry in position 3? To the entry in position 2. Subsequently, α moves the entry in position 2 to the entry in position 4. We now have

$$\gamma = (\ 1\ )(\ 2\ )(\ 3\ \ 4\ \ ?\ ).$$

You can probably guess that 4, as the largest possible entry, will close the cycle, but to be safe we'll check: β moves the entry in position 4 to the entry in position 1, and α moves the entry in position 1 to the entry in position 3. The next entry of the third cycle will be 3, but this is also the first entry of the third cycle, so we close the third cycle and

$$\gamma = (\ 1\ )(\ 2\ )(\ 3\ \ 4\ ).$$

Finally, we simplify γ by not writing cycles of length 1, so

$$\gamma = (\ 3\ \ 4\ ).$$

Hence

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

(2) Now we compute the cycle notation for $\beta\circ\alpha$, but with less detail. Again we start with 1, which α moves to 3, and β then moves to 2. So we start with

$$\beta\circ\alpha = (\ 1\ \ 2\ \ ?\ ).$$

Next, α moves 2 to 4, and β moves 4 to 1. This closes the first cycle:

$$\beta\circ\alpha = (\ 1\ \ 2\ )\ldots?$$

We start the next cycle with position 3: α moves it to position 1, which β moves back to position 3. This generates a length-one cycle, so there is no need to add anything. Likewise, the element in position 4 is also stable under $\beta\circ\alpha$. Hence we need write no more cycles;

$$\beta\circ\alpha = (\ 1\ \ 2\ ).$$

(3) Let's look also at $\beta\circ\gamma$ where $\gamma = (\ 1\ \ 4\ )$. We start with 1, which γ moves to 4, and then β moves to 1. Since $\beta\circ\gamma$ moves 1 to itself, we don't have to write 1 in the cycle. The next smallest number that appears is 2: γ doesn't move it, and β moves 2 to 4. We start with

$$\beta\circ\gamma = (\ 2\ \ 4\ \ ?\ ).$$

Next, γ moves 4 to 1, and β moves 1 to 3. This adds another element to the cycle:

$$\beta\circ\gamma = (\ 2\ \ 4\ \ 3\ \ ?\ ).$$

We already know that 1 won't appear in the cycle, so you might guess that we should not close the cycle. To be certain, we consider what $\beta\circ\gamma$ does to 3: γ doesn't move it, and β moves 3 to 2. The cycle is now complete:

$$\beta\circ\gamma = (\ 2\ \ 4\ \ 3\ ).$$

EXERCISES.

EXERCISE 5.16. For the permutation

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

(a) Evaluate $\alpha(1,2,3,4,5,6)$.
(b) Evaluate $\alpha(1,5,2,4,6,3)$.
(c) Evaluate $\alpha(6,3,5,2,1,4)$.
(d) Write α in cycle notation.
(e) Write α as a piecewise function.

EXERCISE 5.17. For the permutation

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

(a) Evaluate $\alpha(1,2,3,4)$.
(b) Evaluate $\alpha(1,4,3,1)$.
(c) Evaluate $\alpha(3,1,4,2)$.
(d) Write α in tabular notation.
(e) Write α as a piecewise function.

EXERCISE 5.18. Let $\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 4 & 3 & 2 \end{pmatrix}$, and $\gamma = \begin{pmatrix} 1 & 3 \end{pmatrix}$. Compute $\alpha\circ\beta$, $\alpha\circ\gamma$, $\beta\circ\gamma$, $\beta\circ\alpha$, $\gamma\circ\alpha$, $\gamma\circ\beta$, α^2, β^2, and γ^2. (Here $\alpha^2 = \alpha\circ\alpha$.) What is the inverse of each permutation?

EXERCISE 5.19. Construct the cyclic group generated by the permutation

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

Hint: Life will probably be easier if you convert it to cycle notation first.

5.2. GROUPS OF PERMUTATIONS

In Section 5.1 we introduced permutations. For $n \geq 2$, denote by S_n the set of all permutations of a list of n elements. In this section we show that S_n is a group for all $n \geq 2$.

EXAMPLE 5.20. For $n = 2,3$ we have

$$S_2 = \{(1), \begin{pmatrix} 1 & 2 \end{pmatrix}\}$$
$$S_3 = \{(1), \begin{pmatrix} 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 3 \end{pmatrix}, \begin{pmatrix} 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 3 & 2 \end{pmatrix}\}.$$

△

A counting argument based on the multiplication principle shows that S_n has

$$|S_n| = n! = n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1$$

elements: given any list of n elements,

- we have n positions to move the first element, including its current position;
- we have $n-1$ positions to move the second element, since the first element has already taken one spot;
- we have $n-2$ positions to move the third element, since the first and second elements have already take two spots;
- etc.

We observed in Section 5.1 that any permutation is really a one-to-one function; naturally, one can ask whether the set of all permutations on n elements behaves as a group under the operation of composition of functions.

THEOREM 5.21. *For all $n \geq 2$ $(S_n, \circ)$ is a group.*

NOTATION. Normally we just write S_n, understanding from context that the operation is composition of functions. It is common to refer to S_n as the **symmetric group** of n elements.

PROOF. Let $n \geq 2$. We have to show that S_n satisfies (G1)–(G4) under the operation of composition of functions:

(G1): For closure, we must show that the composition of two permutations is a permutation. This is the statement of Lemma 5.3 on page 71.

(G2): The associative property follows from the fact that permutations are functions, and functions are associative.

(G3): The identity function ι such that $\iota(x) = x$ for all $x \in \{1, 2, \ldots, n\}$ is also the identity of S_n under composition: for any $\alpha \in S_n$ and for any $x \in \{1, 2, \ldots, n\}$ we have

$$(\iota \circ \alpha)(x) = \iota(\alpha(x)) = \alpha(x);$$

since x was arbitrary, $\iota \circ \alpha = \alpha$.

(G4): Every one-to-one function has an inverse function, so every element of S_n has an inverse element under composition.

□

In Section 5.38 on page 84 we will show that *all* finite groups are isomorphic to a subgroup of a symmetric group.

EXERCISES.

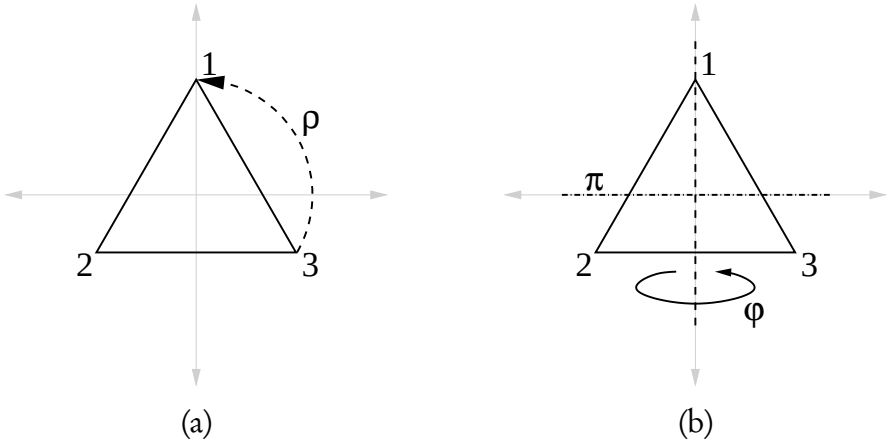
EXERCISE 5.22. Show that all the elements of S_3 can be written as compositions of $\alpha = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$ and $\beta = \begin{pmatrix} 2 & 3 \end{pmatrix}$.

EXERCISE 5.23. For α and β as defined in Exercise 5.22, show that $\beta \circ \alpha = \alpha^2 \circ \beta$. (Notice that $\alpha, \beta \in S_n$ for all $n > 2$, so as a consequence of this exercise S_n is not abelian for $n > 2$.)

EXERCISE 5.24. Write the composition table for S_3. *Hint:* List the six elements of S_3 as $\begin{pmatrix} 1 \end{pmatrix}$, α, α^2, β, $\alpha\beta$, $\alpha^2\beta$, using the previous exercises both to justify and to simplify this task.

EXERCISE 5.25. Show that $D_3 \cong S_3$ by showing that the function $f : D_3 \to S_3$ by $f\left(\rho^a \varphi^b\right) = \alpha^a \beta^b$ is an isomorphism.

FIGURE 5.1. Rotation and reflection of an equilateral triangle centered at the origin



EXERCISE 5.26. How many elements are there of S_4? List them all using cycle notation.

EXERCISE 5.27. Compute the cyclic subgroup of S_4 generated by $\alpha = \begin{pmatrix} 1 & 3 & 4 & 2 \end{pmatrix}$. Compare your answer to that of Exercise 5.19.

EXERCISE 5.28. Let $\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix} \in S_n$. Show that we can write α^{-1} as

$$\beta = \begin{pmatrix} \alpha_1 & \alpha_n & \alpha_{n-1} & \cdots & \alpha_2 \end{pmatrix}.$$

For example, if $\alpha = \begin{pmatrix} 2 & 5 & 3 & 6 \end{pmatrix}$, $\alpha^{-1} = \begin{pmatrix} 2 & 6 & 3 & 5 \end{pmatrix}$. *Hint:* Try computing $\alpha \circ \beta$ and $\beta \circ \alpha$.

5.3. DIHEDRAL GROUPS

In Section 2.5 we studied the symmetries of a triangle; we represented the group as the products of matrices ρ and φ, derived from the symmetries of *rotation* and *reflection about the y-axis*. Figure 5.1, a copy of Figure 2.3 on page 33, shows how ρ and φ correspond to the symmetries of an equilateral triangle centered at the origin. In Exercises 5.22–5.25 you showed that D_3 and S_3 are isomorphic.

We can develop matrices to reflect the symmetries of a regular n-sided polygon as well (the regular n-gon), motivating the definition of the set D_n of symmetries of the n-gon.

DEFINITION 5.29. The **dihedral set** D_n is the set of symmetries of a regular polygon with n sides.

Is D_n always a group?

THEOREM 5.30. *Let $n \in \mathbb{N}$ and $n \geq 3$. Then $(D_n, \circ)$ is a group with $2n$ elements, called the **dihedral group**.*

The proof of Theorem 5.30 depends on the following proposition, which we accept without proof. We could prove it using an argument from matrices as in Chapter 2.5, but proving it requires more energy than is appropriate for this section.

PROPOSITION 5.31. *All the symmetries of a regular n-sided polygon can be generated by a composition of a power of the rotation ρ of angle $2\pi/n$ and a power of the flip φ across the y-axis. In addition, $\varphi^2 = \rho^n = \iota$ (the identity symmetry) and $\varphi\rho = \rho^{n-1}\varphi$.*

PROOF. We must show that properties (G1)–(G4) are satisfied.

(G1): Closure follows from Proposition 5.31.

(G2): The associative property follows from the fact that permutations are functions, and the associative property applies to functions.

(G3): Certainly there exists an identity element $\iota \in D_n$, which corresponds to the identity symmetry where no vertex is moved.

(G4): It is obvious that the inverse of a symmetry of the regular n-gon is also a symmetry of the regular n-gon.

It remains to show that D_n has $2n$ elements. From the properties of ρ and φ in Proposition 5.31, all other symmetries are combinations of these two, which means that all symmetries are of the form $\rho^a\varphi^b$ for some $a \in \{0,\dots,n-1\}$ and $b \in \{0,1\}$. Since $\varphi^2 = \rho^n = \iota$, a can have n values and b can have 2 values. Hence there are $2n$ possible elements altogether. □

We have two goals in introducing the dihedral group: first, to give you another concrete and interesting group; and second, to serve as a bridge to Section 5.4. The next example starts to lead in the latter direction.

EXAMPLE 5.32. Another way to represent the elements of D_3 is to consider how they re-arrange the vertices of the triangle. We can represent the vertices of a triangle as the list $V = (1,2,3)$. Application of ρ to the triangle moves

- vertex 1 to vertex 2;
- vertex 2 to vertex 3; and
- vertex 3 to vertex 1.

This is equivalent to the permutation $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$.

Application of φ to the triangle moves

- vertex 1 to itself—that is, vertex 1 does not move;
- vertex 2 to vertex 3; and
- vertex 3 to vertex 2.

This is equivalent to the permutation $\begin{pmatrix} 2 & 3 \end{pmatrix}$.

In the context of the symmetries of the triangle, it looks as if we can say that $\rho = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$ and $\varphi = \begin{pmatrix} 2 & 3 \end{pmatrix}$. Recall that ρ and φ generate all the symmetries of a triangle; likewise, these two cycles generate all the permutations of a list of three elements! (See Example 5.20 on page 78 and Exercise 2.63 on page 35.) △

Of course, we can do this with D_4 and S_4 as well.

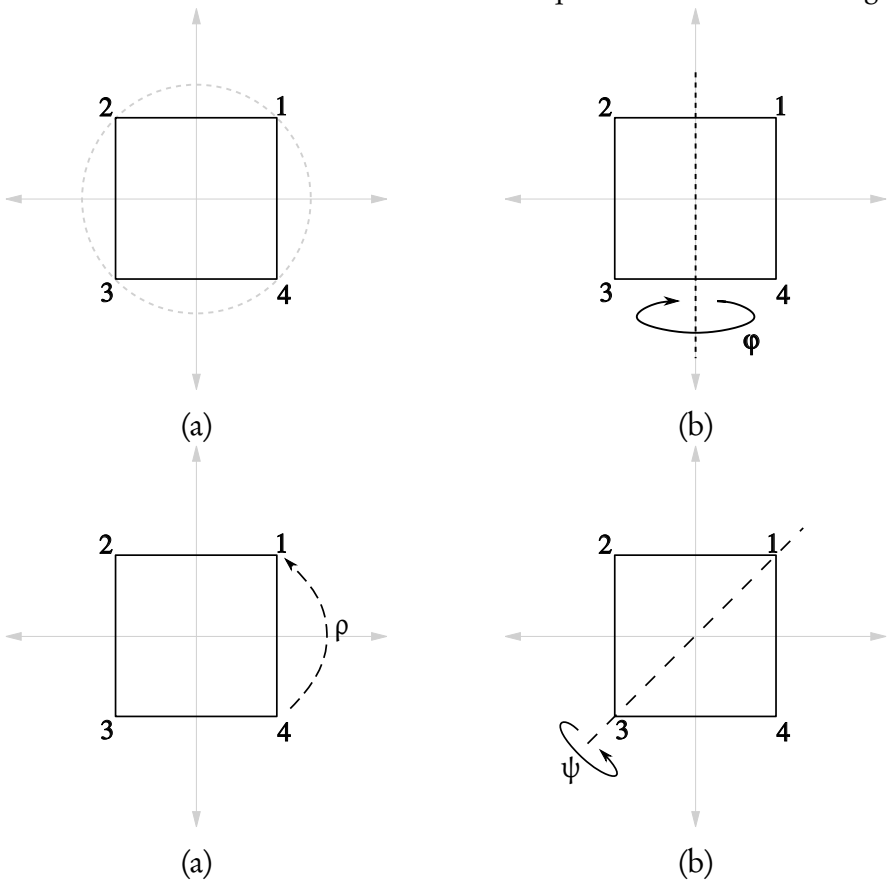
EXAMPLE 5.33. Using the tabular notation for permutations, we identify some elements of D_4, the set of symmetries of a square. Of course we have an identity permutation

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

and a 90° rotation

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

FIGURE 5.2. Rotation and reflection of a square centered at the origin



We can imagine three kinds of flips: one across the y-axis,

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

one across the x-axis,

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

and one across a diagonal,

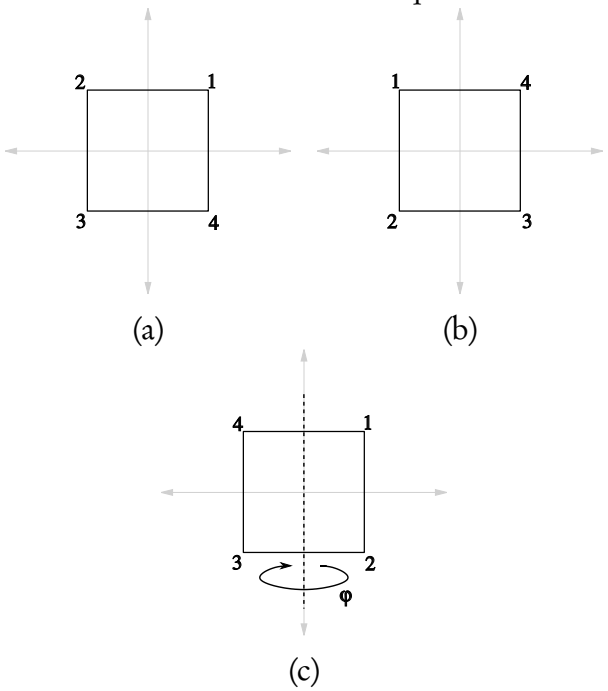
$$\psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}.$$

See Figure 5.2. We can also imagine other diagonals; but they can be shown to be superfluous, just as we show shortly that ϑ and ψ are superflulous. There may be other symmetries of the square, but we'll stop here for the time being.

Is it possible to write ψ as a composition of φ and ρ? It turns out that $\psi = \varphi \circ \rho$. To show this, we consider them as permutations of the vertices of the square, as we did with the triangle above, rather than repeat the agony of computing the matrices of isometries as in Section 2.5.

- Geometrically, ρ moves $(1,2,3,4)$ to $(4,1,2,3)$; subsequently φ moves $(4,1,2,3)$ to $(1,4,3,2)$; see Figure 5.3.

FIGURE 5.3. Rotation and reflection of a square centered at the origin



- We can use the tabular notation for ψ, φ, and ρ to show that the composition of the functions is the same. Starting with the list $(1,2,3,4)$ we see form the tabular notation above that

$$\psi(1,2,3,4) = (1,4,3,2).$$

On the other hand,

$$\rho(1,2,3,4) = (4,1,2,3).$$

Things get a little tricky here; we want to evaluate $\varphi \circ \rho$, and

$$\begin{aligned}(\varphi \circ \rho)(1,2,3,4) &= \varphi(\rho(1,2,3,4)) \\ &= \varphi(4,1,2,3) \\ &= (1,4,3,2).\end{aligned}$$

How did we get that last step? Look back at the tabular notation for φ: the element in the first entry is moved to the second. In the next-to-last line above, the element in the first entry is 4; it gets moved to the second entry in the last line:

$$\begin{array}{cccc} \mathbf{4}, & 1, & 2, & 3 \\ & \searrow & & \\ ?, & \mathbf{4}, & ?, & ? \end{array}$$

The tabular notation for φ also tells us to move the element in the second entry (1) to the first. Thus

$$\begin{array}{cccc} 4, & \mathbf{1}, & 2, & 3 \\ & \swarrow\!\!\searrow & & \\ \mathbf{1}, & 4, & ?, & ? \end{array}$$

Likewise, φ moves the element in the third entry (2) to the fourth, and vice-versa, giving us

$$\begin{array}{cccc} 4, & 1, & \mathbf{2}, & \mathbf{3} \\ & \searrow\!\!\!\times & & \searrow\!\!\!\times \\ 1, & 4, & \mathbf{3}, & \mathbf{2} \end{array}$$

In both cases, we see that $\psi = \varphi \circ \rho$. A similar argument shows that $\vartheta = \varphi \circ \rho^2$, so it looks as if we need only φ and ρ to generate D_4. The reflection and the rotation have a property similar to that in S_3:

$$\varphi \circ \rho = \rho^3 \circ \varphi,$$

so unless there is some symmetry of the square that cannot be described by rotation or reflection on the y-axis, we can list all the elements of D_4 using a composition of some power of ρ after some power of φ. There are four unique 90° rotations and two unique reflections on the y-axis, implying that D_4 has at least eight elements:

$$D_4 \supseteq \left\{\iota, \rho, \rho^2, \rho^3, \varphi, \rho\varphi, \rho^2\varphi, \rho^3\varphi\right\}.$$

Can D_4 have other elements? There are in fact $|S_4| = 4! = 24$ possible permutations of the vertices, but are they all symmetries of a square? Consider the permutation from $(1,2,3,4)$ to $(2,1,3,4)$: in the basic square, the distance between vertices 1 and 3 is $\sqrt{2}$, but in the configuration $(2,1,3,4)$ vertices 1 and 3 are adjacent on the square, so the distance between them has diminished to 1. Meanwhile, vertices 2 and 3 are no longer adjacent, so the distance between them has increased from 1 to $\sqrt{2}$. Since the distances between points on the square was not preserved, the permutation described, which we can write in tabular notation as

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

is *not* an element of D_4. The same can be shown for the other fifteen permutations of four elements.

Hence D_4 has eight elements, making it smaller than S_4, which has $4! = 24$. _____△

COROLLARY 5.34. *For any $n \geq 2$ we have $D_n < S_n$, with equality only when $n = 2$.*

EXERCISES.

EXERCISE 5.35. Write all eight elements of D_4 in cycle notation.

EXERCISE 5.36. Construct the composition table of D_4.

EXERCISE 5.37. Show that the symmetries of any n-gon can be described as a power of ρ and φ, where φ is a flip about the y-axis and ρ is a rotation of $2\pi/n$ radians.

5.4. CAYLEY'S REMARKABLE RESULT

The mathematician Arthur Cayley discovered a lovely fact about the permutation groups.

THEOREM 5.38 (Cayley's Theorem). *Every finite group of n elements is isomorphic to a subgroup of S_n.*

We're going to give an example *before* we give the proof. Hopefully the example will help explain how the proof of the theorem works.

EXAMPLE 5.39. Consider the Klein 4-group; this group has four elements, so it must be isomorphic to a subgroup of S_4. We will build the isomorphism by looking at the multiplication table for the Klein 4-group:

$\times$	e	a	b	ab
e	e	a	b	ab
a	a	e	ab	b
b	b	ab	e	a
ab	ab	b	a	e

To find a permutation appropriate to each element, we'll do the following. First, we label each element with a certain number:

$$e \leftrightsquigarrow 1,$$
$$a \leftrightsquigarrow 2,$$
$$b \leftrightsquigarrow 3,$$
$$ab \leftrightsquigarrow 4.$$

We will use this along with tabular notation to determine the isomorphism. Define a map f from the Klein 4-group to S_4 by

$$f(x) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ \ell(x \cdot e) & \ell(x \cdot a) & \ell(x \cdot b) & \ell(x \cdot ab) \end{pmatrix}, \tag{5.4.1}$$

where $\ell(y)$ is the label that corresponds to y.

First let's compute $f(a)$:

$$f(a) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ ? & ? & ? & ? \end{pmatrix}.$$

The first entry has the value $\ell(a \cdot e) = \ell(a) = 2$, telling us that

$$f(a) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & ? & ? & ? \end{pmatrix}.$$

The next entry has the value $\ell(a \cdot a) = \ell(a^2) = \ell(e) = 1$, telling us that

$$f(a) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & ? & ? \end{pmatrix}.$$

The third entry has the value $\ell(a \cdot b) = \ell(ab) = 4$, telling us that

$$f(a) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & ? \end{pmatrix}.$$

The final entry has the value $\ell(a \cdot ab) = \ell(a^2 b) = \ell(b) = 3$, telling us that

$$f(a) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}.$$

So applying the formula in equation (5.4.1) definitely gives us a permutation.

In fact, we could have filled out the bottom row of the permutation by looking above at the multiplication table for the Klein 4-group, locating the row for the multiples of a (the third row

of the multiplication table), and filling in the labels for the entries in that row! Doing this or applying equation (5.4.1) to the other elements of the Klein 4-group tells us that

$$f(e) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$
$$f(b) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$$
$$f(ab) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}.$$

We now have a subset of S_4; written in cycle notation, it is

$$\begin{aligned} W &= \{f(e), f(a), f(b), f(ab)\} \\ &= \{(1), (1\ \ 2)(3\ \ 4), (1\ \ 3)(2\ \ 4), (1\ \ 4)(2\ \ 3)\}. \end{aligned}$$

Verifying that W is a group, and therefore a subgroup of S_4, is straightforward; you will do so in the homework. What we need to ensure is that f is indeed an isomorphism. Inspection shows that f is one-to-one and onto; the hard part is the homomorphism property. We will use a little cleverness for this. Let x, y in the Klein 4-group.

- Recall that $f(x)$, $f(y)$, and $f(xy)$ are permutations, and by definition one-to-one, onto functions on a list of four elements.
- Notice that ℓ is also a one-to-one function, and it has an inverse.
- Let $m \in (1,2,3,4)$. For any z in the Klein 4-group, $\ell(z) = m$ if we listed z as the mth entry of the group. Thus $\ell^{-1}(m)$ indicates the element of the Klein four-group that is labeled by m. For instance, $\ell^{-1}(b) = 3$.
- Since $f(x)$ is a permutation of a list of four elements, we can look at $(f(x))(m)$ as the place where $f(x)$ moves m.
- By definition, $f(x)$ moves m to $\ell(z)$ where $z = x \cdot \ell^{-1}(m)$. Similar statement hold for how $f(y)$ and $f(xy)$ move m.
- Applying these facts, we observe that

$$\begin{aligned} (f(x) \circ f(y))(m) &= (f(x))(f(y)(m)) \\ &= f(x)\left(\ell\left(y \cdot \ell^{-1}(m)\right)\right) \\ &= \ell\left(x \cdot \ell^{-1}\left(\ell\left(y \cdot \ell^{-1}(m)\right)\right)\right) \\ &= \ell\left(x \cdot \left(y \cdot \ell^{-1}(m)\right)\right) \\ &= \ell\left(xy \cdot \ell^{-1}(m)\right) \\ &= f(xy)(m). \end{aligned}$$

- Since m was arbitrary in $\{1,2,3,4\}$, $f(xy)$ and $f(x) \circ f(y)$ are identical functions.
- Since x, y were arbitrary in the Klein 4-gorup, $f(xy) = f(x) f(y)$.

We conclude that f is a homomorphism; since it is one-to-one and onto, f is an isomorphism. $\triangle$

You should read through Example 5.39 carefully two or three times, and make sure you understand it, since in the homework you will construct a similar isomorphism for a different group, and also because we do the same thing now in the proof of Cayley's Theorem.

PROOF OF CAYLEY'S THEOREM. Let G be a finite group of n elements. Label the elements in any order $G = \{g_1, g_2, \ldots, g_n\}$ and for any $x \in G$ denote $\ell(x) = i$ such that $x = g_i$. Define a relation

$$f : G \to S_n \qquad \text{by} \qquad f(g_i) = \begin{pmatrix} 1 & 2 & \cdots & n \\ \ell(g_i \cdot g_1) & \ell(g_i \cdot g_2) & \cdots & \ell(g_i \cdot g_n) \end{pmatrix}.$$

As we explained in Example 5.39 for the Klein 4-group, this assigns to each $g \in G$ the permutation that, in tabular notation, has the labels for each entry in the row corresponding to g of the operation table for G. By this fact we know that f is one-to-one and onto (see also Theorem 2.29 on page 24). The proof that f is a homomorphism is identical to the proof for Example 5.39: nothing in that argument required x, y, or z to be elements of the Klein 4-group; the proof was for a general group! Hence f is an isomorphism, and $G \cong f(G) < S_n$. □

What's so remarkable about this result? One way of looking at it is the following: since every finite group is isomorphic to a subgroup of a group of permutations, *everything you need to know about finite groups can be learned from studying the groups of permutations!* A more flippant summary is that *the theory of finite groups is all about studying how to rearrange lists.*

In theory, I could go back and rewrite these notes, introducing the reader first to lists, then to permutations, then to S_2, to S_3, to the subgroups of S_4 that correspond to the cyclic group of order 4 and the Klein 4-group, and so forth, making no reference to these other groups, nor to the diherdral group, nor to any other finite group that we have studied.

In reality, it is inconvenient to work only with the symmetric group; it is usually more natural to think in terms other than permutations (geometry for D_n is helpful); and it can be tedious to work only with permutations. While Cayley's Theorem has its uses, it does not suggest that we should always consider groups of permutations in place of the more natural representations.

EXERCISES.

EXERCISE 5.40. In Example 5.39 we found W, a subgroup of S_4 that is isomorphic to the Klein 4-group. It turns out that $W < D_4$ as well. Draw the geometric representations for each element of W, using a square and writing labels in the appropriate places, as we did in Figures 2.3 on page 33 and 5.2.

EXERCISE 5.41. Apply Cayley's Theorem to find a subgroup of S_4 that is isomorphic to $\mathbb{Z}_4$. Write the permutations in both tabular and cycle notations.

EXERCISE 5.42. The subgroup of S_4 that you identified in Exercise 5.41 is also a subgroup of D_4. Draw the geometric representations for each element of this subgroup, using a square and writing labels in the appropriate places.

EXERCISE 5.43. Since S_3 has six elements, we know it is isomorphic to a subgroup of S_6. Can you identify this subgroup *without* using the isomorphism used in the proof of Cayley's Theorem?

5.5. ALTERNATING GROUPS

We close this chapter with a special kind of symmetry group that has very important implications for later topics: the *alternating groups*. To define the alternating group, we need to study permutations a little more closely, in particular the cycle notation.

DEFINITION 5.44. Let $n \in \mathbb{N}^+$. An n-cycle is a permutation that can be written as one cycle with n entries. A **transposition** is a 2-cycle.

EXAMPLE 5.45. The permutation $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \in S_3$ is a 3-cycle. The permutation $\begin{pmatrix} 2 & 3 \end{pmatrix} \in S_3$ is a transposition. The permutation $\begin{pmatrix} 1 & 3 \end{pmatrix}\begin{pmatrix} 2 & 4 \end{pmatrix} \in S_4$ cannot be written as only one n-cycle for any $n \in \mathbb{N}^+$: it is the composition of two disjoint transpositions, and any cycle must move 1 to 3, so it would start as $\begin{pmatrix} 1 & 3 & ? \end{pmatrix}$. If we fill in the blank with anything besides 1, we have a different permutation. So we must close the cycle before noting that 2 moves to 4. △

Thanks to 1-cycles, any permutation can be written with many different numbers of cycles: for example,

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

In addition, a neat trick allows us to write every permutation as a composition of transitions.

EXAMPLE 5.46. $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix}$. Also

$$\begin{pmatrix} 1 & 4 & 8 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 8 \end{pmatrix}\begin{pmatrix} 1 & 4 \end{pmatrix}.$$

Also $(1) = \begin{pmatrix} 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix}$. △

LEMMA 5.47. *Any permutation can be written as a composition of transitions.*

PROOF. You do it! See Exercise (5.57). □

At this point it is worth looking at Example 5.46 and the discussion before it. Can we write ρ with many different numbers of *transpositions*? Yes:

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

Notice something special about the representation of $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$. No matter how you write it, it always has an *even* number of transitions. By contrast, consider

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

No matter how you write it, you always represent $\begin{pmatrix} 2 & 3 \end{pmatrix}$ with an *odd* number of transpositions.

Is this always the case?

THEOREM 5.48. *Let α be a cycle.*

- *If α can be written as the composition of an even number of transpositions, then it cannot be written as the composition of an odd number of transpositions.*

- *If α can be written as the composition of an odd number of transpositions, then it cannot be written as the composition of an even number of transpositions.*

PROOF. Suppose that $\alpha \in S_n$. Consider the polynomials

$$g = \prod_{1 \le i < j \le n} \left(x_i - x_j\right) \qquad \text{and} \qquad g_\alpha := \prod_{1 \le i < j \le n} \left(x_{\alpha(i)} - x_{\alpha(j)}\right).$$

Is $g_\alpha = g$? Not always: If for example $\alpha = \begin{pmatrix} 1 & 2 \end{pmatrix}$ then $g = x_1 - x_2$ and $g_\alpha = x_2 - x_1 \neq g$. Likewise if $\alpha = \begin{pmatrix} 1 & 3 & 2 \end{pmatrix}$ then

$$g = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$$

whereas

$$g_\alpha = (x_3 - x_1)(x_3 - x_2)(x_1 - x_2). \tag{5.5.1}$$

Failing this, can we write g_α *in terms of* g? Try the following. If $\alpha(i) < \alpha(j)$, then the binomial $x_{\alpha(i)} - x_{\alpha(j)}$ appears in g, so we'll leave it alone. (An example would be $(x_1 - x_2)$ in equation (5.5.1).) Otherwise $\alpha(i) > \alpha(j)$ and the binomial $x_{\alpha(i)} - x_{\alpha(j)}$ does not appear in g. (An example would be $(x_3 - x_1)$ in equation (5.5.1).) However, the binomial $x_{\alpha(j)} - x_{\alpha(i)}$ does appear in g, so rewrite g_α by replacing $x_{\alpha(i)} - x_{\alpha(j)}$ as $\left[-\left(x_{\alpha(j)} - x_{\alpha(i)}\right)\right]$.

Recall that α is a one-to-one function: for each i, x_i is mapped to one unique $x_{\alpha(i)}$. In addition, each binomial $x_i - x_j$ in g is unique, so for each i, j, the binomial $x_i - x_j$ is mapped to a binomial $x_{\alpha(i)} - x_{\alpha(j)}$ where the subscripts are unique; that is, in g_α there is are no k, ℓ such that the binomial $x_{\alpha(k)} - x_{\alpha(\ell)}$ has the same pair of subscripts as $x_{\alpha(i)} - x_{\alpha(j)}$. Thus, factoring the constant -1 multiples from the product gives us

$$g_\alpha = (-1)^{\operatorname{swp}\alpha}\, g,$$

where $\operatorname{swp}\alpha \in \mathbb{N}$ is an integer representing the number of swapped indices that α provoked in the binomials of g.

Consider two different representations of α with transpositions. If the first representation has an even number of transpositions, then an even number of binomials in g swapped indices to get g_α, so $\operatorname{swp}\alpha$ is even. Hence $g_\alpha = g$. If the second representation had an odd number of transpositions, there would be an odd number of swaps from g to g_α, , and $\operatorname{swp}\alpha$ would be odd and $g_\alpha = -g$. However, the value of g_α depends on the *permutation* α, not on the choice of *representation* of α. So we cannot have $g_\alpha = g$ and $g_\alpha = -g$. It follows that both representations must have the same number of transpositions.

The statement of the theorem follows: no matter how we choose the representation of α, it will always have either an even or an odd number of transpositions. □

Theorem 5.48 motivates the following. Any permutation can be written as a composition of cycles, and any cycle can be written as either an even or odd number of transpositions. Thus, any permutation can be written as either an even or an odd number of transpositions.

DEFINITION 5.49. If a permutation can be written with an even number of permutations, then we say that the permutation is **even**. Otherwise, we say that the permutation is **odd**.

EXAMPLE 5.50. The permutation $\rho = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \in S_3$ is even, since as we saw earlier $\rho = \begin{pmatrix} 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix}$. So is the permutation $\iota = (1) = \begin{pmatrix} 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix}$. The permutation $\varphi = \begin{pmatrix} 2 & 3 \end{pmatrix}$ is odd. △

At this point we are ready to define a new group.

DEFINITION 5.51. Let $n \in \mathbb{N}^+$ and $n \geq 2$. Let $A_n = \{\alpha \in S_n : \alpha \text{ is even}\}$. We call A_n **the set of alternating permutations**.

EXAMPLE 5.52. We briefly mentioned the alternating group A_3 in Example 3.46 on page 47.

THEOREM 5.53. *For all $n \geq 2$, A_n is a group under the operation of composition of functions.*

PROOF. Let $n \geq 2$. We show that A_n satisfies properties (G1)–(G4) of a group.

(G1): For closure, let $\alpha, \beta \in A_n$. Both α and β can be written as the composition of an even number of transpositions. The sum of two even numbers is also even, so $\alpha \circ \beta$ is also the composition of an even number of transpositions.

(G2): The associative property is inherited from S_n, or more generally from the associative property of the composition of functions.

(G3): The identity element is $\iota = (1)$, which Example 5.46 shows is even.

(G4): Let $\alpha \in A_n$. Write α as a composition of transpositions, denoted by

$$\alpha = \tau_1 \tau_2 \cdots \tau_m$$

for some $m \in \mathbb{N}^+$. Since $\alpha \in A_n$, m is even. Let

$$\beta = \tau_m \tau_{m-1} \cdots \tau_1.$$

You will show in Exercise 5.58 that any transposition is its own inverse, so

$$\begin{aligned}
\alpha\beta &= (\tau_1 \tau_2 \cdots \tau_m)(\tau_m \tau_{m-1} \cdots \tau_1) \\
&= (\tau_1 \tau_2 \cdots \tau_{m-1})(\tau_m \tau_m)(\tau_{m-1} \tau_{m-1} \cdots \tau_1) \\
&= (\tau_1 \tau_2 \cdots \tau_{m-1})\,\iota\,(\tau_{m-1} \tau_{m-2} \cdots \tau_1) \\
&= (\tau_1 \tau_2 \cdots \tau_{m-2})(\tau_{m-1} \tau_{m-1})(\tau_{m-2} \tau_{m-3} \cdots \tau_1) \\
&= (\tau_1 \tau_2 \cdots \tau_{m-2})\,\iota\,(\tau_{m-2} \tau_{m-3} \cdots \tau_1) \\
&\vdots \\
&= \tau_1 \tau_1 \\
&= \iota.
\end{aligned}$$

Hence $\alpha\beta = (\ 1\ )$. A similar argument shows that $\beta\alpha = (\ 1\ )$, so $\beta = \alpha^{-1}$. We have written β with m transpositions. Recall that m is even, so $\alpha^{-1} = \beta \in A_n$.

□

How large is A_n, relative to S_n? Before we answer this, we need to an interesting and useful fact about the function swp α that we defined in Theorem 5.48.

THEOREM 5.54. *For any $n \geq 2$, there are half as many even permutations as there are permutations. That is, $|A_n| = |S_n|/2$.*

PROOF. We use Lagrange's Theorem from page 45, and show that there are two cosets of $A_n < S_n$.

Let $X \in S_n/A_n$. Let $\alpha \in S_n$ such that $X = \alpha A_n$. If α is an even permutation, then $X = A_n$. Otherwise, α is odd. Let β be any other odd permutation. Write out the odd number of transpositions of α^{-1}, followed by the odd number of transpositions of β, to see that $\alpha^{-1}\beta$ is an even permutation. Hence $\alpha^{-1}\beta \in A_n$, and by Lemma 3.24 on page 43 $\alpha A_n = \beta A_n$.

Thus there are only two cosets of A_n in S_n: A_n itself, and the coset of odd permutations. By Lagrange's Theorem,

$$\frac{|S_n|}{|A_n|} = |S_n / A_n| = 2,$$

and a little algebra rewrites this equation to $|A_n| = |S_n| / 2$. □

COROLLARY 5.55. *For any $n \geq 2$, $A_n \triangleleft S_n$.*

PROOF. You do it! See Exercise 5.60. □

Unfortunately, we have only shown a few rather dull facts regarding A_n. —I say that, but these facts are quite striking really: A_n is half the size of S_n, and it is a normal subgroup of S_n. If I call these facts "rather dull", that tells you that some very interesting things are in store for the future.

EXERCISES.

EXERCISE 5.56. List the elements of A_2, A_3, and A_4 in cycle notation.

EXERCISE 5.57. Show that any permutation can be written as a product of transpositions. *Hint:* You know that any permutation can be written as a product of cycles, so it will suffice to show that any cycle can be written as a product of transpositions. For that, take an arbitrary cycle $\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}$ and write it as a product of transpositions, as suggested by Example 5.46. Be sure to explain why this product does in fact equal α.

EXERCISE 5.58. Show that the inverse of any transposition is a transposition. *Hint:* You can do this by showing that any transposition is its own inverse. Take an arbitrary transposition $\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix}$ and show that $\alpha^2 = \iota$.

EXERCISE 5.59. Show that the function $\operatorname{swp} \alpha$ defined in Theorem 5.48 satisfies the property that for any two cycles α, β we have $(-1)^{\operatorname{swp}(\alpha\beta)} = (-1)^{\operatorname{swp}\alpha}(-1)^{\operatorname{swp}\beta}$. *Hint:* Let α and β be arbitrary cycles. Consider the four possible cases where α and β are even or odd.

EXERCISE 5.60. Show that for any $n \geq 2$, $A_n \triangleleft S_n$. *Hint:* See a previous exercise about subgroups or cosets.

5.6. THE 15-PUZZLE

The 15-puzzle is similar to a toy you probably played with as a child. It looks like a 4×4 square, with all the squares numbered except one. The numbering starts in the upper left and proceeds consecutively until the lower right; the only squares that aren't in order are the last two numbers squares, which are swapped:

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

The challenge is to find a way to rearrange the squares so that they are in order, like so:

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

The only permissible moves are those where one "slides" a square left, right, above, or below the empty square. Given the starting position above, the following moves are permissible:

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

or

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

but the following moves are *not* permissible:

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

or

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

.

We will use groups of permutations to show that that the challenge is impossible.

How can we do this? Since the problem is one of rearranging a list of elements, it is a problem of permutations. Every permissible move consists of transpositions τ in S_{16} where:

- $\tau = \begin{pmatrix} x & y \end{pmatrix}$ where
  - $x < y$;
  - $[x] = [0]$ in $\mathbb{Z}_4$ implies $[y] = [0]$ in $\mathbb{Z}_4$;
  - one of x or y is the position of the empty square in the current list; and
  - legal moves imply that either
    - $y = x + 1$; or
    - $y = x + 4$.

EXAMPLE 5.61. The legal moves illustrated above correspond to the transpositions

- $\begin{pmatrix} 15 & 16 \end{pmatrix}$, because square 14 was in position 15, and the empty space was in position 16: notice that $16 = 15 + 1$; and
- $\begin{pmatrix} 12 & 16 \end{pmatrix}$, because square 12 was in position 12, and the empty space was in position 16: notice that $16 = 12 + 4$ and since $[12] = [0]$ in $\mathbb{Z}_4$, $[16] = [0]$ in $\mathbb{Z}_4$.

The illegal moves illustrated above correspond to the transpositions

- $\begin{pmatrix} 11 & 16 \end{pmatrix}$, because square 11 was in position 11, and the empty space was in position 16: notice that $16 = 11 + 5$; and
- $\begin{pmatrix} 13 & 14 \end{pmatrix}$, because in the original configuration, neither 13 nor 14 contains the empty square.

Likewise $\begin{pmatrix} 12 & 13 \end{pmatrix}$ would be an illegal move in any configuration, because it crosses rows: notice that even though $y = x + 1$, $[12] = [0]$ in $\mathbb{Z}_4$ but $[13] \neq [0]$ in $\mathbb{Z}_4$. _____△

How can we use this to show that it is impossible to solve 15-puzzle? Answering this requires several steps. The first shows that if there is a solution, it must belong to a particular group.

LEMMA 5.62. *If there is a solution to the 15-puzzle, it is a permutation $\sigma \in A_{16}$, where A_{16} is the alternating group.*

PROOF. Any permissible move corresponds to a transposition τ as described above. Now any solution contains the empty square in the lower right hand corner. As a consequence, we must have the following: For any move $\begin{pmatrix} x & y \end{pmatrix}$, there must eventually be a corresponding move $\begin{pmatrix} x' & y' \end{pmatrix}$ where $[x'] = [x]$ in $\mathbb{Z}_4$ and $[y'] = [y]$ in $\mathbb{Z}_4$. If not:

- for above-below moves, the empty square could never return to the bottom row; and
- for left-right moves, the empty square could never return to the rightmost row unless we had some $\begin{pmatrix} x & y \end{pmatrix}$ where $[x] = [0]$ and $[y] \neq [0]$, a contradiction.

Thus moves come in pairs, and the solution is a permutation σ consisting of an even number of transpositions. By Theorem 5.48 on page 88 and Definitions 5.49 and 5.51, $\sigma \in A_{16}$. □

We can now show that there is no solution to the 15-puzzle.

THEOREM 5.63. *The 15-puzzle has no solution.*

PROOF. By way of contradiction, assume that it has a solution σ. Then $\sigma \in A_{16}$. Because A_{16} is a subgroup of S_{16}, and hence a group in its own right, $\sigma^{-1} \in A_{16}$. Notice $\sigma^{-1}\sigma = \iota$, the permutation which corresponds to the configuration of the solution.

Now σ^{-1} is a permutation corresponding to the moves that change the arrangement

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

into the arrangement

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

which corresponds to $\begin{pmatrix} 14 & 15 \end{pmatrix}$. So regardless of the transpositions used in the representation of σ^{-1}, the composition must simplify to $\sigma^{-1} = \begin{pmatrix} 14 & 15 \end{pmatrix} \notin A_{16}$, a contradiction. □

As a historical note, the 15-puzzle was developed in 1878 by an American puzzlemaker, who promised a $1,000 reward to the first person to solve it. Most probably, the puzzlemaker knew that no one would ever solve it: if we account for inflation, the reward would correspond to $22,265 in 2008 dollars.[1]

EXERCISES.

EXERCISE 5.64. Determine which of these configurations, if any, is solvable by the same rules as the 15-puzzle:

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

,

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

,

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

.

[1] According to the website `http://www.measuringworth.com/ppowerus/result.php`.

Hint: Use the same strategy as that of the proof of Theorem 5.63: find the permutation σ^{-1} that corresponds to the current confiuration, and decide whether $\sigma^{-1} \in A_{16}$. If not, you know the answer is no. If so, you must still check that it can be written as a product of transpositions that satisfy the rules of the puzzle.

CHAPTER 6

Some elementary number theory

6.1. The Greatest Common Divisor and the Euclidean Algorithm

In grade school, you learned how to compute the greatest common divisor of two integers. For example, given the integers 36 and 210, you should be able to determine that the greatest common divisor is 6.

Computing greatest common divisors—not only of integers, but of other objects as well—turns out to be one of the most interesting problems in mathematics, with a large number of important applications. Besides, many of the concepts underlying greatest common divisors turn out to be deeply interesting topics on their own. Because of this, we review them as well, starting with a definition which you probably don't expect.

Definition 6.1. Let $n \in \mathbb{Z}$ and assume $n > 1$. We say that n is **irreducible** if the only integers that divide n are ± 1 and $\pm n$.

You may read this and think, "Oh, he's talking about prime numbers." Well, yes and no. More on that in the next section.

Example 6.2. The integer 36 is not irreducible, because $36 = 6 \times 6$. The integer 7 is irreducible, because the only integers that divide 7 are ± 1 and ± 7. △

Definition 6.3. Let $m, n \in \mathbb{Z}$. We say that $d \in \mathbb{Z}$ is a **common divisor of m and n if $d \mid m$ and $d \mid n$**. The **greatest common divisor of m and n**, written $\gcd(m,n)$, is the largest of the common divisors of m and n.

Example 6.4. Common divisors of 36 and 210 are 1, 2, 3, and 6. △

One way to compute the list of common divisors is to list all possible divisors of both integers, and identify the largest possible positive divisor. In practice, this takes a Very Long Time^{TM}, so it would be nice to have a different method. One such method was described by the Greek mathematician Euclid many centuries ago.

Theorem 6.5 (The Euclidean Algorithm). *Let $m, n \in \mathbb{Z}$. One can compute the greatest common divisor of m, n in the following way:*

(1) *Let $s = \max(m,n)$ and $t = \min(m,n)$.*
(2) *Repeat the following steps until $t = 0$:*
   (a) *Let q be the quotient and r the remainder after dividing s by t.*
   (b) *Assign s the current value of t.*
   (c) *Assign t the current value of r.*

The final value of s is $\gcd(m,n)$.

Before proving that the Euclidean algorithm gives us a correct answer, let's do an example.

EXAMPLE 6.6. We compute $\gcd(36,210)$ using the Euclidean algorithm. Start by setting $s = 210$ and $t = 36$. Subsequently:

(1) Dividing 210 by 36 gives $q = 5$ and $r = 30$. Set $s = 36$ and $t = 30$.
(2) Dividing 36 by 30 gives $q = 1$ and $r = 6$. Set $s = 30$ and $t = 6$.
(3) Dividing 30 by 6 gives $q = 5$ and $r = 0$. Set $s = 6$ and $t = 0$.

Now that $t = 0$, we stop, and conclude that $\gcd(36,210) = s = 6$. △

When we prove that the Euclidean algorithm generates a correct answer, we will argue that it computes $\gcd(m,n)$ by claiming

$$\gcd(m,n) = \gcd(m,r_1) = \gcd(r_1,r_2) = \cdots = \gcd(r_{k-1},0)$$

where r_i is the remainder from division of the previous two integers in the chain and r_{k-1} is the final non-zero remainder from division. Lemma 6.7 proves this claim.

LEMMA 6.7. *Let $s,t \in \mathbb{Z}$. Let q and r be the quotient and remainder, respectively, of division of s by t, as per the Division Theorem from page 13. Then* $\gcd(s,t) = \gcd(s,r)$.

PROOF. Let $d = \gcd(s,t)$. First we show that d is a divisor of r. From Definition 13, there exist $a,b \in \mathbb{Z}$ such that $s = ad$ and $t = bd$. From the Division Theorem, we know that $s = qt + r$. Substitution gives us $ad = q(bd) + r$; rewriting the equation, we have

$$r = (a - qb)\,d.$$

Hence $d \mid r$.

Since d is a common divisor of s, t, and r, it is a common divisor of t and r. Now we show that $d = \gcd(t,r)$. Let $d' = \gcd(t,r)$; since d is also a common divisor of t and r, the definition of *greatest* common divisior implies that $d \leq d'$. Since d' is a common divisor of t and r, Definition 13 again implies that there exist $x,y \in \mathbb{Z}$ such that $t = d'x$ and $r = d'y$. Substituting into the equation $s = qt + r$, we have $s = q(d'x) + d'y$; rewriting the equation, we have

$$s = (qx + y)\,d'.$$

So $d' \mid s$. We already knew that $d' \mid t$, so d' is a common divisor of s and t.

Recall that $d = \gcd(s,t)$; since d' is also a common divisor of t and r, the definition of *greatest* common divisor implies that $d' \leq d$. Earlier, we showed that $d \leq d'$. Hence $d \leq d' \leq d$, which implies that $d = d'$.

Substitution gives the desired conclusion: $\gcd(s,t) = \gcd(t,r)$. □

We can now finally prove that the Euclidean algorithm gives us a correct answer. This requires two stages, necessary for proving that any algorithm terminates correctly.

(1) **Termination.** To prove that any algorithm provides a correct answer, you must prove that it gives *some* answer. How can this be a problem? If you look at the Euclidean algorithm, you see that one of its instructions asks us to "repeat" some steps "until $t = 0$." What if t never attains the value of zero? It's conceivable that its values remain positive at all times, or jump over zero from positive to negative values. That would mean that we never receive any answer from the algorithm, let alone a correct one.
(2) **Correctness.** Even if the algorithm terminates, we have to guarantee that it terminates with the correct answer.

We will identify both stages of the proof clearly. In addition, we will refer back to the the Division Theorem as well as the well-ordering property of the integers from Section 1.3; you may wish to review those.

PROOF OF THE EUCLIDEAN ALGORITHM. First we show that the algorithm terminates. The only repetition in the algorithm occurs in step 2. The first time we compute step 2(a), we compute the quotient q and remainder r of division of s by t. By the Division Theorem,

$$0 \leq r < t. \tag{6.1.1}$$

Denote this value of r by r_1. In step 2(b) we set s to t, and in step 2(c) we set the value of t to $r_1 = r$. Thanks to equation (6.1.1), the value of $t_{\text{new}} = r$ is smaller than $s_{\text{new}} = t_{\text{old}}$. If $t \neq 0$, then we return to 2(a) and divide s by t, again obtaining a new remainder r. Denote this value of r by r_2; by the Division Theorem $r_2 = r < t$, so

$$0 \leq r_2 < r_1.$$

As long as we repeat step 2, we generate a set of integers $R = \{r_1, r_2, \ldots\} \subset \mathbb{N}_{\geq 0}$. The well-ordering property of the integers implies that R has a smallest element r_i; this implies in turn that after i repetitions, step 2 of the algorithm must stop repeating; otherwise, we would generate $r_{i+1} < r_i$, contradicting the fact that r_i is the smallest element of R. Since step 2 of the algorithm terminates, the algorithm itself terminates.

Now we show that the algorithm terminates *with the correct answer*. If step 2 of the algorithm repeated k times, then applying Lemma 6.7 repeatedly to the remainders of the divisions gives us the chain of equalities

$$\begin{aligned}
\gcd(r_{k-1}, r_{k-2}) &= \gcd(r_{k-2}, r_{k-3}) \\
&= \gcd(r_{k-3}, r_{k-4}) \\
&\vdots \\
&= \gcd(r_2, r_1) \\
&= \gcd(r_1, s) \\
&= \gcd(t, s) \\
&= \gcd(m, n).
\end{aligned}$$

What is $\gcd(r_{k-1}, r_{k-2})$? The final division of s by t is the division of r_{k-1} by r_{k-2}; since the algorithm terminates after the kth repetition, $r_k = 0$. By Definition 1.6, $r_{k-1} \mid r_{k-2}$, making r_{k-1} a common divisor of r_{k-1} and r_{k-2}. No integer larger than r_{k-1} divides r_{k-1}, so the greatest common divisor of r_{k-1} and r_{k-2} is r_{k-1}. Following the chain of equalities, we conclude that $\gcd(m, n) = r_{k-1}$: the Euclidean Algorithm terminates with the correct answer. □

EXERCISES.

EXERCISE 6.8. Compute the greatest common divisor of 100 and 140 by (a) listing all divisors, then identifying the largest; and (b) the Euclidean Algorithm.

EXERCISE 6.9. Compute the greatest common divisor of 4343 and 4429 by using the Euclidean Algorithm.

EXERCISE 6.10. In Lemma 6.7 we showed that $\gcd(m,n) = \gcd(m,r)$ where r is the remainder after division of m by n. Prove the following more general statement: for all $m,n,q \in \mathbb{Z}$ $\gcd(m,n) = \gcd(m-qn)$.

6.2. THE CHINESE REMAINDER THEOREM; THE CARD TRICK EXPLAINED

Go back and reread the card trick from Section 1.1. In this section we explain how the card trick works. The result is based on an old, old Chinese observation.

THEOREM 6.11 (The Chinese Remainder Theorem, simple version). *Let $m,n \in \mathbb{Z}$ such that $\gcd(m,n) = 1$. Let $\alpha,\beta \in \mathbb{Z}$. There exists a solution $x \in \mathbb{Z}$ to the system of linear congruences*

$$\begin{cases} [x] &= [\alpha] \text{ in } \mathbb{Z}_m; \\ [x] &= [\beta] \text{ in } \mathbb{Z}_n; \end{cases}$$

and x is unique in $\mathbb{Z}_N$ where $N = mn$.

Before giving the proof, let's look at an example.

EXAMPLE 6.12 (The card trick). In the card trick, we took twelve cards and arranged them

- once in groups of three; and
- once in groups of four.

Each time, the player identified the *column* in which the mystery card lay. This gave the remainders α from division by three and β from division by four, leading to the system of linear congruences

$$\begin{cases} [x] &= [\alpha] \text{ in } \mathbb{Z}_3; \\ [x] &= [\beta] \text{ in } \mathbb{Z}_m; \end{cases}$$

where x is the location of the mystery card. The simple version of the Chinese Remainder Theorem guarantees us that there is a solution for x, and that this solution is unique in $\mathbb{Z}_{12}$. Since there are only twelve cards, the solution is unique in the game: as long as the dealer can compute x, s/he can identify the card infallibly. △

The reader may be thinking, "Well, and good, but knowing only the existence of a solution seems rather pointless. I also need to know *how* to compute x, so that I can pinpoint the location of the card. How does the Chinese Remainder Theorem help with that?" This emerges from the proof. However, the proof requires us to revisit our friend, the Euclidean Algorithm.

THEOREM 6.13 (The Extended Euclidean Algorithm). *Let $m,n \in \mathbb{Z}$. There exist $a,b \in \mathbb{Z}$ such that $am + bn = \gcd(m,n)$. Both a and b can be found by reverse-substituting the chain of equations obtained by the repeated division in the Euclidean algorithm.*

EXAMPLE 6.14. Recall Example 6.6 the computation of $\gcd(210,36)$. The divisions gave us a series of equations:

$$210 = 5 \cdot 36 + 30 \tag{6.2.1}$$

$$36 = 1 \cdot 30 + 6 \tag{6.2.2}$$

$$30 = 5 \cdot 6 + 0.$$

We concluded from the Euclidean Algorithm that $\gcd(210,36) = 6$. We start by rewriting the equation 6.2.1:

$$36 - 1 \cdot 30 = 6. \tag{6.2.3}$$

This looks a little like what we want, but we need 210 instead of 30. Equation 6.2.2 allows us to rewrite 30 in terms of 210 and 36:

$$30 = 210 - 5 \cdot 36. \tag{6.2.4}$$

Substituting this result into equation 6.2.3, we have

$$36 - (210 - 5 \cdot 36) = 6 \quad \Longrightarrow \quad 6 \cdot 36 + (-1) \cdot 210 = 6.$$

We have found integers $m = 6$ and $n = -1$ such that for $a = 36$ and $b = 210$, $\gcd(a,b) = 6$. △

PROOF OF THE EXTENDED EUCLIDEAN ALGORITHM. Recall that the Euclidean algorithm computes a chain of k quotients $\{q_i\}$ and remainders $\{r_i\}$ such that

$$\begin{aligned} m &= q_1 n + r_1 \\ n &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{k-3} &= q_{k-1} r_{k-2} + r_{k-1} \qquad (6.2.5) \\ r_{k-2} &= q_k r_{k-1} + r_k \qquad (6.2.6) \\ r_{k-1} &= q_{k+1} r_k + 0 \\ \text{and } r_k &= \gcd(m,n). \end{aligned}$$

Using the last equation, we can rewrite equation 6.2.6 as

$$r_{k-2} = q_k r_{k-1} + \gcd(m,n).$$

Solving for $\gcd(m,n)$, we have

$$r_{k-2} - q_k r_{k-1} = \gcd(m,n). \tag{6.2.7}$$

Now solve equation 6.2.5 for r_{k-1} to obtain

$$r_{k-3} - q_{k-1} r_{k-2} = r_{k-1}.$$

Substitute this into equation 6.2.7 to obtain

$$\begin{aligned} r_{k-2} - q_k \left(r_{k-3} - q_{k-1} r_{k-2}\right) &= \gcd(m,n) \\ \left(q_{k-1} + 1\right) r_{k-2} - q_k r_{k-3} &= \gcd(m,n). \end{aligned}$$

Proceeding in this fashion, we can exhaust the list of equations, concluding by solving the first equation in the chain for m, and substituting for r_1 to obtain $am + bn = \gcd(m,n)$ for some integers a, b. □

This ability to write $\gcd(m,n)$ as a sum of integer multiples of m and n is the key to unlocking the Chinese Remainder Theorem. Before doing so, we need an important lemma about numbers whose gcd is 1.

LEMMA 6.15. *Let $d, m, n \in \mathbb{Z}$. If $m \mid nd$ and $\gcd(m,n) = 1$, then $m \mid d$.*

PROOF. Assume that $m \mid nd$ and $\gcd(m,n) = 1$. By definition, there exists $q \in \mathbb{Z}$ such that $qm = nd$. Use the Extended Euclidean Algorithm to choose $a, b \in \mathbb{Z}$ such that $am + bn = \gcd(m,n) = 1$. Then

$$\begin{aligned} amd + bnd &= d \\ adm + bqm &= d \\ (ad + bq)\, m &= d. \end{aligned}$$

Hence $m \mid d$. □

We finally prove the Chinese Remainder Theorem.

PROOF OF THE CHINESE REMAINDER THEOREM, SIMPLE VERSION. Recall that the system is

$$\begin{cases} [x] &= [\alpha] \text{ in } \mathbb{Z}_m; \text{ and} \\ [x] &= [\beta] \text{ in } \mathbb{Z}_n. \end{cases}$$

We have to prove two things: first, that a solution x exists; second, that $[x]$ is unique in $\mathbb{Z}_{12}$.

Existence: Because $\gcd(m,n) = 1$, the Extended Euclidean Algorithm tells us that there exist $a, b \in \mathbb{Z}$ such that $am + bn = 1$. Rewriting this equation two different ways, we have $bn = 1 + (-a)\, m \in 1 + \langle m \rangle$ and $am = 1 + (-b)\, n \in 1 + \langle n \rangle$. Hence $[bn] = [1]$ in $\mathbb{Z}_m$ and $[am] = [1]$ in $\mathbb{Z}_n$. Hence $[\alpha bn] = [\alpha]$ in $\mathbb{Z}_m$ and $[\beta am] = [\beta]$ in $\mathbb{Z}_n$. Moreover, $[\alpha bn] = [0]$ in $\mathbb{Z}_n$ and $[\beta am] = [0]$ in $\mathbb{Z}_m$. Hence

$$\begin{cases} [\alpha bn + \beta am] &= [\alpha] \text{ in } \mathbb{Z}_m; \text{ and} \\ [\alpha bn + \beta am] &= [\beta] \text{ in } \mathbb{Z}_n. \end{cases}$$

Thus $x = \alpha bn + \beta am$ satisfies the requirements of the system.

Uniqueness: Suppose that there exist $[x], [y] \in \mathbb{Z}_N$ that both satisfy the system. Since $[x] = [y]$ in $\mathbb{Z}_m$, $[x - y] = [0]$, so $m \mid (x - y)$. By definition of divisibility, there exists $q \in \mathbb{Z}$ such that $mq = (x - y)$. Since $[x] = [y]$ in $\mathbb{Z}_n$, $[x - y] = [0]$, so $n \mid (x - y)$. By substitution, $n \mid mq$. By Lemma 6.15, $n \mid q$. By definition of divisibility, there exists $q' \in \mathbb{Z}$ such that $q = nq'$. By substitution,

$$x - y = mq = mnq' = Nq'.$$

Hence $N \mid (x - y)$, so that $[x - y] = [0]$ in $\mathbb{Z}_N$, so that $[x] = [y]$ in $\mathbb{Z}_N$, as desired. □

The existence part of the proof gives us an algorithm to solve problems involving the Chinese Remainder Theorem:

COROLLARY 6.16 (Chinese Remainder Theorem Algorithm, simple version). *Let $m, n \in \mathbb{Z}$ such that $\gcd(m,n) = 1$. Let $\alpha, \beta \in \mathbb{Z}$. Write $N = mn$. We can solve the system of linear congruences*

$$\begin{cases} [x] &= [\alpha] \text{ in } \mathbb{Z}_m; \\ [x] &= [\beta] \text{ in } \mathbb{Z}_n; \end{cases}$$

for $[x] \in \mathbb{Z}_N$ by the following steps:

1. *Use the Extended Euclidean Algorithm to find $a, b \in \mathbb{Z}$ such that $am + bn = 1$.*
2. *The solution is $[\alpha bn + \beta am]$ in $\mathbb{Z}_N$.*

PROOF. The proof follows immediately from the existence proof of Theorem 6.11. □

EXAMPLE 6.17. The algorithm of Corollary 6.16 finally explains the method of the card trick. We have $m = 3$, $n = 4$, and $N = 12$. Suppose that the player indicates that his card is in the first column when they are grouped by threes, and in the third column when they are grouped by fours; then $\alpha = 1$ and $\beta = 3$.

The Extended Euclidean Algorithm tells us that $a = 1$ and $b = -1$ give us us $am + bn = 1$; hence $am = 4$ and $bn = -3$. We can therefore find the mystery card by computing

$$x = 1 \cdot 4 + 3 \cdot (-3) = -5;$$

but this isn't helpful. By adding 12, we obtain another representation for $[x]$ in $\mathbb{Z}_{12}$:

$$[x] = [-5 + 12] = [7],$$

which implies that the player chose the 7th card. In fact, $[7] = [1]$ in $\mathbb{Z}_3$, and $[7] = [3]$ in $\mathbb{Z}_4$, which agrees with the information given. △

The Chinese Remainder Theorem can be generalized to larger systems with more than two equations under certain circumstances.

THEOREM 6.18 (Chinese Remainder Theorem on $\mathbb{Z}$). *Let $m_1, m_2, \ldots, m_n \in \mathbb{Z}$ such that $\gcd\left(m_i, m_j\right) = 1$ for all $1 \leq i < j \leq n$. Let $\alpha_1, \alpha_2, \ldots \alpha_n \in \mathbb{Z}$. There exists a solution $x \in \mathbb{Z}$ to the system of linear congruences*

$$\begin{cases} [x] &= [\alpha_1] \text{ in } \mathbb{Z}_{m_1}; \\ [x] &= [\alpha_2] \text{ in } \mathbb{Z}_{m_2}; \\ &\vdots \\ [x] &= [\alpha_n] \text{ in } \mathbb{Z}_{m_n}; \end{cases}$$

and x is unique in $\mathbb{Z}_N$ where $N = m_1 m_2 \cdots m_n$.

Before we can prove this version of the Chinese Remainder Theorem, we need to make an observation of $m_1, m_2, \ldots, m_n$.

LEMMA 6.19. *Let $m_1, m_2, \ldots, m_n \in \mathbb{Z}$ such that $\gcd\left(m_i, m_j\right) = 1$ for all $1 \leq i < j \leq n$. For each $i = 1, 2, \ldots, n$ define $N_i = N / m_i$ where $N = m_1 m_2 \cdots m_n$; that is, N_i is the product of all the m's except m_i. Then $\gcd(m_i, N_i) = 1$.*

PROOF. We show that $\gcd(m_1, N_1) = 1$; for $i = 2, \ldots, n$ the proof is similar.

Use the Extended Euclidean Algorithm to choose $a, b \in \mathbb{Z}$ such that $am_1 + bm_2 = 1$. Use it again to choose $c, d \in \mathbb{Z}$ such that $cm_1 + dm_3 = 1$. Then

$$\begin{aligned} 1 &= (am_1 + bm_2)(cm_1 + dm_3) \\ &= (acm_1 + adm_3 + bcm_2)m_1 + (bd)(m_2 m_3). \end{aligned}$$

Let $x = \gcd(m_1, m_2 m_3)$; the previous equation shows that x is also a divisor of 1. However, the only divisors of 1 are ± 1; hence $x = 1$. We have shown that $\gcd(m_1, m_2 m_3) = 1$.

Rewrite the equation above as $1 = a' m_1 + b' m_2 m_3$; notice that $a', b' \in \mathbb{Z}$. Use the Extended Euclidean Algorithm to choose $e, f \in \mathbb{Z}$ such that $em_1 + fm_4 = 1$. Then

$$\begin{aligned} 1 &= \left(a' m_1 + b' m_2 m_3\right)(em_1 + fm_4) \\ &= \left(a' em_1 + a' fm_4 + b' em_2 m_e\right) m_1 + \left(b' f\right)(m_2 m_3 m_4). \end{aligned}$$

An argument similar to the one above shows that $\gcd(m_1, m_2 m_3 m_4) = 1$.

Repeating this process with each m_i, we obtain $\gcd(m_1, m_2 m_3 \cdots m_n) = 1$. Since $N_1 = m_2 m_3 \cdots m_n$, we have $\gcd(m_1, N_1) = 1$. □

We can now prove the Chinese Remainder Theorem for integers.

PROOF. *Existence:* Write $N_i = N/m_i$ for $i = 1, 2, \ldots, n$. By Lemma 6.19, $\gcd(m_i, N_i) = 1$. Use the Extended Euclidean Algorithm to compute $a_1, b_1, a_2, b_2, \ldots, a_n, b_n$ such that

$$\begin{aligned} a_1 m_1 + b_1 N_1 &= 1 \\ a_2 m_2 + b_2 N_2 &= 1 \\ &\vdots \\ a_n m_n + b_n N_n &= 1. \end{aligned}$$

Put $x = \alpha_1 b_1 N_1 + \alpha_2 b_2 N_2 + \cdots + \alpha_n b_n N_n$. Now $b_1 N_1 = 1 + (-a_1)\, m_1$ so $[b_1 N_1] = [1]$ in $\mathbb{Z}_{m_1}$, so $[\alpha_1 b_1 N_1] = [\alpha_1]$ in $\mathbb{Z}_{m_1}$. Moreover, for $i = 2, 3, \ldots, n$ inspection of N_i verifies that $m_1 \mid N_i$, so $\alpha_i b_i N_i = q_i m_1$ for some $q_i \in \mathbb{Z}$, implying that $[\alpha_i b_i N_i] = [0]$. Hence

$$\begin{aligned} [x] &= [\alpha_1 b_1 N_1 + \alpha_2 b_2 N_2 + \cdots + \alpha_n b_n N_n] \\ &= [\alpha_1] + [0] + \cdots + [0] \end{aligned}$$

in $\mathbb{Z}_{m_1}$, as desired. A similar argument shows that $[x] = [\alpha_i]$ in $\mathbb{Z}_{m_i}$ for $i = 2, 3, \ldots, n$.

Uniqueness: As in the previous case, let $[x], [y]$ be two solutions to the system in $\mathbb{Z}_N$. Then $[x - y] = [0]$ in $\mathbb{Z}_{m_i}$ for $i = 1, 2, \ldots, n$, implying that $m_i \mid (x - y)$ for $i = 1, 2, \ldots, n$.

Since $m_1 \mid (x - y)$, the definition of divisibility implies that there exists $q_1 \in \mathbb{Z}$ such that $x - y = m_1 q_1$.

Since $m_2 \mid (x - y)$, substitution implies $m_2 \mid m_1 q_1$, and Lemma 6.15 implies that $m_2 \mid q_1$. The definition of divisibility implies that there exists $q_2 \in \mathbb{Z}$ such that $q_1 = m_2 q_2$. Substitution implies that $x - y = m_1 m_2 q_2$.

Since $m_3 \mid (x - y)$, substitution implies $m_3 \mid m_1 m_2 q_2$. Lemma 6.19 implies that $\gcd(m_1 m_2, m_3) = 1$, and Lemma 6.15 implies that $m_3 \mid q_2$. The definition of divisibility implies that there exists $q_3 \in \mathbb{Z}$ such that $q_2 = m_3 q_3$. Substitution implies that $x - y = m_1 m_2 m_3 q_3$.

Continuing in this fashion, we show that $x - y = m_1 m_2 \cdots m_n q_n$ for some $q_n \in \mathbb{Z}$. By substition, $x - y = N q_n$, so $[x - y] = [0]$ in $\mathbb{Z}_N$, so $[x] = [y]$ in $\mathbb{Z}_n$. That is, the solution to the system is unique in $\mathbb{Z}_N$. □

The algorithm to solve such systems is similar to that given for the simple version, in that it can be obtained from the proof of existence of a solution.

EXERCISES.

EXERCISE 6.20. Solve the system of linear congruences

$$\begin{cases} [x] &= [2] \text{ in } \mathbb{Z}_4; \\ [x] &= [2] \text{ in } \mathbb{Z}_9. \end{cases}$$

Express your answer so that $0 \leq x < 36$.

EXERCISE 6.21. Solve the system of linear congruences

$$\begin{cases} [x] &= [2] \text{ in } \mathbb{Z}_5; \\ [x] &= [2] \text{ in } \mathbb{Z}_6; \\ [x] &= [2] \text{ in } \mathbb{Z}_7. \end{cases}$$

EXERCISE 6.22. Solve the system of linear congruences

$$\begin{cases} [x] &= [33] \text{ in } \mathbb{Z}_{16}; \\ [x] &= [-4] \text{ in } \mathbb{Z}_{33}; \\ [x] &= [17] \text{ in } \mathbb{Z}_{504}. \end{cases}$$

Hint: This problem is a little tougher, since $\gcd(16, 504) \neq 1$ and $\gcd(33, 504)$. At least $\gcd(16, 33) = 1$, so you can apply the Chinese Remainder Theorem to the first two equations and find a solution in $\mathbb{Z}_{16 \cdot 33}$. Now you have to extend your solution so that it also solves the third equation; use your knowledge of cosets to do that.

EXERCISE 6.23. Give directions for a similar card trick on all 52 cards, where the cards are grouped first by 4's, then by 13's. Do you think this would be a practical card trick?

EXERCISE 6.24. Is it possible to modify the card trick to work with only ten cards instead of 12? If so, how; if not, why not?

EXERCISE 6.25. Is it possible to modify the card trick to work with only eight cards instead of 12? If so, how; if not, why not?

6.3. A NEW GROUP

In this section we find a subset of $\mathbb{Z}_n$ that we can turn into a multiplicative group. Before that, we need a little more number theory.

Before finishing that, we need some more technical details from number theory. *Warning:* the following definition is guaranteed to offend your sensibilities.

DEFINITION 6.26. Let $p \in \mathbb{Z}$ and assume $p > 1$. We say that p is **prime** if for any two integers a, b

$$p \mid ab \quad \Longrightarrow \quad p \mid a \text{ or } p \mid b.$$

EXAMPLE 6.27. Let $a = 68$ and $b = 25$. It is easy to recognize that 10 divides $ab = 1700$. However, 10 divides neither a nor b, so 10 is not a prime number.

It is also easy to recognize that 17 divides $ab = 1700$. Here, 17 must divide one of a or b, because it is prime. In fact, $17 \times 4 = 68 = a$. △

The definition of a prime number may surprise you, since ordinarily people think of a *prime* number as being *irreducible*. In fact, you will prove for homework:

THEOREM 6.28. *A positive integer is irreducible if and only if it is prime.*

If the two definitions are equivalent, why would we give a different definition? It turns out that the concepts are equivalent *only for the integers,* and not for some other sets; you will encounter one such set later in the course.

Primes are useful because every integer has a unique factorization into primes:

THEOREM 6.29. *Let $n \in \mathbb{Z}$ and assume $n > 1$. We can write*

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$$

where $p_1, p_2, \ldots, p_r$ are irreducible (hence, prime) and $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{N}_{\geq 0}$. Moreover, this representation is unique.

PROOF. The proof has two parts: a proof of existence and a proof of uniqueness.

Existence: We proceed by induction on the integers larger than or equal to two.

Inductive base: If $n = 2$, n is irreducible, and we are finished.

Inductive hypothesis: The integers $2, 3, \ldots, n-1$ satisfy the theorem, although each may have its own factorization.

Inductive step: If n is irreducible, then we are finished. Otherwise, n is not irreducible, so there exists an integer p_1 such that $p_1 \mid n$ and $p \neq \pm 1, n$. Choose the largest $\alpha_1 \in \mathbb{N}_{\geq 0}$ such that $p_1^{\alpha_1} \mid n$. Use the definition of divisibility (Definition 1.6 on page 13) to find $q \in \mathbb{Z}$ such that $n = q p_1$. By the definition of irreducible, we know that $p_1 \neq 1$, so $q < n$. Since p_1 is not negative, $q > 1$. Thus q satisfies the inductive hypothesis, and we can write $q = p_2^{\alpha_2} p_3^{\alpha_3} \cdots p_r^{\alpha_r}$. Thus

$$n = q p_1 = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$$

as claimed.

Uniqueness: Suppose that there exist $\alpha_1, \ldots, \alpha_r$ and $\beta_1, \ldots, \beta_r$ such that

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} = p_1^{\beta_1} p_2^{\beta_2} \cdots p_r^{\beta_r}.$$

Without loss of generality, we may assume that $\alpha_1 \leq \beta_1$. It follows that

$$p_2^{\alpha_2} p_3^{\alpha_3} \cdots p_r^{\alpha_r} = p_1^{\beta_1 - \alpha_1} p_2^{\beta_2} p_3^{\beta_3} \cdots p_r^{\beta_r}.$$

This equation implies that $p_1^{\beta_1 - \alpha_1}$ divides the expression on the left hand side of the equation. Since p_1 is irreducible, hence prime, $\beta_1 - \alpha_1 > 0$ implies that p_1 divides one of $p_2, p_3, \ldots, p_r$. This contradicts the irreducibility of $p_2, p_3, \ldots, p_r$. Hence $\beta_1 - \alpha_1 = 0$. A similar argument shows that $\beta_i = \alpha_i$ for all $i = 1, 2, \ldots, r$; hence the representation of n as a product of irreducible integers is unique. □

To turn $\mathbb{Z}_n$ into a multiplicative group, we would like to define multiplication in an "intuitive" way. By "intuitive", we mean that we would like to say

$$[2] \cdot [3] = [2 \cdot 3] = [6] = [1].$$

Before we can address the questions of whether $\mathbb{Z}_n$ can become a group under this operation, we have to remember that cosets can have various representations, and different representations may lead to different results: is this operation well-defined?

LEMMA 6.30. *The proposed multiplication of elements of $\mathbb{Z}_n$ as*

$$[a]\,[b] = [ab]$$

is well-defined.

PROOF. Let $x, y \in \mathbb{Z}_n$ and represent $x = [a] = [c]$ and $y = [b]$. Then

$$xy = [a][b] = [ab] \quad \text{and} \quad xy = [c][b] = [cb].$$

We need to show that $[ab] = [cb]$. Since these are sets, we have to show that each is a subset of the other.

By assumption, $[a] = [c]$; this notation means that $a + n\mathbb{Z} = c + n\mathbb{Z}$. Lemma 3.24 on page 43 tells us that $a - c \in n\mathbb{Z}$. Hence $a - c = nt$ for some $t \in \mathbb{Z}$. Now $(a-c)b = nu$ where $u = tb \in \mathbb{Z}$, so $ab - cb \in n\mathbb{Z}$. Lemma 3.24 again tells us that $[ab] = [cb]$ as desired, so the proposed multiplication of elements in $\mathbb{Z}_n$ is well-defined. □

EXAMPLE 6.31. Recall that $\mathbb{Z}_5 = \mathbb{Z}/\langle 5\rangle = \{[0], [1], [2], [3], [4]\}$. The elements of $\mathbb{Z}_5$ are cosets; since $\mathbb{Z}$ is an additive group, we were able to define easily an addition on $\mathbb{Z}_5$ that turns it into an additive group in its own right.

Can we also turn it into a multiplicative group?In that case, we need to identify an identity, and inverses. Certainly $[0]$ won't have a multiplicative inverse, but what about $\mathbb{Z}_5^* = \mathbb{Z}_5 \setminus \{[0]\}$? This generates a multiplication table that satisfies the properties of an abelian (but non-additive) group:

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

In fact, $\mathbb{Z}_5^* \cong \mathbb{Z}_4$; they are both the cyclic group of four elements. In $\mathbb{Z}_5^*$, however, the operation is multiplication, whereas in $\mathbb{Z}_4$ the operation is addition. ——————△

You might think that this trick of dropping zero and building a multiplication table always works, *but it doesn't*.

EXAMPLE 6.32. Recall that $\mathbb{Z}_4 = \mathbb{Z}/\langle 4\rangle = \{[0], [1], [2], [3]\}$. Consider the set $\mathbb{Z}_4 \setminus \{[0]\} = \{[1], [2], [3]\}$. The multiplication table for this set *is not closed* because

$$[2] \cdot [2] = [4] = [0] \notin \mathbb{Z}_4 \setminus \{[0]\}.$$

The next natural question: Is *any* subset of $\mathbb{Z}_4$ a multiplicative group? Try to fix the problem by removing $[2]$ as well: set $\mathbb{Z}_4^* = \mathbb{Z}_4 \setminus \{[0], [2]\} = \{[1], [3]\}$. This time the multiplication table works out:

×	1	3
1	1	3
3	3	1

In fact, $\mathbb{Z}_4^* \cong \mathbb{Z}_2$; they are both the cyclic group of two elements. In $\mathbb{Z}_4^*$, however, the operation is multiplication, whereas in $\mathbb{Z}_2$, the operation is addition. ——————△

You can determine for yourself that $\mathbb{Z}_2^* = \{[1]\}$ and $\mathbb{Z}_3^* = \{[1], [2]\}$ are also multiplicative groups. In this case, as in $\mathbb{Z}_5^*$, we need remove only 0. For $\mathbb{Z}_6^* = \{[1], [5]\}$, however, we have to remove nearly all the elements!

Why do we need to remove more numbers from $\mathbb{Z}_n$ for certain values of n than for others? Aside from zero, which clearly has no inverse under the operation specified, the elements we've had to remove are invariably those elements whose multiplication tries to re-introduce zero into the group. That already seems strange: we have non-zero elements that, when multiplied by

other non-zero elements, produce a product of zero. Here is an instance where $\mathbb{Z}_n$ superficially behaves *very differently* from the integers. Can we find a criterion to detect this?

LEMMA 6.33. *Let $x \in \mathbb{Z}_n$, with $x \neq [0]$. The following are equivalent:*

(A) There exists $y \in \mathbb{Z}_n$, $y \neq [0]$, such that $xy = [0]$.
(B) For any representation $[a]$ of x, there exists a common divisor d of a and n such that $d \neq \pm 1$.

PROOF. *That (B) implies (A):* If a and n share a common divisor d, use the definition of divisibility (Definition 1.6 on page 13) to choose q such that $n = qd$. Likewise choose t such that $a = td$. Then

$$qx = q[a] = q[td].$$

Exercise 6.38 implies that

$$q[td] = [qtd] = t[qd] = t[n] = [0].$$

Using the same Exercise6.38, we conclude that if $y = [q]$ then $xy = [0]$.

That (A) implies (B): Let $y \in \mathbb{Z}_n$, and suppose that $y \neq [0]$ but $xy = [0]$. Choose $a, b \in \mathbb{Z}$ such that $x = [a]$ and $y = [b]$. Since $xy = [0]$, we can find $k \in \mathbb{Z}$ such that $ab = kn$. Let p_0 be any irreducible number that divides n. Then p_0 also divides kn. Since $kn = ab$, we see that $p_0 \mid ab$. Since p_0 is irreducible, hence prime, it must divide one of a or b. If it divides a, then a and n have a common divisor p_0 that is not ± 1, and we are done; otherwise, it divides b. Use the definition of divisibility to find $n_1, b_1 \in \mathbb{Z}$ such that $n = n_1 p_0$ and $a = b_1 p_0$; it follows that $ab_1 = kn_1$. Again, let p_2 be any irreducible number that divides n_2; the same logic implies that p_2 divides ab_2; being prime, p_2 must divide a or b_2.

As long as we can find prime divisors of the n_i that divide b_i but not a, we repeat this process to find triplets $(n_2, b_2, p_2), (n_3, b_3, p_3), \ldots$ satisfying for all i the properties

- $ab_i = kn_i$; and
- $b_{i-1} = p_i b_i$ and $n_{i-1} = p_i n_i$.

By the well-ordering property, the set $\{n, n_1, n_2, \ldots\}$ has a least element; since $n > n_1 > n_2 \cdots$, we cannot continue finding pairs indefinitely, and must terminate with the least element (n_r, b_r). Observe that

$$b = p_1 b_1 = p_1(p_2 b_2) = \cdots = p_1(p_2(\cdots(p_r b_r))) \tag{6.3.1}$$

and

$$n = p_1 n_1 = p_1(p_2 n_2) = \cdots = p_1(p_2(\cdots(p_r n_r))).$$

Case 1. If $n_r > 1$, then n and a must have a common divisor that is not ± 1.

Case 2. If $n_r = 1$, then $n = p_1 p_2 \cdots p_r$. By substitution into equation 6.3.1, $b = nb_r$. By the definition of divisibility, $n \mid b$. By the definition of $\mathbb{Z}_n$, $y = [b] = [0]$. This contradicts the hypothesis.

Hence n and a share a common divisor that is not ± 1. □

Let's try then to make a *multiplicative* group out of the set of elements of $\mathbb{Z}_n$ that do not violate the zero product rule.

DEFINITION 6.34. Let $n \in \mathbb{Z}$. Let $x, y \in \mathbb{Z}_n$, and represent $x = [a]$ and $y = [b]$.

(1) Define a multiplication operation on $\mathbb{Z}_n$ by $xy = [ab]$.

(2) We say that $a, b \in \mathbb{Z}_n$ are **zero divisors** if $ab = 0$. (Of course, this is true also if $aa = 0$.) That is, zero divisors are the elements of $\mathbb{Z}_n$ that violate the zero-product property of multiplication.
(3) Define the set to be the set of elements in $\mathbb{Z}_n$ that are neither zero nor *not* zero divisors. That is,

$$\mathbb{Z}_n^* := \{x \in \mathbb{Z}_n \setminus \{0\} : \forall y \in \mathbb{Z}_n \, ab \neq 0\}.$$

By removing elements that share non-trivial common divisors with n, we have managed to eliminate those elements that do not satisfy the zero-product rule, and would break closure by trying to re-introduce zero in the multiplication table.

We claim that $\mathbb{Z}_n^*$ is a group under multiplication. Note that while it is a sub*set* of $\mathbb{Z}_n$, it is not a sub*group*: $\mathbb{Z}_n$ is not a group under multiplication, and subgroups maintain the operation of the parent group.

THEOREM 6.35. *$\mathbb{Z}_n^*$ is an abelian group under its multiplication.*

PROOF. We showed in Lemma 6.30 that the operation is well-defined. We check each of the requirements of a group:

(G1): Let $x, y \in \mathbb{Z}_n^*$; represent $x = [a]$ and $y = [b]$. By definition of $\mathbb{Z}_n^*$, a and b have no common divisors with n aside from ± 1; thus ab also has no common divisors with n aside from ± 1. As a result, $xy = [ab] \in \mathbb{Z}_n^*$.

(G2): Let $x, y, z \in \mathbb{Z}_n^*$; represent $x = [a]$, $y = [b]$, and $z = [c]$. Then

$$x(yz) = [a][bc] = [a(bc)] = [(ab)c] = [ab][c] = (xy)z.$$

(G3): We claim that $[1]$ is the identity of this group. Let $x \in \mathbb{Z}_n^*$; represent $x = [a]$. Then

$$x \cdot [1] = [a \cdot 1] = [a] = x;$$

a similar argument shows that $[1] \cdot x = x$.

(G4): Let $x \in \mathbb{Z}_n^*$. By definition of $\mathbb{Z}_n^*$, $x \neq 0$ and x is not a zero divisor in $\mathbb{Z}_n$. Represent $x = [m]$. Since $x \neq 0$, $m \notin \mathbb{Z}_n$, so $n \nmid m$. From Lemma 6.33, m and n have no common divisors except ± 1; hence $\gcd(m, n) = 1$. Using the Extended Euclidean Algorithm, find $a, b \in \mathbb{Z}$ such that $am + bn = 1$. Hence

$$\begin{aligned} am &= 1 + n(-b) \\ \therefore am &\in 1 + n\mathbb{Z} \\ \therefore am + n\mathbb{Z} &= 1 + n\mathbb{Z} \\ \therefore [am] &= [1] \\ \therefore [a][m] &= [1] \end{aligned}$$

by (respectively) the definition of the coset $1 + n\mathbb{Z}$, Lemma 3.24 on page 43, the notation for cosets of subgroups of $\mathbb{Z}_n$, and the definition of multiplication in $\mathbb{Z}_n^*$ given above. Let $y = [a]$; by substitution, the last equation becomes

$$yx = [1]$$

as claimed.

(G5) Let $x, y \in \mathbb{Z}_n^*$; represent $x = [a]$ and $y = [b]$. Then

$$xy = [ab] = [ba] = yx.$$

□

EXERCISES.

EXERCISE 6.36. List the elements of $\mathbb{Z}_7$ using representations between 0 and 7, and construct its multiplication table. Use the table to identify the inverse of each element.

EXERCISE 6.37. List the elements of $\mathbb{Z}_{15}$ using representations between 0 and 15, and construct its multiplication table. Use the table to identify the inverse of each element.

EXERCISE 6.38. Show that for any $a, x, n \in \mathbb{Z}$, $a[x] = [ax]$ in $\mathbb{Z}_n$. *Hint:* Remember that we are talking about repeated addition of cosets. Prove this fact for $a > 0$ using induction; for $a \leq 0$, it follows from the notation for additive sets.

EXERCISE 6.39. Let $p \in \mathbb{Z}$, and $p > 1$. Show that p is irreducible iff p is prime. *Hint:*

6.4. EULER'S NUMBER, EULER'S THEOREM, AND FAST EXPONENTIATION

In Section 6.3 we defined the group $\mathbb{Z}_n^*$ for all $n \in \mathbb{N}_{>1}$. This group satisfies an important property called *Euler's Theorem*. Much of what follows is related to some work of Euler, pronounced in a way that rhymes with "oiler". Euler was a very influential mathematician: You already know of Euler's number $e = \lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x \approx 2.718$; Euler is well-known for contributions to Calculus, Differential Equations, and to Number Theory. He was extremely prolific, and is said to have calculated the way "ordinary" men breathe. After losing his sight in one eye, he expressed his happiness at being only half as distracted from his work as he was before. He sired a large number of children, and used to work with one child sitting on one knee, and another child sitting on the other knee. He is, in short, the kind of historical figure that greatly lowers my self-esteem as a mathematician. Amazingly, Euler is well-known for his work in just about every area of mathematics *except* algebra.

DEFINITION 6.40. **Euler's φ-function** is $\varphi(n) = \left|\mathbb{Z}_n^*\right|$.

THEOREM 6.41 (Euler's Theorem). *For all* $x \in \mathbb{Z}_n^*$, $x^{\varphi(n)} = 1$.

Proofs of Euler's Theorem based only on Number Theory are not very easy. They aren't particularly difficult, either: they just aren't easy. See for example the proof on pages 18–19 of the text, *Concrete Abstract Algebra: from Numbers to Gröbner Bases* by Niels Lauritzen.

On the other hand, a proof of Euler's Theorem using algebra is trivial.

PROOF. Let $x \in \mathbb{Z}_n^*$. By Corollary 3.37 to Lagrange's Theorem, $\mathrm{ord}x \mid \left|\mathbb{Z}_n^*\right|$. Hence $\mathrm{ord}x \mid \varphi(n)$; use the definition of divisibility to write $\varphi(n) = d\,\mathrm{ord}x$ for some $d \in \mathbb{Z}$. Hence

$$x^{\varphi(n)} = x^{d\mathrm{ord}x} = \left(x^{\mathrm{ord}x}\right)^d = 1^d = 1.$$

□

COROLLARY 6.42. *For all* $x \in \mathbb{Z}_n^*$, $x^{\varphi(n)-1} = x^{-1}$.

PROOF. You do it! See Exercise 6.51. □

It thus becomes an important computational question to ask, how large is this group? For irreducible integers this is easy: if p is irreducible, $\varphi(p) = p-1$. For reducible integers, it is not so easy: using Definitions 6.40 and 6.34, $\varphi(n)$ is the number of positive integers smaller than n and sharing no common divisors with n. Checking a few examples, no clear pattern emerges:

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\lvert\mathbb{Z}_n^*\rvert$	1	2	2	4	2	6	4	6	4	10	4	12	6	8

Computing $\varphi(n)$ turns out to be quite hard for arbitrary $n \in \mathbb{N}_{>0}$. This difficulty is what makes the RSA algorithm secure (see Section 6.5.

One way to do it would be to factor n and compute all the positive integers that do not share any common factors. For example,

$$28 = 2^2 \cdot 7,$$

so to compute $\varphi(28)$, we could look at all the positive integers smaller than 28 that do not have 2 or 7 as factors. However, this is unsatisfactory: it requires us to try two divisions on all the positive integers between 2 and 28. That takes too long, and becomes even more burdensome when dealing with large numbers. There has to be a better way! Unfortunately, no one knows it.

One thing we can do is break n into its factors. Presumably, it would be easier to compute $\varphi(m)$ for these smaller integers m, but how to recombine them?

LEMMA 6.43. *Let* $n \in \mathbb{N}_{>0}$. *If* $n = pq$ *and* $\gcd(p,q) = 1$, *then* $\varphi(n) = \varphi(p)\varphi(q)$.

EXAMPLE 6.44. In the table above, we have $\varphi(15) = 8$. Notice that this satisfies

$$\varphi(15) = \varphi(5 \times 3) = \varphi(5)\varphi(3) = 4 \times 2 = 8.$$

△

PROOF. Recall from Exercise 2.42 on page 26 that $\mathbb{Z}_p^* \times \mathbb{Z}_q^*$ is a group; a counting argument shows that the size of this group is $\left|\mathbb{Z}_p^*\right| \times \left|\mathbb{Z}_q^*\right| = \varphi(p)\varphi(q)$. We show that $\mathbb{Z}_n^* \cong \mathbb{Z}_p^* \times \mathbb{Z}_q^*$.

Let $f : \mathbb{Z}_n^* \to \mathbb{Z}_p^* \times \mathbb{Z}_q^*$ by $f([a]_n) = \left([a]_p, [a]_q\right)$ where $[a]_i$ denotes the congruence class of a in $\mathbb{Z}_i$. First we show that f is a homomorphism: Let $a, b \in \mathbb{Z}_n^*$; then

$$\begin{aligned} f\left([a]_n [b]_n\right) = f\left([ab]_n\right) &= \left([ab]_p, [ab]_q\right) \\ &= \left([a]_p [b]_p, [a]_q [b]_q\right) \\ &= \left([a]_p, [a]_q\right)\left([b]_p, [b]_q\right) \\ &= f\left([a]_n\right) f\left([b]_n\right) \end{aligned}$$

(where Lemma 6.30 on page 104 and the definition of the operation in $\mathbb{Z}_p^* \times \mathbb{Z}_q^*$ justify the second two equations).

It remains to show that f is one-to-one and onto. We claim that this follows from the simple version of the Chinese Remainder Theorem, since the mapping f corresponds precisely to the

system of linear congruences

$$[x] = [a] \text{ in } \mathbb{Z}_p^*;$$
$$[x] = [b] \text{ in } \mathbb{Z}_q^*.$$

That f is onto follows from the fact that any such x exists in $\mathbb{Z}_n$; that f is one-to-one follows from the fact that x is unique in $\mathbb{Z}_n$.

We are not quite done; we have shown that a solution x exists in $\mathbb{Z}_n$, but we must show that more specifically $x \in \mathbb{Z}_n^*$. To see that indeed $x \in \mathbb{Z}_n^*$, let d by any common divisor of x and n. Let $c \in \mathbb{Z}$ such that $n = cd$. Let r be an irreducible divisor of d; then $r \mid n$. Now $n = pq$, so $r \mid pq$, so $r \mid p$ or $r \mid q$. Then d shares a common divisor with p or with q. However, $x \in \mathbb{Z}_p^*$ implies that $\gcd(x, p) = 1$; likewise, $\gcd(x, q) = 1$. Since d is a common divisor of x and p or q, $d = 1$. Since it was an arbitrary common divisor of x and n, $\gcd(x, n) = 1$; hence $x \in \mathbb{Z}_n^*$ and f is one-to-one. □

Corollary 6.42 on the preceding page gives us an "easy" way to compute the inverse of any $x \in \mathbb{Z}_n^*$. However, it can take a long time to compute $x^{\varphi(n)}$, so we conclude with a brief discussion of how to compute exponents in this group. We will take two steps towards a fast exponentiation in $\mathbb{Z}_n^*$.

LEMMA 6.45. *For any $n \in \mathbb{N}_{>1}$, $[x^a] = [x]^a$ in $\mathbb{Z}_n^*$.*

PROOF. You do it! See Exercise 6.53 on page 112. □

EXAMPLE 6.46. In $\mathbb{Z}_{15}^*$ we can easily determine that $\left[4^{20}\right] = [4]^{20} = \left([4]^2\right)^{10} = [16]^{10} = [1]^{10} = [1]$. Notice that this is a *lot* faster than computing $4^{20} = 1099511627776$ and dividing to compute the remainder. △

THEOREM 6.47. *Let $a \in \mathbb{N}$ and $x \in \mathbb{Z}$. We can compute x^a in the following way:*

(1) *Let b be the largest integer such that $2^b \leq a$.*
(2) *Use the Division Theorem to divide a repeatedly by 2^b, 2^{b-1}, …, 2^1, 2^0 in that order; let the quotients of each division be q_b, q_{b-1}, …, q_1, q_0.*
(3) *Write $a = q_b 2^b + q_{b-1} 2^{b-1} + \cdots + q_1 2^1 + q_0 2^0$.*
(4) *Let $y = 1$, $z = x$ and $i = 0$.*
(5) *Repeat the following until $i > b$:*
    (a) *If $q_i \neq 0$ then replace y with the product of y and z.*
    (b) *Replace z with z^2.*
    (c) *Replace i with $i + 1$.*

When the repetition stops, $x^a = y$.

Theorem6.47 effectively computes the *binary representation* of a and uses this to square x repeatedly, multiplying the result only by those powers that matter for the representation. Its algorithm is especially effective on computers, whose mathematics is based on binary arithmetic. Combining it with Lemma 6.45 gives an added bonus.

EXAMPLE 6.48. Since $10 = 2^3 + 2^1$, we can compute

$$4^{10} = 4^{2^3 + 2^1}$$

by following the algorithm of Theorem 6.47:

(1) We have $q_3 = 1$, $q_2 = 0$, $q_1 = 1$, $q_0 = 0$.
(2) Let $y = 1$, $z = 4$ and $i = 0$.
(3) When $i = 0$:
    (a) We do not change y because $q_0 = 0$.
    (b) Put $z = 4^2 = 16$.
    (c) Put $i = 1$.
(4) When $i = 1$:
    (a) Put $y = 1 \cdot 16 = 16$.
    (b) Put $z = 16^2 = 256$.
    (c) Put $i = 2$.
(5) When $i = 2$:
    (a) We do not change y because $q_2 = 0$.
    (b) Put $z = 256^2 = 65,536$.
    (c) Put $i = 3$.
(6) When $i = 3$:
    (a) Put $y = 16 \cdot 65,536 = 1,048,576$.
    (b) Put $z = 65,536^2 = 4,294,967,296$.
    (c) Put $i = 4$.

We conclude that $4^{10} = 1,048,576$. Hand computation the long way, or a half-decent calculator, will verify this. △

PROOF OF FAST EXPONENTIATION. *Termination:* Termination follows from the fact that b is a finite number, and the algorithm assigns to i the values $0, 1, \ldots, b+1$ in succession.

Correctness: Since b is the largest integer such that $2^b \leq a$, $q_b \in \{0,1\}$; otherwise, $2^{b+1} = 2 \cdot 2^b \leq a$, contradicting the choice of b. For $i = b-1, \ldots, 1, 0$, we have the remainder from division by 2^{i+1} smaller than 2^i, and we immediately divide by $2^b = 2^{i-1}$, so that $q_i \in \{0,1\}$ as well. Hence $q_i \in \{0,1\}$ for $i = 0, 1, \ldots, b$ and if $q_i \neq 0$ then $q_i = 1$. The algorithm therefore multiplies $z = x^{2^i}$ to y only if $q_i \neq 0$, which agrees with the binary representation

$$x^a = x^{q_b 2^b + q_{b-1} 2^{b-1} + \cdots + q_1 2^1 + q_0 2^0}.$$

□

EXERCISES.

EXERCISE 6.49. Compute 3^{28} in $\mathbb{Z}$ using fast exponentiation. Show the steps of exponentiation.

EXERCISE 6.50. Compute 24^{28} in $\mathbb{Z}_7^*$ using fast exponentiation. Show the steps of exponentiation.

EXERCISE 6.51. Prove that for all $x \in \mathbb{Z}_n^*$, $x^{\varphi(n)-1} = x^{-1}$.

EXERCISE 6.52. Prove that for all $x \in \mathbb{N}_{>0}$, if x and n have no common divisors, then $n \mid \left(x^{\varphi(n)} - 1\right)$.

EXERCISE 6.53. Prove that for any $n \in \mathbb{N}_{>1}$, $[x^a] = [x]^a$ in $\mathbb{Z}_n^*$. *Hint:* Consider the factorization of a into irreducibles, and Lemma 6.30 on page 104.

6.5. THE RSA ENCRYPTION ALGORITHM

From the viewpoint of practical applications, some of the most important results of group theory and number theory are those that enable security in internet commerce. We described this problem in Section 1.1: when you buy something online, you usually submit some private information, in the form either of a credit card number or a bank account number. There is no guarantee that, as this information passes through the internet, it passes through trustworthy computers. In fact, it is quite likely that the information sometimes passes through a computer run by at least one ill-intentioned hacker, and possibly even organized crime. Identity theft has emerged in the last few decades as an extremely profitable pursuit.

Given the inherent insecurity of the internet, the solution is to disguise your private information so that disreputable snoopers cannot understand it. Mathematicians discovered a long time ago that mathematics provides a highly reliable method both of analyzing and of creating methods of encryption. A common method in use today is the RSA encryption algorithm.[1] First we describe the algorithms for encryption and decryption; afterwards we explain the ideas behind each stage, illustrating with an example; finally we prove that it succesfully encrypts and decrypts messages.

THEOREM 6.54 (RSA algorithm). *Let M be a list of positive integers obtained by converting the letters of a message. Let p, q be two irreducible integers that satisfy the following two criteria:*

- $\gcd(p,q) = 1$; *and*
- $(p-1)(q-1) > \max\{m : m \in M\}$.

Let $N = pq$, and let $e \in \mathbb{Z}^*_{\varphi(N)}$, where φ is the Euler phi-function. If we apply the following algorithm to M:

(1) *Let C be a list of positive integers found by computing $[m^e] \in \mathbb{Z}_N$ for each $m \in M$.*

and subsequently apply the following algorithm to C:

(1) *Let $d = e^{-1} \in \mathbb{Z}^*_{\varphi(N)}$.*

(2) *Let D be a list of positive integers found by computing $c^d \in \mathbb{Z}_N$ for each $c \in C$.*

then $D = M$.

EXAMPLE 6.55. Consider the text message

ALGEBRA RULZ.

We will convert the letters to integers in the fashion that you might expect: A=1, B=2, ..., Z=26. We will also assign 0 to the space. Thus

$$M = (1, 12, 7, 5, 2, 18, 1, 0, 18, 21, 12, 26).$$

Let $p = 5$ and $q = 11$; then $N = 55$. Let $e = 3$; note that

$$\gcd(3, \varphi(N)) = \gcd(3, 4 \times 10) = \gcd(3, 40) = 1.$$

[1] RSA stands for Rivest, Shamir, and Adleman, three researchers at MIT.

We encrypt by computing m^e for each $m \in M$:

$$\begin{aligned} C &= \left(1^3, 12^3, 7^3, 5^3, 2^3, 18^3, 1^3, 0^3, 18^3, 21^3, 12^3, 26^3\right) \\ &= (1, 23, 13, 15, 8, 2, 1, 0, 2, 21, 23, 31)\,. \end{aligned}$$

A snooper who intercepts C and tries to read it as a plain message would have a problem, since it contains a number that does not fall in the range 0 and 26. If he gave that number the symbol _, he would see

AWMOHBA BUW_

which is not an obvious encryption of ALGEBRA RULZ.

The inverse of $3 \in \mathbb{Z}_{40}^*$ is $d = 27$ (since $3 \times 27 = 81$ and $[81] = [1]$ in $\mathbb{Z}_{40}^*$). We decrypt by computing c^d for each $c \in C$:

$$\begin{aligned} D &= \left(1^{27}, 23^{27}, 13^{27}, 15^{27}, 8^{27}, 2^{27}, 1^{27}, 0^{27}, 2^{27}, 21^{27}, 23^{27}, 31^{27}\right) \\ &= (1, 12, 7, 5, 2, 18, 1, 0, 18, 21, 12, 26)\,. \end{aligned}$$

Trying to read this as a plain message, we have

ALGEBRA RULZ.

It does, doesn't it?. △

A few observations are in order.

(1) Usually encryption is not done letter-by-letter; instead, letters are grouped together and converted to integers that way. For example, the first four letters of the secret message above are

ALGE

and we can convert this to a number using any of several methods; for example

$$\text{ALGE} \quad \to \quad 1 \times 26^3 + 12 \times 26^2 + 7 \times 26 + 5 = 25{,}785.$$

In order to encrypt this, we would need larger values for p and q. We give an example of this in the homework.

(2) RSA is an example of a *public-key cryptosystem.* In effect that means that person A broadcasts to the world, "Anyone who wants to send me a secret message can use the RSA algorithm with values $N = \ldots$ and $e = \ldots$." Even the snooper knows N and e!

(3) If even the snooper knows N and e, what makes RSA safe? To decrypt, the snooper needs to compute $d = e^{-1} \in \mathbb{Z}_{\varphi(N)}^*$. This would be relatively easy if he knew $\varphi(N)$. There is no known method of computing $\varphi(N)$ "quickly". If p and q are small, of course this isn't hard: one simply tries to factor N; Lemma 6.43 tells us that $\varphi(N) = (p-1)(q-1)$. In practice, however, p and q are *very* large numbers (many digits long). There is a careful science to choosing p and q in such a way that it is hard to determine their values from N and e.

(4) It is time-consuming to perform these computations by hand; a computer algebra system will do the trick nicely. At the end of this section, after the exercises, we list programs that will help you perform these computations in the Sage and Maple computer algebra systems. The programs are:

- `scramble`, which accepts as input a plaintext message like "ALGEBRA RULZ" and turns it into a list of integers;

- `descramble`, which accepts as input a list of integers and turns it into plaintext;
- `en_de_crypt`, which encrypts or decrypts a message, depending on whether you feed it the encryption or decryption exponent.

Examples of usage:

- in Sage:
  - to determine the list of integers M, type `M = scramble("ALGEBRA RULZ")`
  - to encrypt M, type `C = en_de_crypt(M,3,55)`
  - to decrypt C, type `en_de_crypt(C,27,55)`
- in Maple:
  - to determine the list of integers M, type `M := scramble("ALGEBRA RULZ");`
  - to encrypt M, type `C := en_de_crypt(M,3,55);`
  - to decrypt C, type `en_de_crypt(C,27,55);`

Now, *why* does the RSA algorithm work?

PROOF OF THE RSA ALGORITHM. Let $i \in \{1,2,\ldots,|C|\}$. Let $c \in C$. By definition of C, $c = m^e \in \mathbb{Z}_N^*$ for some $m \in M$. We need to show that $c^d = (m^e)^d = m$.

Since $\gcd(e, \varphi(N)) = 1$, the Extended Euclidean Algorithm tells us that there exist $a, b \in \mathbb{Z}$ such that

$$1 = ae + b\varphi(N).$$

Rearranging the equation, we see that

$$1 - ae = b\varphi(N);$$

in other words, $[1-ae] = [0] \in \mathbb{Z}_{\varphi(N)}$, so that $[1] = [a][e] \in \mathbb{Z}_{\varphi(N)}$. By definition of an inverse, $[a] = [e]^{-1} = [d] \in \mathbb{Z}_{\varphi(N)}^*$. (Notice that we omitted the star previously, but now we include it.)

Without loss of generality, $d, e > 0$, which implies that $b < 0$. Let $c = -b$. Substitution gives us

$$(m^e)^d = m^{ed} = m^{ae} = m^{1-b\varphi(N)} = m^{1+c\varphi(N)}.$$

We claim that $[m]^{1+c\varphi(N)} = [m] \in \mathbb{Z}_N$. This requires us to show two subclaims.

CLAIM. $[m]^{1+c\varphi(N)} = [m] \in \mathbb{Z}_p$.

If $p \mid m$, then $[m] = [0] \in \mathbb{Z}_p$, and

$$[m]^{1+c\varphi(N)} = [0]^{1+c\varphi(N)} = [0] = [m] \in \mathbb{Z}_p.$$

Otherwise, recall that p is irreducible; then $\gcd(m, p) = 1$ and by Euler's Theorem on page 108

$$[m]^{\varphi(p)} = [m]^{p-1} = [1] \in \mathbb{Z}_p^*.$$

Thus

$$[m]^{1+c\varphi(N)} = [m] \cdot [m]^{c\varphi(N)} = [m] \left([m]^{\varphi(N)}\right)^c = [m] \cdot [1]^c = [m] \in \mathbb{Z}_p^*.$$

What is true for $\mathbb{Z}_p^*$ is also true in $\mathbb{Z}_p$, since the former is a subset of the latter. Hence

$$[m]^{1+c\varphi(N)} = [m] \in \mathbb{Z}_p.$$

CLAIM. $[m]^{1+c\varphi(N)} = [m] \in \mathbb{Z}_q$.

The argument is similar to that of the first claim.

Since $[m]^{1+c\varphi(N)} = [m]$ in both $\mathbb{Z}_p$ and $\mathbb{Z}_q$, properties of the quotient groups $\mathbb{Z}_p$ and $\mathbb{Z}_q$ tell us that $\left[m^{1+c\varphi(N)} - m\right] = [0]$ in both $\mathbb{Z}_p$ and $\mathbb{Z}_q$ as well. In other words, both p and q divide $m^{1+c\varphi(N)} - m$. You will show in Exercise 115 that this implies that N divides $m^{1+c\varphi(N)} - m$. □

EXERCISES.

EXERCISE 6.56. The phrase

$$[574, 1, 144, 1060, 1490, 0, 32, 1001, 574, 243, 533]$$

is the encryption of a message using the RSA algorithm with the numbers $N = 1535$ and $e = 5$. You will decrypt this message.

(a) Factor N.
(b) Compute $\varphi(N)$.
(c) Find the appropriate decryption exponent. *Hint:* Using the Extended Euclidean Algorithm might make this go faster. The proof of the RSA algorithm outlines how to use it.
(d) Decrypt the message.

EXERCISE 6.57. In this exercise, we encrypt a phrase using more than one letter in a number.

(a) Rewrite the phrase GOLDEN EAGLES as a list M of three positive integers, each of which combines four consecutive letters of the phrase.
(b) Find two prime numbers whose product is larger than the largest number you would get from four letters. *Hint:* That largest number should come from encrypting ZZZZ.
(c) Use those two prime numbers to compute an appropriate N and e to encrypt M using RSA.
(d) Find an appropriate d that will decrypt M using RSA. *Hint:* Using the Extended Euclidean Algorithm might make this go faster. The proof of the RSA algorithm outlines how to use it.
(e) Decrypt the message to verify that you did this correctly.

EXERCISE 6.58. Let $m, p, q \in \mathbb{Z}$ and suppose that $\gcd(p, q) = 1$. Show that if $p \mid m$ and $q \mid m$, then $pq \mid m$. *Hint:* There are a couple of ways to argue this. The best way for you is to explain why there exist a, b such that $ap + bq = 1$. Next, explain why there exist integers d_1, d_2 such that $m = d_1 a$ and $m = d_2 b$. Observe that $m = m \cdot 1 = m \cdot (ap + bq)$. Put all these facts together to show that $ab \mid m$.

SAGE PROGRAMS. The following programs can be used in Sage to help make the amount of computation involved in the exercises less burdensome:

```
def scramble(s):
  result = []
  for each in s:
    if ord(each) >= ord("A") and ord(each) <= ord("Z"):
      result.append(ord(each)-ord("A")+1)
    else:
      result.append(0)
  return result


def descramble(M):
  result = ""
  for each in M:
    if each == 0:
      result = result + " "
    else:
      result = result + chr(each+ord("A") - 1)
  return result


def en_de_crypt(M,p,N):
  result = []
  for each in M:
    result.append((each^p).mod(N))
  return result
```

MAPLE PROGRAMS. The following programs can be used in Maple to help make the amount of computation involved in the exercises less burdensome:

```
scramble := proc(s)
  local result, each, ord;
  ord := StringTools[Ord];
  result := [];
  for each in s do
    if ord(each) >= ord("A") and ord(each) <= ord("Z") then
      result := [op(result),
        ord(each) - ord("A") + 1];
    else
      result := [op(result), 0];
    end if;
  end do;
  return result;
end proc:


descramble := proc(M)
  local result, each, char, ord;
  char := StringTools[Char];
  ord := StringTools[Ord];
  result := "";
  for each in M do
    if each = 0 then
      result := cat(result, " ");
    else
      result := cat(result, char(each + ord("A") - 1));
    end if;
  end do;
  return result;
end proc:


en_de_crypt := proc(M,p,N)
  local result, each;
  result := [];
  for each in M do
    result := [op(result), (each^p) mod N];
  end do;
  return result;
end proc:
```

Part 2

Elementary ring theory

CHAPTER 7

Rings and ideals

7.1. RINGS

Groups are simple in the following respect: a group is defined by a set and *one* operation. When we studied the set of matrices $\mathbb{R}^{m\times n}$ as a group, for example, we considered only the operation of addition. Likewise, when we studied $\mathbb{Z}$ as a group, we considered only the operation of addition. With other groups, we studied other operations, but we only studied one operation at a time.

Besides adding matrices or integers, one can also multiply matrices or integers. We can deal with multiplication independently of addition by restricting the set in certain ways—using the subset $\mathrm{GL}_m(\mathbb{R})$, for example. In some cases, however, we want to analyze how both addition and multiplication interact in a given set. This motivates the study of a structure that incorporates common properties of both operations.

DEFINITION 7.1. Let R be a set *with at least two elements*, and $+$ and $\times$ two operations on that set. We say that $(R,+,\times)$ is a **ring** if it satisfies the following properties:

(R1) $(R,+)$ is an abelian group.
(R2) R is closed under multiplication: that is,
for all $a,b\in R$, $ab\in R$.
(R3) R is associative under multiplication: that is,
for all $a,b,c\in R$, $(ab)c=a(bc)$.
(R4) R satisfies the distributive property of addition over multiplication: that is,
for all $a,b,c\in R$, $a(b+c)=ab+ac$ and $(a+b)c=ac+bc$.

NOTATION. As with groups, we usually refer simply to R as a group, rather than $(R,+,\times)$.

Since $(R,+)$ is an abelian group, the ring has an additive identity, 0. We sometimes write 0_R to emphasize that it is the additive identity of a ring. Likewise, if there is a multiplicative identity, we write 1 or 1_R, *not e*.

Notice the following:

- While the addition is guaranteed to be commutative by (R1), we have *not* stated that multiplication is commutative. Indeed, our first example ring has non-commutative for multiplication.
- There is not requirement that a multiplicative identity exists.
- There is no requirement that multiplicative inverses exist.
- There is no guarantee (yet) that the additive identity satisfies any properties that you remember from past experience: in particular, there is *no guarantee* that
  - the zero-product rule holds; or even that
  - $0_R\cdot a=0_R$ for any $a\in R$.

EXAMPLE 7.2. Let $R = \mathbb{R}^{m\times m}$ for some positive integer m. It turns out that R is a ring under the usual addition and multiplication of matrices. We pass over the details, but they can be found in any reputable linear algebra book.

We do want to emphasize the following. Let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Routine computation shows that

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

even though $A, B \neq 0$. Hence

We can never assume in any ring R the **zero product property** that

$$\forall a, b \in R \qquad ab = 0 \implies a = 0 \text{ or } b = 0.$$

△

The previous observation motivates a definition that we will explore later:

DEFINITION 7.3. Let D be a ring. If the elements of D satisfy the zero product property, then we call D a **domain**.

EXAMPLE 7.4. Although $\mathbb{R}^{m\times m}$ is not a domain, $\mathrm{GL}_m(\mathbb{R})$ is: Let $A, B \in \mathrm{GL}_m(\mathbb{R})$. Assume $AB = 0$ but $A \neq 0$. Thus A^{-1} exists and

$$\begin{aligned} AB &= 0 \\ A^{-1}(AB) &= A^{-1} \cdot 0 \\ B &= 0. \end{aligned}$$

Since A, B were arbitrary in $\mathrm{GL}_m(\mathbb{R})$, $\forall A, B \in \mathrm{GL}_m(\mathbb{R})$ if $AB = 0$ then one of A, B is also the zero matrix. That is, $\mathrm{GL}_m(\mathbb{R})$ is a domain. △

Likewise, the following sets with which you are long familiar are also rings:

- $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ under their usual addition and multiplication;
- the sets of univariate polynomials $\mathbb{Z}[x]$, $\mathbb{Q}[x]$, $\mathbb{R}[x]$, $\mathbb{C}[x]$ under their usual addition and multiplication;
- the sets of multivariate polynomials $\mathbb{Z}[x_1, \ldots, x_n]$, etc. under their usual addition and multiplication.

You will study other example rings in the exercises. For now, we prove a familiar property of the additive identity.

PROPOSITION 7.5. *For all $r \in R$, $r \cdot 0_R = 0_R \cdot r = 0_R$.*

PROOF. Since $(R, +)$ is an abelian group, we know that $0_R + 0_R = 0_R$. Let $r \in R$. By substitution, $r(0_R + 0_R) = r \cdot 0_R$. By distribution, $r \cdot 0_R + r \cdot 0_R = r \cdot 0_R$. Since $(R, +)$ is an abelian group, $r \cdot 0_R$ has an additive inverse; call it s. Substitution followed by the associative, inverse, and identity properties implies that

$$\begin{aligned} s + (r \cdot 0_R + r \cdot 0_R) &= s + r \cdot 0_R \\ (s + r \cdot 0_R) + r \cdot 0_R &= s + r \cdot 0_R \\ 0_R + r \cdot 0_R &= 0_R \\ r \cdot 0_R &= 0_R. \end{aligned}$$

A similar argument shows that $0_R \cdot r = 0_R$. □

We now turn our attention to two properties that, while pleasant, are not necessary for a ring.

DEFINITION 7.6. Let R be a ring. If R has a multiplicative identity 1_R such that

$$r \cdot 1_R = 1_R \cdot r = r \quad \forall r \in R,$$

we say that R is a **ring with unity**. (Another name for the multiplicative identity is **unity**.)

If R is a ring and the multiplicative operation is commutative, so that

$$rs = sr \quad \forall r \in R,$$

then we say that R is a **commutative ring**.

EXAMPLE 7.7. The set of matrices $\mathbb{R}^{m\times m}$ is a ring with unity, with the identity matrix I_m as the multiplicative identity. However, it is not a commutative ring.

You will show in Exercise 7.9 that $2\mathbb{Z}$ is a ring. It is also a commutative ring, but it is not a ring with unity.

For a commutative ring with unity, we have $\mathbb{Z}$. △

We conclude this section by characterizing all rings with only two elements.

EXAMPLE 7.8. Let R be a ring with only two elements. There are two possible structures for R.

Why? Since $(R,+)$ is an abelian group, by Section 2.1 the addition table of R has the form

$+$	0_R	a
0_R	0_R	a
a	a	0_R

.

By Proposition 7.5, we know that the multiplication table *must* have the form

$\times$	0_R	a
0_R	0_R	0_R
a	0_R	?

where $a \cdot a$ is undetermined. Nothing in the properties of a ring tell us whether $a \cdot a = 0_R$ or $a \cdot a = a$; in fact, rings exist with both properties:

- if $R = \mathbb{Z}_2^*$ (see Exercise 7.10 to see that this is a ring) then $a = [1]$ and $a \cdot a = a$; but
- if

$$R = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, a = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right\} \subset \left(\mathbb{Z}_2^*\right)^{2\times 2}$$

(two-by-two matrices whose entries are elements of $\mathbb{Z}_2^*$), then $a \cdot a = 0 \neq a$. △

EXERCISES.

EXERCISE 7.9. (a) Show that $2\mathbb{Z}$ is a ring under the usual addition and multiplication of integers.

(b) Show that $n\mathbb{Z}$ is a ring for all $n \in \mathbb{Z}$ under the usual addition and multiplication of integers. *Hint:* The cases where $n = 0$ and $n = 1$ can be disposed of rather quickly; the case where $n \neq 0, 1$ is similar to (a).

(c) Is $n\mathbb{Z}$ a domain for all non-zero $n \in \mathbb{Z}$?

EXERCISE 7.10. (a) Show that $\mathbb{Z}_2$ is a ring under the addition and multiplication of cosets defined in Sections 3.5 and 6.3.
(b) Show that $\mathbb{Z}_n$ is a ring for all $n \in \mathbb{Z}$ where $n > 1$, under the addition and multiplication of cosets defined in Sections 3.5 and 6.3.
(c) Is $\mathbb{Z}_n$ a domain for all non-zero $n \in \mathbb{Z}$?

EXERCISE 7.11. Let R be a ring.
(a) Show that for all $r, s \in R$, $(-r)s = r(-s) = -(rs)$. *Hint:* This is short, but not trivial. You need to show that $(-r)s + rs = 0_R$. Try using the distributive property.
(b) Suppose that R has unity. Show that $-r = -1_R \cdot r$ for all $r \in R$. *Hint:* You need to show that $-1_R \cdot r + r = 0$. Try using a proof similar to part (a), but work in the additive identity as well.

EXERCISE 7.12. Let R be a ring with unity. Show that $1_R \neq 0_R$. *Hint:* Proceed by contradiction. Show that if $r \in R$ and $r \neq 0, 1$, then something goes terribly wrong with multiplication in the ring.

EXERCISE 7.13. Consider the two possible ring structures from Example 7.8. Show that if a ring R has only two elements, one of which is unity, then it can have only one of the structures. *Hint:* Use the result of Exercise 7.12.

EXERCISE 7.14. Let $R = \{T, F\}$ with the additive operation $\oplus$ (Boolean xor) where

$$F \oplus F = F$$
$$F \oplus T = T$$
$$T \oplus F = T$$
$$T \oplus T = F$$

and a multiplicative operation $\wedge$ (Boolean and) where

$$F \wedge F = F$$
$$F \wedge T = F$$
$$T \wedge F = F$$
$$T \wedge T = T.$$

(see also Exercises 2.10 and 2.11 on page 18). Is $(R, \oplus, \wedge)$ a ring? If it is a ring, what is the zero element? Is it a domain? *Hint:* You already know that $(B, \oplus)$ is an additive group, so it remains to decide whether $\wedge$ satisfies the requirements of multiplication in a ring.

EXERCISE 7.15. Let R be a ring.
(a) Show that $R[x]$ is a ring.
(b) Show that $R[x, y]$ is a ring.
(c) Show that $R[x_1, x_2, \ldots, x_n]$ is a ring.

7.2. IDEALS AND QUOTIENT RINGS

Just as groups have subgroups, rings have subrings:

DEFINITION 7.16. Let R be a ring, and $S \subset R$. If S is also a ring under the same operations as R, then S is a subring of R.

EXAMPLE 7.17. Recall from Exercise 7.9 that $2\mathbb{Z}$ is a ring. It is also a subset of $\mathbb{Z}$, another ring. Hence $2\mathbb{Z}$ is a subring of $\mathbb{Z}$.

To show that a subset of a ring is a subring, do we have to show all four ring properties? No: as with subgroups, we can simplify the characterization, but to two properties:

THEOREM 7.18 (The Subring Theorem). *Let R be a ring and $S \subset R$. The following are equivalent:*

(A) S is a subring of R.
(B) S is closed under subtraction and multiplication. That is, for all $a, b \in S$
  (S1) $a - b \in S$, and
  (S2) $ab \in S$.

PROOF. That (A) implies (B) is clear, so assume (B). From (B) we know that for any $a, b \in S$ we have (S1) and (S2). Now (S1) is essentially the Subgroup Theorem (Theorem 3.5 on page 37) so S is an additive subgroup of the additive group R. On the other hand, (S2) only tells us that S satisfies property $(R2)$ of a ring, but any elements of S are elements of R, so that the associative and distributive properties follow from inheritance. Thus S is a ring in its own right, which makes it a subring of R. □

You might think that, just as we moved from subgroups to quotient groups via cosets, we will move from subrings to "quotient rings" via the ring analogue of cosets. No, actually: although we are moving to something called a "quotient ring", and we will build an analogue of cosets, we won't do it with subrings! Instead, we will use a special class of subrings called *ideals*.

DEFINITION 7.19. Let A be a subring of R that satisfies the **absorption property**:

$$\forall r \in R \; \forall a \in A \quad ra \in A.$$

We say that A is an **ideal subring** of R, or simply, an **ideal**. We write .

EXAMPLE 7.20. Recall the subring $2\mathbb{Z}$ of the ring $\mathbb{Z}$. We show that $2\mathbb{Z} \triangleleft \mathbb{Z}$: let $r \in \mathbb{Z}$, and $a \in 2\mathbb{Z}$. By definition of $2\mathbb{Z}$, there exists $d \in \mathbb{Z}$ such that $a = 2d$. Substitution gives us

$$ra = r \cdot 2d = 2(rd) \in 2\mathbb{Z},$$

so $2\mathbb{Z}$ "absorbs" multiplication by $\mathbb{Z}$. This makes $2\mathbb{Z}$ an ideal of $\mathbb{Z}$.

Naturally, we can generalize this proof to arbitrary $n \in \mathbb{Z}$: see Exercise . △

The absorption property of ideals distinguishes them from other subrings, and makes them useful for applications.

EXAMPLE 7.21. Let $\mathbb{C}[x,y]$ be the set of all polynomials in x and y with complex coefficients. You showed in Exercise 7.15 that this is a ring.

Now let $f = x^2 + y^2 - 1$, $g = xy - 1$. Define $A = \{hf + kg : h, k \in \mathbb{C}[x,y]\}$. We claim that A is an ideal:

- For any $a, b \in A$, we can by definition of A write $a = h_a f + k_a g$ and $b = h_b f + k_b g$ for some $h_a, h_b, k_a, k_b \in \mathbb{C}[x,y]$. Thus

- $a-b=(h_af+k_ag)-(h_bf+k_bg)=(h_a-h_b)f+(k_a-k_b)g\in A$; and
- $ab=(h_af+k_ag)(h_bf+k_bg)=h_ah_bf^2+h_ak_bfg+h_bk_afg+k_ak_bg^2=h'f+k'g$ where

$$h'=h_ah_bf+h_ak_bg+h_bk_ag \quad \text{and} \quad k'=k_ak_bg,$$

which shows that ab has the form of an element of A. Thus $ab\in A$ as well.

By the Subring Theorem, A is a subring of $\mathbb{C}[x,y]$.

- For any $a\in A$, $r\in\mathbb{C}[x,y]$, write a as before; then

$$ra=r(h_af+k_ag)=(rh_a)f+(rk_a)g=h'f+k'g$$

where $h'=rh_a$ and $k'=rk_a$. This shows that ra has the form of an element of A, so $ra\in A$.

We have shown that A satisfies the subring and absorption properties; thus, $A\triangleleft\mathbb{C}[x,y]$.

What's most interesting about A is the following algebraic fact: *the common of f and g are roots of any element of A.* To see this, let (α,β) be a common root of f and g; that is, $f(\alpha,\beta)=g(\alpha,\beta)=0$. Let $p\in A$; by definition of A we can write $p=hf+kg$ for some $h,k\in\mathbb{C}[x,y]$. Substitution shows us that

$$\begin{aligned}p(\alpha,\beta)&=(hf+kg)(\alpha,\beta)\\&=h(\alpha,\beta)\cdot f(\alpha,\beta)+k(\alpha,\beta)\cdot g(\alpha,\beta)\\&=h(\alpha,\beta)\cdot 0+k(\alpha,\beta)\cdot 0\\&=0;\end{aligned}$$

that is, (α,β) is a root of p. △

You should recall from linear algebra that *vector spaces* are an important tool for the study of systems of linear equations: finding a *triangular basis* of the vector space spanned by a system of linear polynomials allows us to analyze the solutions. Example 7.21 illustrates why ideals are an important tool for the study of non-linear polynomial equations. If one can compute a "triangular basis" of a polynomial ideal, then one can analyze the solutions in a method very similar to methods for linear systems.

We conclude with a theorem that allows us to decide easily if a subset of a ring is an ideal.

THEOREM 7.22 (The Ideal Theorem). *Let R be a ring and $A\subset R$. The following are equivalent:*

(A) A is an ideal subring of R.

(B) A is closed under subtraction and absorption. That is,

(I1) for all $a,b\in A$, $a-b\in A$; and

(I2) for all $a\in A$, $r\in R$, $ar\in A$.

PROOF. You do it! See Exercise . □

EXERCISES.

EXERCISE 7.23. Show that for any $n\in\mathbb{N}^+$, $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$.

EXERCISE 7.24. Prove Theorem 7.22 (the Ideal Theorem).

EXERCISE 7.25. Let R be a ring and A and I two ideals of R. Decide whether the following subsets of R are also ideals, and explain your reasoning:
(a) $A \cap I$
(b) $A \cup I$
(c) $A + I = \{x + y : x \in A, y \in I\}$
(d) $AI = \{xy : x \in A, y \in I\}$

7.3. PRIME AND MAXIMAL IDEALS

7.4. RING HOMOMORPHISMS

7.5. FIELDS AND DOMAINS

7.6. FINITE FIELDS

CHAPTER 8

Rings and polynomial factorization

CHAPTER 9

Ideals, Varieties, and Gröbner bases

Bibliography

[AF05] Marlow Anderson and Todd Feil. *A First Course in Abstract Algebra*. Chapman and Hall/CRC, second edition, 2005.

[Bah08] Tavmjong Bah. *Inkscape: Guide to a Vector Drawing Program*. Prentice-Hall, second edition, 2008. Software available online at `www.inkscape.org`.

[Bri] Rogério Brito. The `algorithms` bundle. Retrieved 16 October 2008 from `http://www.ctan.org/tex-archive/macros/latex/contrib/algorithms/`. Version 0.1.

[CLO97] David Cox, John Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms*. Springer-Verlag New York, Inc., New York, second edition, 1997.

[Grä04] George Grätzer. *Math into LaTeX*. Birkhäuser, Boston, third edition, 2004.

[HA88] Abraham P. Hillman and Gerald L. Alexanderson. *A First Undergraduate Course in Absract Algebra*. Wadsworth Publishing Company, Belmont, California, 1988.

[Lam86] Leslie Lamport. *LaTeX: a Document Preparation System*. Addison-Wesley Publishing Company, 1986.

[Lau03] Niels Lauritzen. *Concrete Abstract Algebra: from Numbers to Gröbner Bases*. Cambridge University Press, Cambridge, 2003.

[LP98] Rudolf Lidl and Günter Pilz. *Applied Abstract Algebra*. Springer-Verlag, New York, second edition edition, 1998.

[Lyx09] Lyx Team. *Lyx*, 2008–2009. Retrieved too many times to count from `http://www.lyx.org`.

[Pic06] Paul Pichaureau. *The mathdesign package*, 2006. Retrieved from `http://www.ctan.org/tex-archive/help/Catalogue/entries/mathdesign.html`.

[RO08] Sebastian Rahtz and Heiko Oberdiek. *Hypertext marks in LaTeX: a manual for* `hyperref`, 2008. Retrieved 21 April 2009 from `http://www.tug.org/applications/hyperref/manual.html`.

[Rot06] Joseph J. Rotman. *A First Course in Abtract Algebra with Applications*. Pearson Education, Inc., New Jersey, third edition, 2006.

[Soc02] Americal Mathematical Society. *User's Guide for the* `amsmath` *Package*, version 2.0 edition, 2002. Retrieved 21 April 2009 from `http://www.ams.org/tex/amslatex.html`.

[Ste08] William Stein. *Sage: Open Source Mathematical Software (Version 3.1.1)*. The Sage Group, 2008. `http://www.sagemath.org`.

Index