Lecture Notes on Abstract Algebra

S.R. Doty

21 November 2024

Doty, Stephen Richard
Department of Mathematics and Statistics
Loyola University Chicago
Chicago, Illinois 60660 U.S.A.
https://doty.math.luc.edu

Contents

Preface

These are my lecture notes for a first course in abstract algebra, which I have taught a number of times over the years. Typically, the course attracts students of varying background and ability. The notes assume some familiarity with linear algebra, in that matrices are used frequently.

The main focus of the course is on group theory, with the goal of getting to the Sylow theorems and the classification of finite abelian groups. The very beginnings of ring theory are also treated here, with a focus on commutative rings, in order to discuss the finite rings and fields inherent in modular arithmetic. Students have little trouble understanding the ring axioms, probably due to prior exposure to the number systems of basic mathematics.

The organization of the material is somewhat novel, in that the main classes of examples are introduced and studied first, before the abstract group axioms are given. This seems to offer some advantages over the standard approach of clobbering unsuspecting students with the abstract group axioms before they have had any experience with examples. In my experience, even very capable students find the group axioms quite difficult at first.

Modulo preliminaries, the course starts with permutation groups, which are defined as nonempty sets of permutations closed under products and inverses; this of course includes the symmetric and alternating groups. Next the dihedral groups are introduced as symmetry groups of regular polygons and the groups of rotational symmetries of the platonic solids are also discussed briefly without proof. Next comes modular arithmetic, including axioms for commutative rings and fields, and a proof that the ring of integers modulo n is a field if and only if n is prime. The last class of concrete examples are linear groups, defined as nonempty sets of matrices closed under products and inverses. A fairly detailed analysis of the rotation group $\mathrm{SO}(2)$ is given, along with the full orthogonal group $\mathrm{O}(2)$ of orthogonal 2×2

matrices.

Only then are the axioms for abstract groups introduced. By this point students have seen enough examples of groups to be able to appreciate the utulity value of the axiomatic approach. The rest of the course proceeds as usual, covering all the standard main topics, including subgroups, cyclic groups, quotients, homomorphisms, products, group actions, Sylow theorems, and finite abelian groups. Brief discussions of simple groups, composition series, and generators and relations are included.

I wish to thank all the students over the years who used various incarnations of these notes; their feedback has been incorporated into the notes in many ways.

Chapter 0

Preliminaries

1 Logic

We start by discussing some basic terminology from mathematical logic. These definitions and notation are used throughout mathematics.

1.1 Definition. In mathematics, a *statement* is an assertion which is either true or false.

1.2 Definition. A *conditional* statement is any statement of the form "if P then Q" where P and Q are statements. Conditional statements are also called *implications*. The implication "if P then Q" is also commonly written as "P implies Q" or "$P \implies Q$."

To prove an implication $P \implies Q$, we assume P is given (the hypothesis) and show by logical deduction that one can derive Q (the conclusion) from the given hypothesis. Such an approach is called a direct proof of the implication.

1.3 Definition. A statement of the form "P if and only if Q" is called a *biconditional* or *equivalence*. It is commonly written as "P iff[1] Q" or "$P \iff Q$." By definition, $P \iff Q$ means that both $P \implies Q$ and $Q \implies P$.

Thus, to show that "P if and only if Q" is true you must prove two conditionals: that $P \implies Q$ and that $Q \Rightarrow P$. The implication $Q \implies P$ is called the *converse* of $P \implies Q$. So proving the equivalence $P \iff Q$

[1] Thus, *iff* is an abbreviation for the phrase "if and only if."

amounts to proving both the implication $P \implies Q$ and its converse $Q \implies P$.

Be careful about converses. It is a fallacy to assume that if $P \implies Q$ then also $Q \implies P$. This is often false. For instance, the implication "all dogs are mammals" is true, but its converse "all mammals are dogs" is plainly false.

1.4 Remark. By standard convention, *all definitions in mathematics are considered to be biconditionals*, even if not stated as such. Thus, in a definition, the word *if* should always be interpreted as *if and only if*.

1.5 Definition. The *contrapositive* of $P \implies Q$ is $(\neg Q) \implies (\neg P)$. Here, the symbol $\neg$ means "not." I.e., $\neg P$ means "not P."

It is a well known that *every implication is logically equivalent to its contrapositive*, and mathematicians routinely use that fact without comment.

1.6 Example. To prove the implication: (n^2 is odd) $\implies$ (n is odd), it suffices to show the contrapositive statement: (n is even) $\implies$ (n^2 is even). The contrapositive is easy to see by a direct proof, as follows. If n is an even integer then $n = 2k$ for some integer k, and hence $n^2 = 4k^2 = 2(2k^2) = 2m$ is even, because $m = 2k^2$ is an integer.

1.7 Definition (quantifiers)**.** The symbol $\forall$ is the *universal quantifier*. It means "for all." The symbol $\exists$ is the *existential quantifier*. It means "there exists."

A *universal statement* is one which is universally quantified. For example, anything of the form $\forall x, P(x)$. By definition, this is true whenever $P(x)$ is true for all possible values of x (in some domain).

An *existential statement* is one which is existentially quantified. For example, anything of the form $\exists x, P(x)$. By definition, this is true whenever $P(x)$ is true for at least one value of x.

Exercises

1.1. Prove that if n^2 is odd then n must be odd.
1.2. Prove that if n^3 is odd then n must be odd.
1.3. Prove that n is even if and only if n^2 is even.
1.4. Prove that n is odd if and only if n^2 is odd.
1.5. Prove that for all real numbers x, $x^2 \geq 0$.
1.6. Prove that there exists a complex number whose square is -2.

2 Sets

Next we discuss the basic notions of set theory, which are used throughout mathematics. We take the naive approach, in which the notion of a set is left somewhat imprecise.

2.1 Definition. The naive concept is that a set is a collection of *elements*, which is determined by its elements, in the following sense: two sets are considered to be *equal* if[2] they have precisely the same elements. Order of the elements doesn't matter, and in a set duplicates are not allowed.

In computer science the concept of a *list* is quite important, and one might get the impression that lists are the same thing as sets. This is erroneous, however, since a list can have repetitions and order matters, while a set cannot have repetitions and order doesn't matter.

2.2 Remark. There are inherent difficulties with this naive concept of set; see the discussion of Russell's paradox below. Rather than allowing a set to be any collection of elements, in order to avoid paradoxes we should only allow collections which are not "too big" in a certain sense. Fortunately, most of the sets we deal with in basic mathematics are not too big, so in practice we don't worry very much about this issue.

2.3 Definition. If A is a set and a is one of its elements then we write $a \in A$ (i.e., $\in$ means "is in"). If b is not an element in the set A then we write $b \notin A$ (i.e., $\notin$ means "is not in").

2.4 Definition. A set is called *finite* if it has finitely many elements. The number of distinct elements is called the *cardinality* of the set. Usually we write $|A|$ for the cardinality of a set A.

A set is *infinite* if it is not finite, and in that case we can write $|A| = \infty$.[3]

2.5 Definition. The *empty set* or *null set* is the set $\varnothing$ with no elements. Of course the cardinality of the empty set is zero: $|\varnothing| = 0$. The empty set may also be written as $\{\ \}$.

Often we write a set by listing its elements. Thus $A = \{2, 5, 1, 9\}$ is the set consisting of the elements $1, 2, 5, 9$. We could also correctly write

[2] We follow the convention of 1.4 here. Since we are making a definition, the word "if" means "if and only if" in this context.

[3] There is a whole theory of infinite cardinals (due to Georg Cantor), but for our purposes it is usually not necessary to distinguish between different orders of infinity.

$A = \{1, 2, 5, 9\}$ since only the elements, and not their order, is important. So for this set A it is correct to write $2 \in A$, $3 \notin A$.

For infinite sets we sometimes use the ... notation, to indicate that the displayed pattern continues as indicated. For example, the set written $\{1, 3, 5, 7, 9, \dots\}$ stands for the set of all odd natural numbers and the set $\{0, \pm 2, \pm 4, \pm 6, \dots\}$ is the set of all even integers.

2.6 Definition (set builder notation)**.** Often we define sets by listing some *condition* for membership in the set. This is sometimes called *set builder notation.* If $P(x)$ is some condition on x then the set

$$\{x \mid P(x)\} = \{x : P(x)\}$$

should be read as "the set of all x such that $P(x)$ is true." In particular, the equivalent symbols | and : usually mean "such that" when they appear inside sets.

2.7 Definition (some standard sets of numbers)**.** The following notations have become the standard for various common sets of numbers used in much of mathematics:

$$\begin{aligned}
&\mathbb{N} = \{0, 1, 2, 3, 4, \dots\} \quad \text{natural numbers} \\
&\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\} \quad \text{integers} \\
&\mathbb{Q} = \{\tfrac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0\} \quad \text{rational numbers} \\
&\mathbb{R} = \text{real numbers} \\
&\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\} \quad \text{complex numbers}
\end{aligned}$$

where it is understood that $i^2 = -1$. (The number i is called the *imaginary unit.*)

2.8 Examples. (a) $\{x \in \mathbb{R} \mid 1 < x \leq 3\}$ defines the interval $(1, 3]$.

(b) $\{2k + 1 \mid k \in \mathbb{N}\}$ is the set of all odd natural numbers.

(c) $\{x \in \mathbb{R} : x^2 - 4 = 0\}$ is the set of all real numbers x satisfying the equality $x^2 - 4 = 0$. As we know, this is just the set $\{2, -2\}$.

Russell's paradox. Bertrand Russell pointed out the problem in trying to define the following set:

$$X = \{A \mid A \notin A\}.$$

The set X is the set consisting of all sets which are not elements of themselves. There are lots of such sets, so X is really big. Then Russell asked the question: Is $X \in X$?

It turns out that this question has no answer, since it contradicts itself! If we assume $X \in X$, then $X \notin X$ and we have a contradiction. On the other hand, if we assume $X \notin X$ then $X \in X$ and we again have a contradiction. Since either $X \in X$ or $X \notin X$ we cannot avoid a contradiction if X exists as a set.

Because of this paradox, and others like it, it has been found necessary to exclude certain "large" sets from the realm of set theory. It turns out that to give a precise definition of the notion of a set, which avoids such paradoxes and contradictions, is indeed a very difficult problem. This problem was solved by Russell and Whitehead in their tome *Principia Mathematica*. To find out more on such matters, see a decent modern text on set theory.

The sets that we use in ordinary mathematical discourse are almost always small enough to be free of such difficulties, so in practice we usually don't need to worry very much about this problem. In general, so long as a given set can be realized as a subset of some existing set, things are okay.

Exercises

2.1. Is it true that $\{0, 10, 0, 1, 10, 8, 1, 10\} = \{0, 1, 8, 10\}$? Explain.

2.2. Is it true that $\{1, 2, 3, 4\} = \{4, 3, 2, 1\}$? Explain.

2.3. Is it true that $\{\{1, 2, 3\}, \{3, 2, 1\}, \{1\}, \{2\}, \{3\}\} = \{1, 2, 3\}$? Explain and justify.

2.4. Find the cardinality of the following sets:
   (a) $\{0, 10, 0, 1, 10, 8, 1, 10\}$.
   (b) $\{1, 2, 3, 4\}$.
   (c) $\{\{1, 2, 3\}, \{3, 2, 1\}, \{1\}, \{2\}, \{3\}\}$.
   (d) $\mathbb{Z}$.

2.5. (a) Is $1 \in \{\{1, 2, 3\}, \{3, 2, 1\}, \{1\}, \{2\}, \{3\}\}$? Explain.
   (b) Is $\{1\} \in \{\{1, 2, 3\}, \{3, 2, 1\}, \{1\}, \{2\}, \{3\}\}$? Explain.

2.6. (a) Is $1 \in \{1\}$? Explain.
   (b) Is $1 \in \{\{1\}\}$? Explain.
   (c) Is $\emptyset \in \{1\}$? Explain.

2.7. Let $A = \{1, 2, 3, 4, 5, 6\}$. Compute $B = \{4n - 1 \mid n \in A\}$. What is $|A|$ and $|B|$?

2.8. Is $\{3n \mid n \in \mathbb{Z}\} = \{0, \pm 3, \pm 6, \pm 9, \dots\}$? Explain.

3 Set operations

We now consider the basic relations among sets as well as some fundamental operations on sets.

3.1 Definition (subset)**.** We say that A *is a subset of* B (written $A \subset B$ or equivalently $A \subseteq B$) if $x \in A \implies x \in B$.

We also sometimes say "A *is contained in* B" as a synonym for "A is a subset of B."

Note that the statements $A \subset B$ and $A \in B$ do *not* have the same meaning. Note also that by definition $A \subset A$: every set is a subset of itself. Also, $\varnothing \subset A$; i.e., the empty set is a subset of every set.

3.2 Remark. By definition, $B \supset A \iff A \subset B$. Thus, the symbol $\supset$ means "contains."

3.3 Definition (proper inclusions)**.** If $A \subset B$ but $A \neq B$ then we will write $A \subsetneqq B$. In this case we say that A is a *proper* subset of B, or that A is *strictly* contained within B.

3.4 Theorem (set equality)**.** *Let A, B be sets. Then $A = B$ if and only if both $A \subset B$ and $B \subset A$.*

This theorem is often used in proofs to show equality of two sets. In other words, to prove that $A = B$, you have to prove two things: that $A \subset B$ and $B \subset A$.

We allow sets whose elements are themselves sets. Let A be a set. We distinguish between the set A and $\{A\}$, which is the set with one element, A. The latter object is the set consisting of the set A, and that is different from A itself.

Thus if A is a set, we can form the set $\{A, \{A\}\}$, the set whose elements are A and $\{A\}$. As another example along these lines, consider the set $X = \{1, \{1\}, \{1, 2\}\}$. Then X has 3 elements. It is true that $1 \in X$, but the statement $2 \in X$ is false: 2 is part of the third element, but it is not an element. These structural distinctions may seem pedantic but they are quite important.

3.5 Definition (union of sets)**.** The *union* or *join* of a collection of sets is the set whose elements are obtained by joining together all the elements in the collection. The union of two sets A, B is written as $A \cup B$. Formally, we can write

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

using the set-builder notation. More generally, if $A_1, A_2, \ldots, A_n$ are sets then

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n = \{x \mid x \in A_i, \text{ for some } i\}.$$

Even more generally, if we have a family A_i of sets, where i varies over all the elements of some given set I, then we can write

$$\bigcup_{i \in I} A_i = \{x \mid x \in A_i, \text{ for some } i \in I\}.$$

In this context, the set I is called an *indexing* set.

3.6 Definition (intersection of sets)**.** The *intersection* or *meet* of a collection of sets is the set of elements common to all sets in the collection. The intersection of two sets is written as $A \cap B$. Formally,

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

More generally, if $A_1, A_2, \ldots, A_n$ are sets then

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n = \{x \mid x \in A_i, \text{ for all } i = 1, \ldots, n\}.$$

More generally still, if we have a family A_i of sets, where i varies over all the elements of some given indexing set I, then we write

$$\bigcap_{i \in I} A_i = \{x \mid x \in A_i, \text{ for all } i \in I\}.$$

Two sets A, B are said to be *disjoint* if their intersection is the empty set.

3.7 Definition (complements of sets)**.** If A and B are given sets the *complement* of B in A is the set

$$A - B = A \setminus B = \{x \in A \mid x \notin B\}.$$

In words, it is the set of all elements of A which are not elements of B.

3.8 Definition (products of sets)**.** Let A, B be sets. Then their *Cartesian product* is the set

$$A \times B = \{(x, y) : x \in A \text{ and } y \in B\}.$$

Elements of $A \times B$ are called *ordered pairs* since order is important: (x, y) is usually different from (y, x).

In case the two sets are the same set, we often write A^2 in place of $A \times A$.

This construction should look familiar. For instance, the set $\mathbb{R}^2$ is the set of points in the standard Euclidean plane.

3.9 Definition (products of sets)**.** More generally, if $A_1, A_2, \ldots, A_n$ are sets then we can form their cross product, the elements of which are called ordered n-tuples. Formally,

$$A_1 \times \cdots \times A_n = \{(x_1, x_2, \ldots, x_n) \mid x_i \in A_i \text{ for all } i = 1, \ldots, n\}.$$

Note that in an ordered n-tuple the *order is important*. Elements of cross products are *not* sets, they are ordered tuples.

The special case $A \times A \times \cdots \times A$ in which all the A_i are the same set A is often written A^n, where n is the number of sets in the cross product. Thus, for example, we have $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$ (this is Euclidean 3-space).

Exercises

3.1. Let $A = \{1, 3, 5, 7\}$, $B = \{2, 4, 6, 8\}$, $C = \{1, 2, 3, 4, 3, 2, 1\}$. Find:
   (a) $A \cap C$ and $A \cup C$.
   (b) $A - C$ and $C - A$.
   (c) $A - B$ and $B - A$.
   (d) $A \cap B \cap C$ and $A \cup B \cup C$.
   (e) $(A \cup B) - C$ and $C - (A \cup B)$.

3.2. (a) Is it true that $\{1, 2, 3\} \subset \{\{1, 2, 3\}, \{3, 2, 1\}, \{1\}, \{2\}, \{3\}\}$? Explain.
   (b) Is it true that $\{\{1\}\} \subset \{\{1, 2, 3\}, \{3, 2, 1\}, \{1\}, \{2\}, \{3\}\}$? Explain.
   (c) Is it true that $\{\{1\}, \{1, 3, 2\}\} \subset \{\{1, 2, 3\}, \{3, 2, 1\}, \{1\}, \{2\}, \{3\}\}$? Explain.

3.3. Explain why $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

3.4. (a) Show that $\{4n + 1 \mid n \in \mathbb{Z}\} = \{4m - 3 \mid m \in \mathbb{Z}\}$.
   (b) Show that $\{4n + 1 \mid n \in \mathbb{Z}\} = \{4m + 9 \mid m \in \mathbb{Z}\}$.

3.5. Show that $\{2n \mid n \in \mathbb{Z}\} \cap \{3n + 7 \mid n \in \mathbb{Z}\} = \{6k + 4 \mid k \in \mathbb{Z}\}$.

3.6. Suppose that A, B are subsets of some bigger set X. Write A^c for the complement of A in X (so $A^c = X - A$). Prove that $B - A = B \cap A^c$.

3.7. Suppose that A, B, C are subsets of some bigger set X. Write A^c for the complement of A in X (so $A^c = X - A$). Prove that:
   (a) $(A \cap B)^c = A^c \cup B^c$.
   (b) $(A \cup B)^c = A^c \cap B^c$.
   (c) $(A \cap B \cap C)^c = A^c \cup B^c \cup C^c$.

(d) $(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$.

3.8. (a) Show by example that it is not always true that $A \times B = B \times A$ for sets A, B.

(b) Is it ever true that $A \times B = B \times A$? Justify your answer with proof.

4 Relations

Relations appear all over mathematics, and are also important in computer science. For instance, *relational databases* are based on the mathematical idea of relations.

4.1 Definition. Let A be a given set. A *relation* on A is a subset R of $A \times A$. Often when R is a relation on A we write xRy instead of $(x, y) \in R$ in order to suggest the idea that x is related to y (by the relation R).

Thus a relation is nothing but a set of ordered pairs. But that is not usually how we think about it. Usually we prefer to visualize the idea behind the ordered pairs instead of the set of ordered pairs.

4.2 Example. On the usual set $\mathbb{R}$ of real numbers, we have the usual inequality relations: $<$, $\leq$, $>$, $\geq$. We could define $<$ as $\{(x, y) \mid y - x \text{ is positive}\}$, and so on.

One can also widen the definition of relation to cover two sets. Thus, a subset R of $A \times B$ could also be called a relation from A to B by some authors.

4.3 Definition. Let R be a relation on a set A. We say that R is an *equivalence relation* if the relation R satisfies

reflexivity: xRx,
symmetry: xRy implies yRx, for all $x, y \in A$
transitivity: xRy and yRz implies xRz, for all $x, y, z \in A$.

The *equivalence class* $[x]$ of $x \in A$ is defined by $[x] = \{y \in A \mid xRy\}$.

If R is an equivalence relation, one can show that $[x] = [y] \iff xRy$. Thus, two equivalence classes either coincide or are disjoint.

4.4 Definition. A *partition* of a set A is a collection of subsets $\{P_i\}_{i \in I}$ of A such that:

(a) the subsets are pairwise disjoint: $i \neq j$ implies $P_i \cap P_j = \varnothing$.
(b) the union is all of A: $\bigcup_{i \in I} P_i = A$.

4.5 Theorem (fundamental theorem of equivalence relations)**.** *Any equivalence relation on a set A induces a partition of A into equivalence classes. Conversely, any given partition of A determines an equivalence relation for which the given partition is the induced one.*

4.6 Example. Consider the set P of all living people on earth. Define a relation R on the set P by declaring that xRy if and only if x, y have the same age (in years). It is easy to check that R is an equivalence relation of the set P. The equivalence classes are just a way of grouping people by their age: all 4-year olds would be one such equivalence class. The 18-year olds form another class. Obviously different classes are disjoint, and the union of all the classes is P.

Such "classifications" are what equivalence relations describe.

Exercises

4.1. Let $<$ be the usual "less than" on the set $\mathbb{R}$ of real numbers. That is, for real numbers x, y we write $x < y$ if and only if x is less than y.

(a) Is $<$ reflexive? Explain.

(b) Is $<$ symmetric? Explain.

(c) Is $<$ transitive? Explain.

(d) Is $<$ an equivalence relation on $\mathbb{R}$? Explain.

4.2. Same as the previous questions, except for $\leq$ in place of $<$.

4.3. Define a relation R on the set $\mathbb{Z}$ by declaring that aRb if and only if $a - b$ is even.

(a) Prove that R is an equivalence relation on the set $\mathbb{Z}$.

(b) Describe the equivalence classes. How many equivalence classes are there?

4.4. Define a relation R on the set $\mathbb{Z} \times (\mathbb{Z} - \{0\})$ by declaring that $(a, b)R(c, d)$ if and only if $ad = bc$.

(a) Prove that R is an equivalence relation on the set $\mathbb{Z} \times (\mathbb{Z} - \{0\})$.

(b) Describe the equivalence classes. [Hint: Think about representing rational numbers by fractions.]

4.5. Let S be the set of all infinite sequences $(a_n) = (a_n)_{n=1}^{\infty}$ of real numbers. Define a relation $\sim$ on S by declaring that $(a_n) \sim (b_n)$ if and only if there exists some N such that $a_n = b_n$ for all $n \geq N$. Show that $\sim$ is an equivalence relation on the set S.

5 Functions

Functions are special types of relations in which images are unique. Functions between sets are used throughout mathematics, so it is important to know the precise definitions collected here.

5.1 Definition. Let A, B be two given sets. A *function* f from A to B is a rule which assigns to each element $x \in A$ a unique element $f(x) \in B$. The element $y = f(x)$ is called the *image* of x.

If f is a function from a set A to a set B then A is called the *domain* of the function and B is called the *co-domain*. People write $f : A \to B$ to indicate that f is a function mapping A to B.

The word *mapping* is a synonym for function. Sometimes it is shortened to *map*.

The uniqueness part of the above definition is crucial. If $f(x)$ is not necessarily uniquely determined by the rule then we say that f is not *well-defined*. For example, consider the rule f which defines $y = f(x)$ to be the solution y to the equation $y^2 = x$, for every real number x. This is not well-defined as a function from $\mathbb{R}$ into $\mathbb{R}$, because when $x > 0$ the equation $y^2 = x$ always has two solutions.

I emphasize that the definition of function just given says *nothing at all* about equations or formulas. While we can, and very often do, define functions in terms of some formula, formulas are NOT the same thing as functions. The concept of function is much more general.

For instance, the equation $y = f(x) = x^2 - 1$ defines a function from $\mathbb{R}$ to $\mathbb{R}$. This function is given by a formula. However, consider the function D such that $D(t)$ is the temperature at time t at a certain chosen location in Chicago. Can you write down an explicit formula for this function? How about the function DOW(t) which gives the closing value of the Dow–Jones industrial average, day-by-day?

5.2 Definition (image and preimage)**.** Suppose $f : A \to B$.

(a) *Image of a subset of A.* If $S \subset A$ then

$$f(S) = \{f(x) \mid x \in S\};$$

is called the *image* of S (under the mapping f). By definition, $f(S) \subset B$.

(b) *Preimage of a subset of B.* Given $T \subset B$, we have

$$f^{-1}(T) = \{x \in A \mid f(x) \in T\};$$

this set is called the *preimage* or *inverse-image* of T.

The set $f(A)$, which is the set of all outputs of the function f, is called the *image* or *range* of f, sometimes denoted by $\operatorname{im} f$ or $\operatorname{Im} f$.

5.3 Definition (surjection)**.** Let $f : A \to B$ be a mapping from A to B. We say that f is *surjective* if the image equals the co-domain; in other words if $f(A) = B$. We also say that f maps A *onto* B, or that f is *onto*, in that case.

For example, the rule $f(x) = x^2$ defines a mapping from $\mathbb{R}$ to $\mathbb{R}$ which is *not* surjective since f maps $\mathbb{R}$ *into* $\mathbb{R}$ but not *onto* $\mathbb{R}$, since obviously you can't get any negative real numbers by squaring real numbers.

However, the rule $f(x) = 7x - 23$ defines a surjective mapping $\mathbb{R} \to \mathbb{R}$, since every real number y is obtainable as the image of some real x.

If the mapping f is surjective then we also say that f is a *surjection*.

5.4 Definition (injection)**.** We say that $f : A \to B$ is *injective* if the preimage of every point in the image consists of a single point in the domain. To say it another way: f is injective if $x_1 \neq x_2$ implies that $f(x_1) \neq f(x_2)$. Injective functions are also called *one-to-one*.

For example, the rule $f(x) = x^2$ defines a mapping from $\mathbb{R}$ to $\mathbb{R}$ which is NOT injective since it is a two-to-one mapping: every y except 0 has *two* elements in its preimage.

Injective mappings are also called *injections*.

5.5 Definition (bijection)**.** Let $f : A \to B$ be a mapping. We say that f is *bijective* if it is both surjective and injective. Bijections always set up a one-to-one correspondence between the domain and co-domain.

5.6 Definition (composition of functions)**.** Let $f : A \to B$ and $g : B \to C$ be functions. Then we can define a new function $h : A \to C$ by the rule: $h(x) = g(f(x))$. The function h so defined is called the *composite* of g and f, and we write $h = g \circ f$. Sometimes, by abuse of notation, we will simply write $h = gf$ for the composite function.

Note the functional composition is not commutative: $f \circ g \neq g \circ f$. It is associative, however: $h \circ (g \circ f) = (h \circ g) \circ f$ for any functions f, g, h such that the various composites are defined.

5.7 Definition (identity function)**.** Let A be a given set. The *identity* function on A is the function id such that $id(x) = x$ for all $x \in A$. If we must specify the underlying set A then we write id_A or sometimes 1_A.

5.8 Definition (invertible functions)**.** Let $f : A \to B$ be a function mapping A to B. We say that f is *invertible* if there exists another function $g : B \to A$ such that $f \circ g = id_B$ and $g \circ f = id_A$. When this holds, the function g is called the *inverse* of the function f, and is written as f^{-1}.

Note that $f \circ g = id_B$ if and only if $f(g(y)) = y$ for all $y \in B$. Similarly, $g \circ f = id_A$ if and only if $g(f(x)) = x$ for all $x \in A$.

5.9 Example. The function $f : \mathbb{R} \to \mathbb{R}$ defined by the rule $f(x) = 2x - 7$ is invertible. What is its inverse?

5.10 Theorem (fundamental theorem of invertible functions)**.** *A function is invertible if and only if it is bijective.*

5.11 Definition. If $f : A \to B$ is a function with image I then we can always regard f as a mapping from A into T where T is any set such that $I \subset T \subset B$. This is called *restricting the co-domain.* We can always shrink the co-domain to any such T.

Strictly speaking, we get a new function when we do this, but people often use the same symbol f for this new function, by abuse of notation.

Note that every function always gives us (by restriction) a surjection onto its image. More precisely, if $f : A \to B$ is a function from A to B and if $I = f(A)$ is its image, then the restriction $f : A \to I$ is a surjection.

5.12 Definition. Suppose that $f : A \to B$ is a function. Given any subset $S \subset A$ we can define a new mapping $f_{|S} : S \to B$ by the same rule as for f: $f_{|S}(x) = f(x)$ for all $x \in S$. This is called *restricting the domain.*

We usually need to use a different notation for such a function, in order to avoid confusion.

A familiar example of the use of domain restriction in basic calculus is when you restrict the domain of the sine or cosine function in order to make them invertible. Without restriction of the domain, the inverse sine and cosine functions would not exist.

Exercises

5.1. Let $f : \mathbb{N} \to \mathbb{N}$ be given by the rule $f(x) = x^2$. Compute both the image and preimage of the set $\{1, 2, 3, 4\}$.

5.2. Show that a mapping $f : A \to B$ is injective if and only if $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$.

5.3. (a) Show that a mapping $f : A \to B$ is injective if and only if there exists a mapping $g : B \to A$ such that $g \circ f = id_A$.

(b) Show that a mapping $f : A \to B$ is surjective if and only if there exists a mapping $g : B \to A$ such that $f \circ g = id_B$.

5.4. Give an example of mappings $f : \mathbb{Z} \to \mathbb{Z}$, $g : \mathbb{Z} \to \mathbb{Z}$ such that $g \circ f = id_{\mathbb{Z}}$ but f is not invertible.

5.5. Let $f : A \to B$ be a function and let S, T be subsets of A. Show that $f(S \cup T) = f(S) \cup f(T)$ and that $f(S \cap T) \subset f(S) \cap f(T)$. Give an example to show that $f(S \cap T)$ need not coincide with $f(S) \cap f(T)$.

5.6. Let $f : A \to B$ be a function and let U, V be subsets of B. Show that $f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V)$ and that $f^{-1}(U \cap V) = f^{-1}(U) \cap f^{-1}(V)$.

Chapter 1

Permutation Groups

6 Permutations

Lagrange and Galois studied permutations among the roots of polynomials as a way of understanding solutions of polynomial equations. This eventually led to what is now called *group theory*.

6.1 Definition. A *permutation* of a set X is a bijection $X \to X$. Write $\mathbb{S}_X$ for the set of all permutations of X; i.e.,

$$\mathbb{S}_X = \{\sigma : X \to X \mid \sigma \text{ is a bijection}\}.$$

In the special case $X = \underline{\mathbf{n}} := \{1, 2, \ldots, n\}$ we write $\mathbb{S}_n$ instead of $\mathbb{S}_X$.

So $\mathbb{S}_n$ is the set of all permutations of the numbers from 1 to n. We can think of it as the set of all permutations of any n things, since we can always assign numbers from 1 to n to those things.

If X is a finite set of n elements, then the number of permutations of X is $n!$, the factorial of n. So the cardinality $|\mathbb{S}_n| = n!$. If X is an infinite set then $\mathbb{S}_X$ is also infinite.

6.2 Definition. If $\sigma \in \mathbb{S}_n$, then we can depict the permutation by writing

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

This is called the *two-line notation* for a permutation.

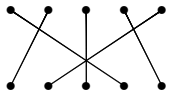
Read down the columns to see images of elements from the top row sitting in the corresponding position in the bottom row. In other words, if α maps i to j then we put i over j in the ith column of the two-line notation.

6.3 Example. $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 3 & 5 & 2 \end{smallmatrix}\right)$ is the permutation which maps $1 \to 4$, $2 \to 1$, $3 \to 3$, $4 \to 5$, and $5 \to 2$. Note that 3 is a *fixed point* for α.

6.4 Definition (permutation diagrams)**.** Permutations can also be expressed by diagrams. A diagram is a directed graph with $2n$ vertices and n edges, with the vertices arranged in two rows of n each, such that each edge connects a single vertex in the top row to a single vertex in the bottom row. The vertices on the top and bottom rows are numbered 1 to n in order from left to right. Then the diagram depicts the permutation α that maps i to j if and only if there is an edge connecting vertex i in the top row with vertex j in the bottom row.

6.5 Example. We can express the permutation α in the previous example by the diagram

$$\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 3 & 5 & 2 \end{smallmatrix}\right) = \quad .$$



Technically, the edges should all have arrows pointing down, but since all the arrows point in the same direction, it is customary to omit them. You should be able to *read* the diagram as a bijection by reading the edges from top to bottom as defining images of each numbered vertex.

6.6 Definition (multiplication of permutations)**.** Given two permutations, say $\alpha, \beta \in \mathbb{S}_n$, we can get a new permutation $\alpha \circ \beta \in \mathbb{S}_n$ by the usual composition of functions. *Convention:* We usually simplify notation and write the composite $\alpha \circ \beta$ as the "product" $\alpha\beta$. *With this convention, we have to read such products as composites.*

Thus, $\alpha \circ \beta$ is the bijection defined by the rule $(\alpha \circ \beta)(j) = \alpha(\beta(j))$ for all $j \in \underline{\mathbf{n}}$. In terms of the above convention, this reads as $(\alpha\beta)(j) = \alpha(\beta(j))$.

We know that $\alpha\beta = \alpha \circ \beta$ is another permutation because the composite of two bijections is always a bijection.

Let me remind you that composition of functions is not always commutative. That is, if f, g are functions then it can happen that $f \circ g \neq g \circ f$. Since permutations are functions, this also applies to permutations. So for permutations $\alpha, \beta \in \mathbb{S}_n$ it can happen that $\alpha\beta \neq \beta\alpha$.

6.7 Example. If $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 1 & 5 & 3 \end{smallmatrix}\right)$ and $\beta = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 4 & 5 & 2 \end{smallmatrix}\right)$ are defined in terms of the two-line notation then we have

$$\alpha\beta = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 1 & 5 & 3 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 4 & 5 & 2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 5 & 3 & 4 \end{smallmatrix}\right).$$

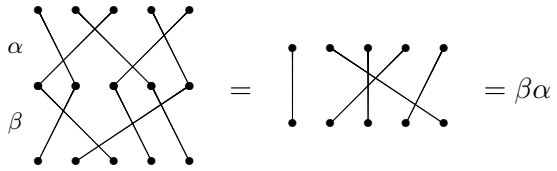
On the other hand, if we do the product the other way around we obtain

$$\beta\alpha = \left(\begin{smallmatrix}1&2&3&4&5\\3&1&4&5&2\end{smallmatrix}\right)\left(\begin{smallmatrix}1&2&3&4&5\\2&4&1&5&3\end{smallmatrix}\right) = \left(\begin{smallmatrix}1&2&3&4&5\\1&5&3&2&4\end{smallmatrix}\right).$$

You should check these results yourself to make sure that you understand the example. Notice that $\alpha\beta \neq \beta\alpha$ in this example.

In particular, you need to read $\alpha\beta = \alpha \circ \beta$ as α of β, which is the same as β followed by α. In a composition $\alpha \circ \beta$, which is defined by $(\alpha \circ \beta)(j) = \alpha(\beta(j))$, the *second* function is the *first* to be applied to the argument.[1]

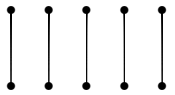
6.8 Example. Multiplication of permutations can also be calculated using permutation diagrams. For instance,



computes $\beta\alpha$ in terms of the diagram, by reading the edges all the way from top to bottom in the joined diagram displayed on the left above. Again, because we are computing $\beta\alpha = \beta \circ \alpha$, we are computing α followed by β, so we must put the diagram of α above the diagram of β.

6.9 Definition (identity permutation)**.** The simplest permutation in $\mathbb{S}_n$ is the *identity* permutation, which is just the identity mapping id_X on the set $X = \underline{\mathbf{n}}$. We will often write id for this permutation.

6.10 Example. For instance, if $n = 5$ we have $id = \left(\begin{smallmatrix}1&2&3&4&5\\1&2&3&4&5\end{smallmatrix}\right)$. The diagram of this permutation is

$$id = \quad$$  .

6.11 Definition (inverse permutation)**.** Since a permutation α is a bijection, it is always an invertible mapping. Thus α^{-1} exists. It is defined by the property $\alpha^{-1}\alpha = id = \alpha\alpha^{-1}$.

6.12 Example. If $\alpha = \left(\begin{smallmatrix}1&2&3&4&5\\4&1&3&5&2\end{smallmatrix}\right)$ in the two-line notation then $\alpha^{-1} = \left(\begin{smallmatrix}1&2&3&4&5\\2&5&3&1&4\end{smallmatrix}\right)$. Note that $\alpha\alpha^{-1} = id = \alpha^{-1}\alpha$.

[1] This confusing order reversal is a consequence of the fact that functions are normally written on the left of their argument. It can be avoided by deciding instead to write functions on the right of their argument.

In terms of diagrams, the diagram of α^{-1} is obtained by turning the diagram of α upside down (and reversing the direction of its arrows).

6.13 Theorem (properties of permutation multiplication)**.** *Let $\alpha, \beta, \gamma \in \mathbb{S}_n$. Then:*

(a) multiplication is associative: $(\alpha\beta)\gamma = \alpha(\beta\gamma)$,
(b) the identity is neutral for multiplication: $\alpha\, id = id\, \alpha = \alpha$,
(c) inverses exist: $\alpha^{-1} \in \mathbb{S}_n$ exists such that $\alpha^{-1}\alpha = id = \alpha\alpha^{-1}$.

Proof. (a) It is well known (and easy to check) that composition of functions is associative. Since permutations are functions, composition of permutations is associative.

(b) This is obvious.

(c) We have already discussed this, in 6.11. □

Another important property of permutation multiplication is: *the inverse of a product is the product of the inverses taken in reverse order:* $(\alpha\beta)^{-1} = \beta^{-1}\alpha^{-1}$. The proof is an exercise.

6.14 Definition (cycles and cycle notation)**.** A permutation in $\mathbb{S}_n$ which maps $i_1 \to i_2 \to \cdots \to i_{r-1} \to i_r \to i_1$ and which fixes all other numbers in the set $\underline{\mathbf{n}}$ is called an r-cycle. In the *cycle notation* it is written as $(i_1, i_2, \ldots, i_r)$.

6.15 Example. The permutation $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 3 & 2 & 4 \end{smallmatrix}\right)$ is a 3-cycle in $\mathbb{S}_5$. In the *cycle notation* we write this permutation as $\alpha = (2, 5, 4)$.

Note that the cycle notation is ambiguous unless n is specified. Also, there is more than one way to write a cycle in the cycle notation, since for instance $(2, 5, 4) = (5, 4, 2) = (4, 2, 5)$ are all the same 3-cycle! Despite these deficiencies, the cycle notation is extremely useful.

6.16 Remark. The identity permutation id in $\mathbb{S}_n$ is (by standard convention) often written as the 1-cycle (1). From now on we will usually write (1) for the identity permutation.

The inverse of a cycle is also a cycle of the same length: it is the cycle obtained by writing the numbers of the original cycle in reverse order. For instance, the inverse of $(2, 5, 4)$ is the cycle $(4, 5, 2)$.

6.17 Definition. A 2-cycle is also known as a *transposition* or *swap*. It simply interchanges two numbers and fixes all others. Every transposition is its own inverse.

6.18 Definition. Two given cycles are said to be *disjoint* if they have no numbers in common.

6.19 Example. The cycles $(2, 5, 4)$ and $(3, 7, 1, 9)$ are disjoint, while $(2, 5, 4)$ and $(3, 5)$ are not.

6.20 Theorem (disjoint cycle factorization)**.** *Any permutation can be written as a product of disjoint cycles. Moreover, disjoint cycles commute with one another, so the product of disjoint cycles can be taken in any order.*

Proof. If all numbers are fixed by the permutation, then it is identity, and can be expressed as a product of disjoint 1-cycles. Otherwise, let i_1 be the first number which is not fixed. It then maps to another number, say i_2, and so on. Because a permutation is a bijection on a *finite* set, eventually we must reach a number i_r which is mapped back to i_1, so we obtain an r-cycle $(i_1, i_2, \ldots, i_r)$. Now continue the argument with the next number which is not fixed by the permutation, and which has not already appeared in some cycle. This process must terminate after finitely many steps since we are permuting finitely many elements. This proves the first claim. The second claim is obvious. □

6.21 Example. The above proof is constructive, and highly computational, in that it provides a *procedure* for computing such a product of disjoint cycles for any given permutation. Here is an illustrative example. Let $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 5 & 3 & 1 & 8 & 9 & 4 & 7 & 6 \end{smallmatrix}\right)$. Then α sends $1 \to 2 \to 5 \to 8 \to 7 \to 4 \to 1$, which gives the 6-cycle $(1, 2, 5, 8, 7, 4)$. The next number that doesn't appear in this cycle is 3, which is a fixed point. The next after that is 6, and α sends $6 \to 9 \to 6$, which gives a 2-cycle $(6, 9)$. At this point every number not fixed by α appears in a cycle, so we are finished. The permutation α has the cycle factorization $\alpha = (1, 2, 5, 8, 7, 4)(6, 9) = (6, 9)(1, 2, 5, 8, 7, 4)$. Some people might include the 1-cycle (3) in order to emphasize the fact that 3 is fixed, writing $\alpha = (1, 2, 5, 8, 7, 4)(6, 9)(3)$. This is fine, too.

It is a fundamental fact that every permutation is expressible as a product of transpositions (not necessarily disjoint). One way to prove this uses the following observation.

6.22 Lemma. *Any r-cycle can be written as a product of (not necessarily disjoint) transpositions.*

Proof. One verifies the identity $(i_1, i_2, \ldots, i_r) = (i_1, i_2)(i_2, i_3) \cdots (i_{r-1}, i_r)$ by direct calculation. □

6.23 Examples. 1. $(3,4,5) = (3,4)(4,5)$.

2. $(3,6,4,2) = (3,6)(6,4)(4,2)$.

It should also be noted that there is *always* more than one way to express a given permutation as a product of transpositions. If σ is any permutation and τ any transposition then $\sigma = \sigma\tau^2$ (because $\tau^2 = id$). Thus, if σ is factored as a product of transpositions, then by tacking on two additional factors of τ you have found another such factorization of σ.

Furthermore, there are *other* ways of factoring into transpositions that do not arise from the formula in the lemma. For instance, check that $(3,4,5) = (4,5)(5,3)$.

6.24 Theorem. *Any permutation can be written as a product of (not necessarily disjoint) transpositions.*

Proof. Combine 6.22 with 6.20. □

As already noted, there are always many different ways (in fact, infinitely many) to factor a permutation as a product of transpositions.

Exercises

6.1. Consider the permutations $\alpha = \left(\begin{smallmatrix}1&2&3&4&5&6\\3&1&5&6&2&4\end{smallmatrix}\right)$ and $\beta = \left(\begin{smallmatrix}1&2&3&4&5&6\\6&3&4&1&5&2\end{smallmatrix}\right)$.

(a) Compute the products $\alpha\beta$ and $\beta\alpha$.

(b) What permutation is α^{-1}?

(c) Compute α^2 and α^3.

(d) What is the smallest positive power of α which equals identity? (I.e., compute the order of α.)

6.2. Compute the order $|\beta|$ of $\beta = \left(\begin{smallmatrix}1&2&3&4&5&6&7&8&9&10&11&12&13&14\\3&1&5&6&2&4&13&11&9&7&10&8&14&12\end{smallmatrix}\right)$. Justify your answer.

6.3. (a) Write out all 3-cycles in $\mathbb{S}_4$. How many are there?

(b) How many 3-cycles are there in $\mathbb{S}_n$?

(c) For any r, how many r-cycles are there in $\mathbb{S}_n$?

6.4. (a) Write the following permutation as a product of disjoint cycles:

$$\alpha = \left(\begin{smallmatrix}1&2&3&4&5&6&7&8&9\\6&1&7&5&4&2&8&9&3\end{smallmatrix}\right).$$

(b) Now write α as a product of transpositions.

6.5. Show that if $\alpha = \alpha_1\alpha_2\cdots\alpha_k$ is a product of disjoint cycles, then $\alpha^t = \alpha_1^t\alpha_2^t\cdots\alpha_k^t$. Show by an example that this may fail for products of non-disjoint cycles.

6.6. (Maps on the right). Suppose we decide to write functions on the *right* of their argument instead of on the left. (This is called *postfix* notation in computer science.) This means that we write $(x)f$ instead of $f(x)$. One usually reads $(x)f$ as *the image of* x *under* f. In this notation, the definition of $f \circ g$ becomes $(x)(f \circ g) = ((x)f)g$. In other words, this fixes the annoying order reversal that we see when maps are written on the left. Show that if we adopt this convention then:

(a) $(1, 2, 3) = (3, 2)(2, 1)$.

(b) $(i_1, i_2, \ldots, i_r) = (i_r, i_{r-1}) \cdots (i_3, i_2)(i_2, i_1)$.

(c) The product $\alpha\beta$ can be computed by placing the diagram of α above the diagram of β and then reading the edges from the top of α to the bottom of β.

7 Permutation Groups

We now come to our first examples of groups, the permutation groups. These were first studied by Lagrange in the late 1700s. Here are the two key definitions.

7.1 Definition. Let G be a nonempty subset of $\mathbb{S}_n$.

(a) We say that G is *closed under products* if $\alpha, \beta \in G$ implies $\alpha\beta \in G$.

(b) We say that G is *closed under inverses* if $\alpha \in G$ implies $\alpha^{-1} \in G$.

7.2 Definition. A nonempty subset $G \subset \mathbb{S}_n$ is a *group of permutations*, or *permutation group* for short, if the set G is closed under products and inverses.

Notice that *any permutation group must contain the identity permutation*, since it contains some element α (because it is nonempty) and contains α^{-1} (because of closure under inverses) and thus must contain $\alpha\alpha^{-1} = id = (1)$ (by closure under products). Hence, if a given subset of $\mathbb{S}_n$ does not include the identity permutation then it cannot be a permutation group.

7.3 Examples. 1. The set $\mathbb{S}_n$ itself is a permutation group, called the *symmetric group* on n letters.

2. The *trivial group* is the group $\{(1)\}$ consisting of just the identity permutation. This group is not interesting; thus the name.

3. The group $K = \{(1), (1,2), (3,4), (1,2)(3,4)\}$ is another example of a permutation group.

7.4 Definition (order of a group)**.** The *order* of a permutation group is the number of elements in the group. In other words, the order of a group is its cardinality as a set. If G is a group, then we always write $|G|$ for its order.

Note that permutation groups always have finite order since they are, by definition, subsets of some $\mathbb{S}_n$, and $\mathbb{S}_n$ is itself a finite set. We will see examples of groups of infinite order later.

7.5 Definition (powers of a permutation)**.** Let $\alpha \in \mathbb{S}_n$ be a permutation and let $m \in \mathbb{Z}$ be a positive integer. We define:

$$\begin{aligned} \alpha^m &= \alpha\alpha\cdots\alpha \quad (m \text{ factors}) \\ \alpha^0 &= (1) = id \\ \alpha^{-m} &= \alpha^{-1}\alpha^{-1}\cdots\alpha^{-1} \quad (m \text{ factors}) \end{aligned}$$

Notice that $(\alpha^m)^{-1} = \alpha^{-m}$. Furthermore, $\alpha^r\alpha^s = \alpha^{r+s}$ and $(\alpha^r)^s = \alpha^{rs}$ for all $r, s \in \mathbb{Z}$. So the usual laws of exponents (but only for integer exponents) are applicable to powers of permutations.

7.6 Lemma. *Let $\alpha \in \mathbb{S}_n$. Then there must be some positive integer p such that $\alpha^p = id$.*

Proof. Since the set $\mathbb{S}_n$ is closed under products, the set $S = \{\alpha, \alpha^2, \alpha^3, \dots\}$ of all positive integer powers of α is contained in $\mathbb{S}_n$. So the set S must be a *finite* set. This implies that there must be repetition in the powers of α. In other words, there exist distinct positive integers r, s such that $\alpha^r = \alpha^s$. We may assume that $r > s$ (otherwise we can just interchange them). Then it follows that

$$\alpha^{r-s} = \alpha^r\alpha^{-s} = \alpha^s\alpha^{-s} = \alpha^0 = id.$$

Since $r - s > 0$, this completes the proof of the lemma. □

7.7 Definition (order of a permutation)**.** Let $\alpha \in \mathbb{S}_n$. The *order* of α (written as order(α) or $|\alpha|$) is the <u>smallest</u> positive integer m such that $\alpha^m = id$.

Note that only verifying the property $\alpha^m = id$ is *not* enough to prove that the order of α is m. You also need to show that the positive exponent m is minimal with respect to that property.

7.8 Example. The order of the identity id is the integer 1: $|id| = 1$.

7.9 Example. If $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 3 & 5 & 2 \end{smallmatrix}\right)$ then $|\alpha| = 4$, as you should verify by computing the successive powers α^2, α^3, and α^4. Since $\alpha^1 \neq id$, $\alpha^2 \neq id$, $\alpha^3 \neq id$, but $\alpha^4 = id$, it follows that 4 is the smallest positive exponent of α giving the identity.

7.10 Proposition. *Let α be a permutation. If $|\alpha| = m$ and $\alpha^k = id$ for some positive integer k then k must be a multiple of m.*

Proof. Divide k by m to get an integer quotient q and remainder r, so that

$$k = qm + r \quad \text{and} \quad 0 \leq r < m.$$

Then $id = \alpha^k = \alpha^{qm+r} = (\alpha^m)^q\alpha^r = id\,\alpha^r = \alpha^r$, so $\alpha^r = id$. If $r > 0$ then we contradict the fact that $|\alpha| = m$, so it follows that $r = 0$ and thus $k = qm$ as required. □

Since any permutation can be written as a product of disjoint cycles, the following result enables us to easily compute the order of any permutation. Note that if $\alpha = \alpha_1\alpha_2\cdots\alpha_k$ is a product of disjoint cycles, then $\alpha^t = \alpha_1^t\alpha_2^t\cdots\alpha_k^t$, because disjoint cycles commute.

7.11 Proposition. *The order of any r-cycle is r. The order of a product $\alpha_1\alpha_2\cdots\alpha_n$ of* disjoint *cycles is the least common multiple (lcm) of their individual orders.*

Proof. Let $\alpha = (i_1, i_2, \ldots, i_r)$ be any r-cycle. Then for any $j < r$, the map α^j sends i_1 to i_j and hence cannot be equal to the identity (1). On the other hand, it is easy to check that $\alpha^r = (1)$. This proves the first claim. The proof of the second claim is left to you as an exercise. □

7.12 Example. If $\alpha = (6, 9, 5)(2, 7, 3, 10)(1, 11)$ then $|\alpha| = \text{lcm}(3, 4, 2) = 12$. This follows from the preceding result.

7.13 Definition (cyclic groups)**.** Let $\alpha \in \mathbb{S}_n$. The smallest permutation group containing α is called the *cyclic group* generated by α. This group is often written as $\langle\alpha\rangle$.

7.14 Proposition. *Let $\alpha \in \mathbb{S}_n$, and let $G = \langle\alpha\rangle$ be the cyclic group generated by α. Then $G = \{id, \alpha, \alpha^2, \ldots, \alpha^{k-1}\}$, where $k = |\alpha|$.*

Note that the order of the cyclic group G generated by α is the same as the order of α itself: if $G = \langle\alpha\rangle$ is cyclic then $|G| = |\alpha|$. This is a characteristic property of cyclic groups. The notation C_k is often used for a cyclic group of order k.

7.15 Proposition. *Let $\alpha \in \mathbb{S}_n$. If $|\alpha| = m$ then $\alpha^{-1} = \alpha^{m-1}$.*

Proof. $\alpha^{m-1} = \alpha^m\alpha^{-1} = (1)\alpha^{-1} = \alpha^{-1}$. □

This implies the following useful fact.

7.16 Theorem (closure under products suffices)**.** *Suppose that G is a nonempty subset of $\mathbb{S}_n$. Then G is a permutation group if and only if the set G is closed under products.*

Proof. Suppose that $\alpha \in G$. Since G is closed under products, it is clear that G must contain the subgroup $\langle\alpha\rangle$ generated by α. So by Proposition 7.15, $\alpha^{-1} = \alpha^{m-1} \in G$, where $m = |\alpha|$. □

Exercises

7.1. Does there exist a permutation group of order r for any given positive integer r? Justify your answer.

7.2. Compute the cyclic group $\langle\alpha\rangle$ generated by α for $\alpha = (1,2,3,4)$ in $\mathbb{S}_4$. What is the order of the group?

7.3. Let $\alpha = (i_1, i_2, \ldots, i_r)$ be an r-cycle. Write α^2 as a product of disjoint cycles. [You may have to distinguish the cases where r is odd or even.]

7.4. Prove the second claim in Proposition 7.11.

7.5. Let $\alpha = (1,2)(3,4)$ and $\beta = (1,2,3,4)$ in $\mathbb{S}_4$. By definition, the group $G = \langle\alpha,\beta\rangle$ *generated* by α, β is the smallest permutation group containing both α, β. Find and list all the elements of G. What is $|G|$?

7.6. Let $\alpha = (1,2)$ and $\beta = (1,2,3)$ in $\mathbb{S}_3$. Let $G = \langle\alpha,\beta\rangle$ be the group generated by α, β. Show that $G = \mathbb{S}_3$.

7.7. (a) Show that $(1,3) = (2,3)(1,2)(2,3)$.

(b) Show that $(1,4) = (3,4)(2,3)(1,2)(2,3)(3,4)$.

(c) Prove that for $j > 1$ we have $(1,j) =$

$$(j-1,j)(j-2,j-1)\cdots(1,2)\cdots(j-2,j-1)(j-1,j).$$

(d) Prove that for $i < j$ we have $(i,j) =$

$$(j-1,j)(j-2,j-1)\cdots(i,i+1)\cdots(j-2,j-1)(j-1,j).$$

Part (d) shows that it is possible to write any transposition as a product of *adjacent* ones; i.e., ones of the form $(k, k+1)$.

7.8. Prove that $\mathbb{S}_n$ is generated by the set $\{(1,2),(2,3),\ldots,(n-1,n)\}$ of adjacent transpositions. [Hint: By Theorem 6.24 it is enough to show that any transposition is expressible as a product of the ones in the given set. Now use the result of Problem 7.7.]

7.9. (a) Show that if $\alpha = (1,2)$, $\beta = (1,2,\ldots,n)$ then for any $1 < i < n$ we have $(i,i+1) = \beta^{i-1}\alpha(\beta^{i-1})^{-1} = \beta^{i-1}\alpha\beta^{n-i+1}$.

(b) Prove that $\mathbb{S}_n$ is generated by the set $S = \{(1,2),(1,2,3,\ldots,n)\}$. [Hint: Use part (a) and the result of the preceding exercise.]

8 The sign of a permutation

We have shown that any permutation can be factored as a product of transpositions, but in infinitely many ways. It is time to investigate this in greater detail.

8.1 Theorem (the identity is even)**.** *Every factorization of the identity id as a product of transpositions must use an even number of transpositions.*

This might seem obvious, but a careful proof is difficult. The proof can be done by induction on the number of transpositions. We omit the proof, and refer to the excellent article[2] by Keith Conrad (see the bibliography at the end).

8.2 Corollary. *Let $\alpha \in \mathbb{S}_n$. Suppose that $\alpha = \sigma_1 \cdots \sigma_r$ and $\alpha = \tau_1 \cdots \tau_s$ are two ways of expressing α as a product of transpositions. Then the difference $r - s$ must be an even number.*

Proof. We have $id = \alpha\alpha^{-1} = \sigma_1 \cdots \sigma_r \tau_s \cdots \tau_1$. By the previous theorem, $r + s$ is even. This implies that $r - s$ must be even. (If $r - s$ is odd then $2r = (r + s) + (r - s)$ would be odd, which is absurd.) □

The corollary implies that if we find one way to factor $\alpha \in \mathbb{S}_n$ as a product of an odd number of transpositions, then all ways of expressing α as a product of transpositions uses an odd number of them. The same holds if we replace "odd" by "even." This means that the following definition makes sense.

8.3 Definition. Let $\alpha \in \mathbb{S}_n$. We say that α is an *odd permutation* if there is some way to express α as a product of an odd number of transpositions. We say that α is an *even permutation* if there is some way to express α as a product of an even number of transpositions. The *sign* of α is defined to be

$$\operatorname{sgn}(\alpha) = \begin{cases} -1 & \text{if } \alpha \text{ is odd} \\ 1 & \text{if } \alpha \text{ is even.} \end{cases}$$

8.4 Examples. 1. The sign of id is 1. (The identity is even.)

2. The sign of any transposition is -1. (A transposition is odd.)
3. The sign of $(1, 2, 3) = (1, 2)(2, 3)$ is 1. The 3-cycle $(1, 2, 3)$ is even.
4. The sign of any r-cycle is $(-1)^{r-1}$. The proof is an exercise.

[2] Keith Conrad, *The sign of a permutation*, `http://www.math.uconn.edu/ kconrad/blurbs/grouptheory/sign.pdf`.

8.5 Theorem (sign is multiplicative)**.** *For any $\alpha, \beta \in \mathbb{S}_n$ we have* $\operatorname{sgn}(\alpha\beta) = \operatorname{sgn}(\alpha)\operatorname{sgn}(\beta)$.

The proof is an easy exercise.

8.6 Definition. For any $n \geq 2$, the *alternating group* $\mathbb{A}_n$ is the set of all even permutations in $\mathbb{S}_n$.

It remains to verify that the definition makes sense. To do so, we must show that the set $\mathbb{A}_n$ of even permutations is closed under products. This is an exercise. Note that $\mathbb{A}_1$ is the empty set.

8.7 Examples. 1. $\mathbb{A}_2 = \{(1)\}$. So $|\mathbb{A}_2| = 1$.

2. $\mathbb{A}_3 = \{(1), (1,2,3), (3,2,1)\}$. This is the same as the cyclic group generated by $(1,2,3)$. So $|\mathbb{A}_3| = 3$.

3. $\mathbb{A}_4 = \{(1), (1,2,3), (1,2,4), (1,3,4), (2,3,4), (3,2,1), (4,2,1), (4,3,1), (4,3,2), (1,2)(3,4), (1,3)(2,4), (1,4)(2,3)\}$. So $|\mathbb{A}_4| = 12$.

3. In general, for any $n \geq 2$, we have $|\mathbb{A}_n| = n!/2$.

8.8 Proposition. $|\mathbb{A}_n| = n!/2$ *for any* $n \geq 2$.

Proof. Let β be any transposition, say $\beta = (1,2)$ for instance. Holding β fixed, it is easy to check that the mapping $f : \mathbb{A}_n \to \mathbb{S}_n$ defined by $f(\alpha) = \alpha\beta$ is a bijection between the disjoint sets $\mathbb{A}_n$ and $\mathbb{S}_n - \mathbb{A}_n$. Hence $|\mathbb{A}_n| = |\mathbb{S}_n - \mathbb{A}_n|$. If we write $k = |\mathbb{A}_n|$ then the fact that $\mathbb{S}_n = \mathbb{A}_n \cup (\mathbb{S}_n - \mathbb{A}_n)$ along with the disjointness of the two sets implies that $|\mathbb{S}_n| = |\mathbb{A}_n| + |\mathbb{S}_n - \mathbb{A}_n|$. In other words, $n! = k + k$, so $k = n!/2$, as required. □

8.9 Remark. It can be shown that $\operatorname{sgn}(\alpha) = (-1)^{I(\alpha)}$, where $I(\alpha)$ is the number of *inversions* in α. By definition, an inversion occurs when $i < j$ but $\alpha(i) > \alpha(j)$. You can count the number of inversions in α by drawing the diagram of α and counting the number of edge crossings.

8.10 Remark. One important application of permutations is the following closed formula for the determinant of an $n \times n$ matrix $A = (a_{ij})$:

$$\det A = \sum_{\alpha \in \mathbb{S}_n} \operatorname{sgn}(\alpha)\, a_{1,\alpha(1)} a_{2,\alpha(2)} \cdots a_{n,\alpha(n)}.$$

This formula says that to compute the determinant one must form a product of entries chosen one from each row and column, and then take the signed sum of all such products.

Exercises

8.1. Show the inverse of a permutation must have the same sign as the permutation.

8.2. Prove that for any $n \geq 2$ the set $\mathbb{A}_n$ of all even permutations is a group.

8.3. Let S be the set of all odd permutations in some fixed $\mathbb{S}_n$. Is S a permutation group? Why or why not?

8.4. Prove that for $n \geq 2$ the order of the alternating group $\mathbb{A}_n$ is $n!/2$. [Hint: Establish a bijection f from $\mathbb{A}_n$ onto the set B_n of all odd permutations in $\mathbb{S}_n$, by choosing any transposition τ and setting $f(\alpha) = \tau\alpha$. Show this is a bijection, and conclude that $|\mathbb{A}_n| = |B_n|$.]

8.5. Prove Theorem 8.5.

8.6. Find a set of generators for $\mathbb{A}_n$, for $n \geq 3$.

Chapter 2

Symmetry Groups

9 Symmetry groups

What are groups good for, anyway? One answer is that they can be used to quantify *symmetry*. Measuring symmetry is what groups do. Symmetry is ubiquitous in nature, and is an important component of art and music. In chemistry, symmetry groups distinguish between different molecular structures and describe properties of crystals; in physics, symmetry groups help us understand interactions of subatomic particles. The symmetry group of the Rubik's cube is helpful for solving the puzzle. The original application of symmetry groups was in the theory of polynomial equations, where Galois showed that the symmetry group of the equation determines whether or not it can be solved in terms of radicals. We will only touch on a few simple examples of symmetry groups. This is a complex topic, which can be studied from several different viewpoints.

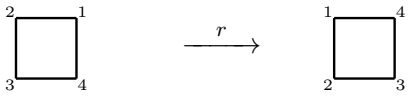
9.1 Example (rotation group of a square)**.** A *rotational symmetry* of a square is a rotation of the square which brings it back to itself; it is always assumed that such rotations fix the center point of the square. If the square has no marks, we would not be able to tell whether or not it was rotated. Such rotations are said to *preserve* the square.

The *rotation group* of the square is the set of all of its rotational symmetries. This set is a group, because we can compose rotations by following one by another.

How many rotational symmetries (of the square) are there? The answer is that it depends on how we decide to count them. It is useful to number the vertices of the square in order to keep track of the effect of rotations.

Let's number the vertices as follows:

We will call this configuration the *start* configuration. Let i be the rotation by zero degrees. This rotation does nothing to the configuration, so i is the identity rotation. Another rotational symmetry is the counterclockwise rotation r by 90 degrees (i.e. $\pi/2$ radians), depicted below:



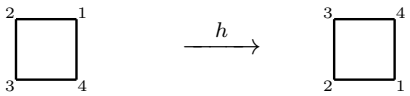
The rotation r actually does something to the square, so it is not the same as the identity i. The rotations r^2 = rotation by 180 degrees, and r^3 = rotation by 270 degrees are also symmetries of the square. Let's adopt the convention that two symmetries are consider the same if and only if they are indistinguishable from one another in terms of the numbering of the vertices. Then r^4 = rotation by 360 degrees will bring the square back to its starting position, so $r^4 = i$.

We have found four distinct rotational symmetries thus far: namely the rotations i, r, r^2, r^3, so the rotation group of the square contains at least the elements $\{i, r, r^2, r^3\}$. Are there any more rotational symmetries? The answer is no. It is true that r^{-1} = rotation by -90 degrees is a symmetry, but it is the same as the rotation by 270 degree; i.e., $r^{-1} = r^3$. So r^{-1} is already counted. Similarly, you may think that r^5 = rotation by 450 degrees is a new rotation, but since $r^4 = i$ it follows that $r^5 = r$. So r^5 is not anything new. It turns out that any rotational symmetry you might think of is already in our list. So the rotation group of a square is the group

$$G = \{i, r, r^2, r^3\} = \langle r \rangle$$

of order 4, generated by the basic rotation r.

9.2 Example (improper symmetries)**.** There are other symmetries of a square besides the rotations. Let h be the reflection of the points on the square across its horizontal axis. The effect of h can be depicted as follows:



Reflections are sometimes called *improper* symmetries; in this parlance the rotations would be called *proper* symmetries.

Are there any other improper symmetries of the square? Yes. There are three other reflections:

v = reflection across the vertical axis,

d_1 = reflection across the diagonal line connecting vertices 1 and 3,

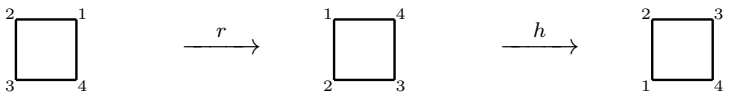
d_2 = reflection across the diagonal line connecting vertices 2 and 4.

It turns out that we have found them all. There are precisely four improper symmetries of the square, namely $\{h, v, d_1, d_2\}$.

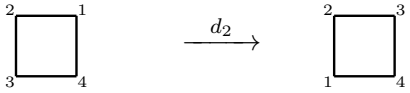
The improper symmetries do *not* form a group, however. To be a group, a set must be closed under products and inverses, and the set $\{h, v, d_1, d_2\}$ of improper symmetries is not closed under products, because the square of a reflection is the identity i and i is not in the set.

However, if we combine the proper and improper symmetries in a set, we do obtain a group. So the set $\mathbb{D}_4 = \{i, r, r^2, r^3, h, v, d_1, d_2\}$ of proper and improper symmetries of the square is a group; it is called the *symmetry group* of the square. One nice way to see that it really is a group is to work out its multiplication table.

As already mentioned, we multiply symmetries by composition, i.e., following one by the other. For example, to compute the product hr means to first do r and then do h. (Recall that $hr = h \circ r$ means r followed by h.) The product hr can be depicted by:



The net result of combining the two operations in succession is the same as the reflection d_2,



so we write $hr = d_2$. By doing similar calculations, one can work out the full multiplication table of the symmetry group $\mathbb{D}_4$. This group is called a *dihedral group.*

9.3 Remark. Improper symmetries can also be viewed as rotations. The reflection of a plane figure in a given line may be regarded as a rotation of the plane containing the figure by π radians about that line, which serves as the axis of the rotation. For this it is necessary to work in three dimensions. The proper rotations, on the other hand, are rotations of the plane containing the figure. Proper rotations take place entirely in the plane, and thus do not require three dimensions to describe.

The following terminology, which is a basic concept, will be used frequently.

9.4 Definition. Let G, H be groups. An *isomorphism* from G onto H is a bijection $f : G \to H$ which preserves products (in the sense that $f(ab) = f(a)f(b)$ for all $a, b \in G$). We say that G is *isomorphic to* H (written as $G \cong H$) if there is an isomorphism $f : G \to H$.

The fact that an isomorphism is a bijection means that if $G \cong H$ then $|G| = |H|$. The fact that it preserves products means that the multiplication tables for the two groups are identical, except for the names of the elements. Of course, names are purely arbitrary labels and have no intrinsic meaning, so it follows that isomorphic groups are essentially the "same" group apart from the names of the elements. So, informally speaking, the word "isomorphic" means "the same structure."

The concept of isomorphism may be initially somewhat abstract, so let's look at a concrete example.

9.5 Example. Let G be the symmetry group of the square, as in the previous example. We can think of each symmetry in G as corresponding to the permutation whose second row records the *position* in which each vertex ends up under the motion. In short, we say that each symmetry in G can be *represented* by a permutation. The representation is given in the following table.

symmetry	permutation
i	(1)
r	$(1,2,3,4)$
r^2	$(1,3)(2,4)$
r^3	$(4,3,2,1)$
h	$(1,4)(2,3)$
v	$(1,2)(3,4)$
d_1	$(2,4)$
d_2	$(1,3)$

Note that the permutations are written in terms of the cycle notation. A bit of thought reveals that the 8 permutations in the rightmost column of the table form a permutation group P. The function $f : \mathbb{D}_4 \to P$ which sends the element in the left column to the corresponding permutation in the right column defines a bijection from $\mathbb{D}_4$ onto P. Although it is tedious do do so at this point, it can be checked that f preserves products. Thus f is an isomorphism and $\mathbb{D}_4 \cong P$. So we may think of $\mathbb{D}_4$ as a permutation group if we so desire.

Furthermore, if we restrict the isomorphism f to the group G of proper symmetries (the rotation group) of the square, then we obtain an isomorphism of G onto the permutation group $H = \langle \alpha \rangle$, where $\alpha = (1, 2, 3, 4)$. The group $H = \{id, \alpha, \alpha^2, \alpha^3\}$ is a cyclic group of order 4, so we have proved that the group G of proper symmetries of a square is isomorphic to a cyclic group of order 4.

9.6 Example. [symmetry group of the regular n-gon] Fix a regular n-sided polygon, where $n \geq 3$. The group of all symmetries of the polygon is denoted by $\mathbb{D}_n$; it is known as the *dihedral group* of order $2n$. When $n = 4$ this is the symmetry group $\mathbb{D}_4$ of a square.

Clearly, r = rotation by $2\pi/n$ radians preserves the polygon. Then r^k = rotation by $2k\pi/n$ radians, and it preserves the polygon as well. Notice that $r^n = i$, the identity. So the proper symmetry group (the rotation group) of the polygon is

$$G = \{i, r, r^2, \ldots, r^{n-1}\} = \langle r \rangle.$$

We have $|G| = n$, so there are n proper symmetries.

There are also n improper symmetries, the reflections. Let ℓ be a line which bisects the polygon (so that the two parts on opposite sides of ℓ are congruent). Then reflection across the line ℓ is an improper symmetry. Let d = the reflection that sends vertex n to itself. Then with a bit of work it is possible to show that the set

$$\{d, dr, \ldots, dr^{n-1}\}$$

is the set of improper symmetries of the polygon. This is due to the fact that following a rotation by a reflection always produces another reflection, and all reflections are produced this way. So the dihedral group $\mathbb{D}_n$ is the group

$$\mathbb{D}_n = \{i, r, r^2, \ldots, r^{n-1}, d, dr, dr^2, \ldots, dr^{n-1}\}.$$

The dihedral group consists of n rotations and n reflections, for a total of $2n$ symmetries. Thus $|\mathbb{D}_n| = 2n$.

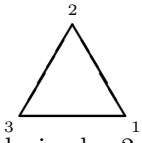
With a bit of effort one can check that the generating rotation r and the reflection d are related by the nice formula

$$rdr = d$$

which is equivalent to the relations $rd = dr^{-1} = dr^{n-1}$. These relations are all one needs to compute products of elements of $\mathbb{D}_n$.

Exercises

9.1. Cut a piece of paper in the shape of an equilateral triangle and number its vertices (on both sides) in order counterclockwise, as shown in the figure below:



Let r = rotation counterclockwise by $2\pi/3$ radians; note that $r^3 = i$ is the identity rotation. Let d_j = reflection in the bisector which fixes the point j, for $j = 1, 2, 3$. Then the symmetries of the triangle are $\mathbb{D}_3 = \{i, r, r^2, d_1, d_2, d_3\}$. Using these labels, work out the multiplication table for the symmetry group $\mathbb{D}_3$.

9.2. Cut a piece of paper into an isosceles triangle which is not equilateral. Describe its symmetry group G, and compute its multiplication table.

9.3. In biology, it is pointed out that many animals have *bilateral symmetry*. Humans are an example. The human body has bilateral symmetry (if the arms and legs are in a symmetric position). What is the symmetry group of the human form? How does it relate to the group G of the previous problem?

9.4. Cut a square piece of paper and number the vertices (on both sides) as in Example 9.1. Using this tool, make a multiplication table for the group $\mathbb{D}_4$.

9.5. Cut a rectangular piece of paper that is *not* square and number the vertices. Using this tool, find the symmetry group of the non-square rectangle. Give the order and a multiplication table of this group.

9.6. How much symmetry does a word have? Let's define a *word* in an alphabet A to be any finite string of symbols from A. For instance, if $A = \{a, b\}$ then $w_1 = abba$, $w_2 = baba$, $w_3 = aba$, and $w_4 = bbb$ are all words over A. Let's define a *symmetry* of a word w to be any permutation of the word w which preserves the word. For instance, the transposition $(1, 3)$, which interchanges the letters in positions 1 and 3, preserves the words w_2, w_3 but not w_1. Let $G(w)$ be the group of all symmetries of a word w.

(a) Compute $G(w)$ for $w = w_1, w_2, w_3, w_4$ as above.
(b) Which of these four words has the *most* symmetry?

9.7. (a) Prove that $rhr = h$ in $\mathbb{D}_4$ by giving a geometric argument.
(b) Similarly, prove that $rd_2r = d_2$ in $\mathbb{D}_4$.

9.8. Given the fact that $\mathbb{D}_5 = \{i, r, r^2, r^3, r^4, d, dr, dr^2, dr^3, dr^4\}$, where r, d are as defined in Example 9.6, compute its multiplication table. Every product in your table must be an element of $\mathbb{D}_5$.

9.9. Use the equation $rdr = d$ from Example 9.6 to prove that in $\mathbb{D}_n$ we have $r^k d = dr^{n-k}$ for $k = 1, 2, \ldots, n-1$.

9.10. Make a table of the permutation representation of $\mathbb{D}_3 = \{i, r, r^2, d_1, d_2, d_3\}$,

using the labels introduced in Problem 9.1. Use the table to show that $\mathbb{D}_3 \cong \mathbb{S}_3$. Explain how your isomorphism is defined.

9.11. Is $\mathbb{D}_4 \cong \mathbb{S}_4$? Justify your answer.

9.12. (a) Find the permutations representing $r, d \in \mathbb{D}_n$, as defined in Example 9.6.

(b) Then describe a permutation group which is isomorphic to $\mathbb{D}_n$.

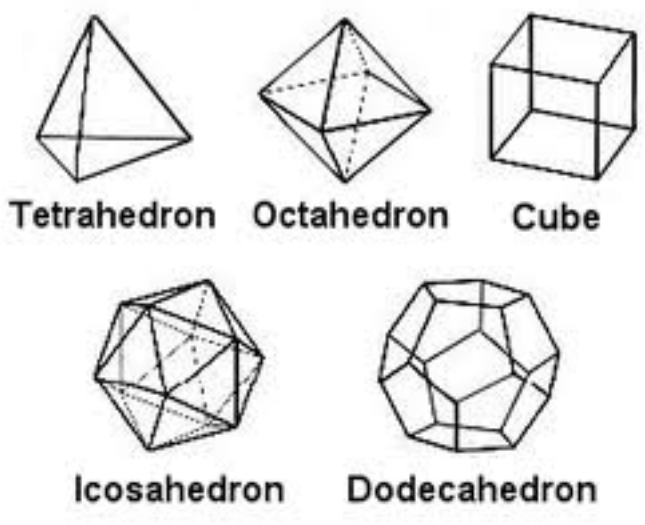
(c) Use the isomorphism from part (b) to show that $rdr = d$.

10 The Platonic solids

Now we turn our attention to symmetry groups in 3-dimensional euclidean geometry. First, we consider the classification of regular polyhedra in solid geometry.

Plane geometry is the study of figures in the plane $\mathbb{R}^2$, while solid geometry means the study of solid figures in $\mathbb{R}^3$ (three dimensions instead of two). A *platonic solid* is a regular figure in three dimensions which is analogous to the regular n-gons in two dimensions. While in two dimensions we have infinitely many such figures (the regular n-gon for each n) in three dimensions *there are only five regular solids.*

We will assume this, as the proof would take too much of our time, but if you want to know more, there is an excellent Wikipedia article on this topic that not only delves into the history but also explains several proofs. The classification of regular solid polyhedra is the final result in Euclid's *Elements*; this was the first published proof.



Here are the names of the five platonic solids, which are pictured above. The *tetrahedron* is a 4-sided solid with equilateral triangles as faces; the *cube* is a 6-sided solid with square faces; the *octahedron* is an 8-sided solid with equilateral triangles as faces; the *dodecahedron* is a 12-sided solid with faces which are regular pentagons; and finally the *icosahedron* is a 20-sided solid with equilateral triangles for its faces.

10.1 Example (proper symmetry groups of the Platonic solids)**.** Let us begin by describing the *proper* symmetries (the rotations) of the five Platonic solids. It is already a difficult problem to count them in some effective way.

The proper symmetry group of the tetrahedron is the *tetrahedral group*, denoted by $\mathbf{T}$. It turns out that $\mathbf{T} \cong \mathbb{A}_4$. This can be demonstrated by considering a permutation representation of the tetrahedron. One can number the vertices and write out the permutations corresponding to the rotations, it turns out that these are precisely the elements of $\mathbb{A}_4$. The details are tedious. The isomorphism $\mathbf{T} \cong \mathbb{A}_4$ means in particular that $|\mathbf{T}| = |\mathbb{A}_4| = 12$. So there are 12 proper rotations of a tetrahedron.

The proper symmetries of the cube give us another group, called the *octahedral group*, and written as $\mathbf{O}$. We have $|\mathbf{O}| = 24$.

The proper symmetries of the octahedron give us a group isomorphic with $\mathbf{O}$, so we don't get a new group from this solid.

This is due to the fact that the cube and the octahedron are *dual* to one another. To obtain the dual of a regular polyhedron, put a dot in the center of each face, and let the dots be the vertices of a new polyhedron. This new polyhedron is the dual of the original one. It is not hard to see that the symmetry group of a polyhedron must be the same as the symmetry group of its dual.

The proper symmetries of the dodecahedron give us a third symmetry group, called the *icosahedral group*, and denoted by the symbol $\mathbf{I}$. This group has 60 elements; in fact it is isomorphic with the alternating group $\mathbb{A}_5$. The proper symmetries of the icosahedron give us a group isomorphic with $\mathbf{I}$, so we don't get a new group from this solid. This is due to the fact that the icosahedron is the dual of the dodecahedron.

Summary: The proper symmetries of the five Platonic solids give us just three new symmetry groups: $\mathbf{T} \cong \mathbb{A}_4$; $\mathbf{O}$; and $\mathbf{I} \cong \mathbb{A}_5$. Two of them are isomorphic to groups we have seen before, but the octahedral group $\mathbf{O}$ is new.

10.2 Example (improper symmetries of the Platonic solids)**.** As in the case of regular polygons, it turns out that there are improper symmetries of all of the regular polyhedra. They are a bit harder to describe, but it turns out that there are always as many improper symmetries as there are proper ones, just as in the polygon case.

When we include the improper symmetries, we get the full symmetry group. The full symmetry group of each Platonic solid therefor has twice as many elements as its proper symmetry group. We need to develop more

theory in order to say more about this topic, so we leave this for now.

Exercises

10.1. Explain why two finite groups of different cardinality cannot be isomorphic.

10.2. Prove that the proper symmetry group of the tetrahedron is isomorphic to the alternating group $\mathbb{A}_4$. One way to do this is via a permutation representation. (Number the vertices of the tetrahedron and regard symmetries as permutations of the vertices.)

10.3. Using the same idea as in the previous problem, prove that the full symmetry group of the tetrahedron is the symmetric group $\mathbb{S}_4$.

10.4. Explain why the symmetry group of a regular solid must be the same as that of its dual.

10.5. Describe the 24 proper symmetries of the cube in words.

A Appendix: Symmetries of polynomials

Symmetry also plays an important role in the study of polynomial equations, where the symmetry groups turn out to be permutation groups. This was Lagrange's motivation to study permutations. But the definition of the symmetry group of a polynomial, which is now known as its *Galois group*, is not as accessible as the symmetry group of a geometric object such as a polygon. In this appendix, we indicate the definition only in an intuitive way.

Our approach is difficult to use, so it turns out to be useful to replace it by something less intuitive, but ultimately easier to work with. The modern approach uses the theory of *field extensions*. Field extensions are typically studied in a second course in abstract algebra, but they will not be considered here.

Suppose we have an nth degree polynomial $p(x)$ with real number coefficients, say

$$p(x) = x^n + c_1x^{n-1} + c_2x^{n-2} + \cdots + c_{n-1}x + c_n. \qquad (*)$$

The coefficients are the real numbers c_j for $j = 1, \ldots, n$. According to the fundamental theorem of algebra , the polynomial $p(x)$ has exactly n roots $z_1, z_2, \ldots, z_n$ in the complex number system $\mathbb{C}$, where we agree to list multiple roots according to their multiplicity.

A.1 Examples. 1. The polynomial $(x-1)^3 = x^3 - 3x^2 + 3x - 1$ has three identical roots $1, 1, 1$. There is one root of multiplicity three (a triple root).

2. The polynomial $(x^2-2)^2 = x^4 - 4x^2 + 4$ has roots $\sqrt{2}$, $\sqrt{2}$, $-\sqrt{2}$, $-\sqrt{2}$. There are two real roots, each of multiplicity two (two double roots).

3. The polynomial $(x^2+1)^2 = x^4 + 2x^2 + 1$ has roots $i, i, -i, -i$. There are two non-real complex roots, each of multiplicity two (two double roots). Here i stands for the *imaginary unit* in the complex number system $\mathbb{C}$. (By definition, $i^2 = -1$.)

4. The polynomial $x(x^3 - 3x + 2) = x^4 - 3x^2 + 2x$ has roots $0, 1, 1, -2$. This has one root of multiplicity two (one double root). You can verify the roots by expanding the product $x(x-1)^2(x+2)$.

For definiteness, we assume that the polynomial $p(x)$ in equation $(*)$ above has *rational* coefficients, and that $p(x)$ is irreducible, meaning that it cannot be factored into polynomials (with rational coefficients) of strictly smaller degree. If $p(x)$ had such a factorization, then we could deal sepa-

rately with the two smaller polynomials in order to solve $p(x) = 0$, so it makes sense to restrict our attention to irreducible polynomials.

We also assume, for simplicity, that the roots of $p(x)$ are all *distinct*, i.e., all the roots have multiplicity one.

A.2 Definition. Under the above assumptions, we define the *Galois group* of the polynomial $p(x)$ to be the permutation group consisting of all permutations of the roots $\{z_1, \ldots, z_n\}$ which preserve any algebraic relations among them.

For a general polynomial[1] of degree n there will be only trivial relations among the roots, and thus the Galois group of the polynomial will be unconstrained; it will thus be the full symmetric group $\mathbb{S}_n$ consisting of all permutations of the n roots. For special polynomials, however, there can be non-trivial relations among the roots imposing constraints on the permutations in the group; in such cases the Galois group of the polynomial is often smaller than $\mathbb{S}_n$. In all cases, the Galois group of an nth degree polynomial will be a permutation group contained in $\mathbb{S}_n$.

Let's look at some concrete examples.

A.3 Example. Consider the polynomial $p(x) = x^4+x^3+x^2+x+1$. Clearly we have

$$p(x) = (x^5 - 1)/(x - 1)$$

as you can see by clearing denominators and expanding the resulting product. The roots of $p(x)$ are $z_1 = e^{2\pi i/5}$, $z_2 = e^{4\pi i/5}$, $z_3 = e^{6\pi i/5}$, $z_4 = e^{8\pi i/5}$. These complex numbers all lie on the unit circle in the complex plane $\mathbb{C}$. (Recall that $e^{i\theta} = \cos\theta + i\sin\theta$ for any angle θ, where the imaginary unit i satisfies $i^2 = -1$.) The roots satisfy the relations

$$z_2 = z_1^2, \quad z_3 = z_1^3, \quad z_4 = z_1^4$$

and any other relations (e.g., $z_3^2 = z_1$) are consequences of these, along with the fact that $z_i^5 = 1$ for any $i = 1, \ldots, 4$. We are looking for permutations of the four roots that preserve these relations. If α is such a permutation, then $\alpha(z_2)$, $\alpha(z_3)$, and $\alpha(z_4)$ will be determined by $\alpha(z_1)$. Now $\alpha(z_1)$ can be z_1, z_2, z_3, or z_4. We consider these possibilities separately.

If $\alpha(z_1) = z_1$ then $\alpha(z_2) = \alpha(z_1)^2 = z_2$, $\alpha(z_3) = \alpha(z_1)^3 = z_3$, and $\alpha(z_4) = \alpha(z_1)^4 = z_4$. Thus $\alpha = (1)$ is the identity permutation.

[1] A general polynomial is one whose coefficients are represented by variables. For instance, $ax^2 + bx + c$ is the general polynomial of degree 2 (the general quadratic), $ax^3 + bx^2 + cx + d$ is the general polynomial of degree 3 (the general cubic), and so on.

If $\alpha(z_1) = z_2$ then $\alpha(z_2) = \alpha(z_1)^2 = z_2^2 = z_4$, $\alpha(z_3) = \alpha(z_1)^3 = z_2^3 = z_1$, and $\alpha(z_4) = \alpha(z_1)^4 = z_2^4 = z_3$. Thus α is the 4-cycle $(1, 2, 4, 3)$.

If $\alpha(z_1) = z_3$ then $\alpha(z_2) = \alpha(z_1)^2 = z_3^2 = z_1$, $\alpha(z_3) = \alpha(z_1)^3 = z_3^3 = z_4$, and $\alpha(z_4) = \alpha(z_1)^4 = z_3^4 = z_2$. Thus α is the 4-cycle $(1, 3, 4, 2)$.

Finally, if $\alpha(z_1) = z_4$ then $\alpha(z_2) = \alpha(z_1)^2 = z_4^2 = z_3$, $\alpha(z_3) = \alpha(z_1)^3 = z_4^3 = z_2$, and $\alpha(z_4) = \alpha(z_1)^4 = z_4^4 = z_1$. Thus $\alpha = (1, 4)(2, 3)$.

From these calculations it follows that the symmetry group of the polynomial $x^4+x^3+x^2+x+1$ is the cyclic group generated by the cycle $(1, 2, 4, 3)$. This Galois group G has order 4.

A.4 Example. Now consider the polynomial $p(x) = x^4 - 10x^2 + 1$. We can factor $p(x)$ as follows:

$$\begin{aligned} p(x) = (x^4 - 2x^2 + 1) - 8x^2 &= (x^2 - 1)^2 - (x\sqrt{8})^2 \\ &= (x^2 - 1 - 2\sqrt{2}x)(x^2 - 1 + 2\sqrt{2}x) \\ &= (x^2 - 2\sqrt{2}x - 1)(x^2 + 2\sqrt{2}x - 1) \end{aligned}$$

and we can find its roots by setting each quadratic factor to zero and using the quadratic formula. The roots are $z_1 = \sqrt{2} + \sqrt{3}$, $z_2 = -\sqrt{2} + \sqrt{3}$, $z_3 = \sqrt{2} - \sqrt{3}$, $z_4 = -\sqrt{2} - \sqrt{3}$. Notice that all the roots are real numbers in this case.

The roots satisfy the relations:

$$\begin{aligned} z_1 + z_4 &= 0 \\ z_2 + z_3 &= 0 \\ (z_1 + z_2)^2 &= 12 \\ (z_1 + z_3)^2 &= 8 \\ (z_2 + z_4)^2 &= 8 \\ (z_3 + z_4)^2 &= 12 \end{aligned}$$

and again all other relations are consequences of these. We want all permutations of the four roots which preserve these relations. You can check that the permutations $(1, 2)(3, 4)$, $(1, 3)(2, 4)$, $(1, 4)(2, 3)$, and the identity (1) preserve the relations. No other permutation does. Thus the symmetry group of the polynomial $p(x) = x^4 - 10x^2 + 1$ is

$$G = \{(1), (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\}.$$

This is a group of order four. It turns out that it is isomorphic with the Klein four-group.

The cubic formula

We now consider the cubic formula, the degree three analogue of the quadratic formula. Given a general cubic polynomial equation:

$$x^3 - bx^2 + cx - d = 0 \tag{A.1}$$

with undetermined (i.e., general) coefficients, we denote its complex roots by z_1, z_2, and z_3. To solve the equation, we first substitute $x = y + b/3$ and obtain (after expanding) the *reduced cubic* equation

$$y^3 + py - q = 0 \tag{A.2}$$

where $p = c - b^2/3$ and $q = d - bc/3 + 2b^3/27$. The roots of the reduced cubic are given by *Cardano's formula*, first published in the year 1545.

Theorem (Cardano 1545)**.** *The roots of the reduced cubic $y^3 + py - q = 0$ are given by*

$$\begin{aligned} y_1 &= \sqrt[3]{\frac{q}{2} + \sqrt{R}} + \sqrt[3]{\frac{q}{2} - \sqrt{R}} \\ y_2 &= \omega^2 \sqrt[3]{\frac{q}{2} + \sqrt{R}} + \omega \sqrt[3]{\frac{q}{2} - \sqrt{R}} \\ y_3 &= \omega \sqrt[3]{\frac{q}{2} + \sqrt{R}} + \omega^2 \sqrt[3]{\frac{q}{2} - \sqrt{R}} \end{aligned} \tag{A.3}$$

where

$$R = \frac{q^2}{4} + \frac{p^3}{27}; \qquad \omega = \frac{-1 + i\sqrt{3}}{2}.$$

Note that $\omega = e^{i2\pi/3}$ and therefore $\omega^3 = 1$. Moreover, $\omega^2 + \omega + 1 = 0$. The complex number ω is called a *primitive cube root of unity*.

To find the roots z_i $(i = 1, 2, 3)$ of the original cubic we only have to add $b/3$ to the y_i. This solves the original cubic equation, and the solution is in terms of certain radicals (square and cube roots), namely those that occur in the expressions for the y_i.

Lagrange's method for solving a cubic

In an important paper published in 1770, Lagrange noticed that the radicals appearing in the solution to the original cubic are expressible as functions of the roots z_i themselves. This observation was the beginning of the link between group theory and polynomial equations.

To see this for the cube roots, we multiply the equations for the y_i by 1, ω, ω^2 respectively and add, obtaining

$$3\sqrt[3]{\frac{q}{2}+\sqrt{R}} = y_1 + \omega y_2 + \omega^2 y_3.$$

Thus if we substitute $y_i = z_i - b/3$ we obtain

$$3\sqrt[3]{\frac{q}{2}+\sqrt{R}} = z_1 + \omega z_2 + \omega^2 z_3.$$

Denote this value by φ_1, and set $\varphi_3 = \omega^2\varphi_1$, $\varphi_5 = \omega\varphi_1$, so that we have equalities

$$\begin{aligned}
\varphi_1 &= 3\sqrt[3]{\frac{q}{2}+\sqrt{R}} \\
\varphi_3 &= 3\omega^2\sqrt[3]{\frac{q}{2}+\sqrt{R}} \\
\varphi_5 &= 3\omega\sqrt[3]{\frac{q}{2}+\sqrt{R}}.
\end{aligned}$$

Going back and multiplying the equations for y_1, y_2, y_3 by 1, ω^2, ω respectively, we obtain after adding them and substituting for $y_i = z_i - b/3$ the equality

$$3\sqrt[3]{\frac{q}{2}-\sqrt{R}} = z_1 + \omega^2 z_2 + \omega z_3.$$

Let us denote this value by φ_2, and set $\varphi_4 = \omega\varphi_2$, $\varphi_6 = \omega^2\varphi_2$. Then we have equalities

$$\begin{aligned}
\varphi_2 &= 3\sqrt[3]{\frac{q}{2}-\sqrt{R}} \\
\varphi_4 &= 3\omega\sqrt[3]{\frac{q}{2}-\sqrt{R}} \\
\varphi_6 &= 3\omega^2\sqrt[3]{\frac{q}{2}-\sqrt{R}}.
\end{aligned}$$

We can also express $\sqrt{R}$ in terms of the roots z_i by cubing the equations for φ_1 and φ_2 and subtracting. Doing this, we obtain

$$54\sqrt{R} = (z_1 + \omega z_2 + \omega^2 z_3)^3 - (z_1 + \omega^2 z_2 + \omega z_3)^3$$

and after expanding, simplifying, and factoring we have

$$18\sqrt{R} = 3i(z_1 - z_2)(z_1 - z_3)(z_2 - z_3).$$

Thus all the radicals appearing in the formulas for the roots are expressible as nice functions of the roots themselves.

Lagrange noticed that the roots of the general cubic are obtainable from the six values φ_i $(i = 1, \ldots, 6)$, because we have the system of linear equations

$$\begin{aligned} \varphi_1 &= z_1 + \omega z_2 + \omega^2 z_3 \\ \varphi_2 &= z_1 + \omega^2 z_2 + \omega z_3 \\ b &= z_1 + z_2 + z_3 \end{aligned} \tag{A.4}$$

which can easily be solved for the roots z_1, z_2, z_3 by first adding them as they stand, then adding them after multiplying by ω^2, ω, 1, and then adding them again after multiplying by ω, ω^2, 1, respectively.

Lagrange also pointed out that we can compute the six values φ_i as follows. Set $A_1 = q/2 + \sqrt{R}$ and $A_2 = q/2 - \sqrt{R}$. Then we have the equations

$$\left(\frac{\varphi_i}{3}\right)^3 = A_1 \qquad (i = 1, 3, 5)$$

and

$$\left(\frac{\varphi_i}{3}\right)^3 = A_2 \qquad (i = 2, 4, 6).$$

In other words, the six φ_i are obtained by solving the equations $X^3 = 27A_1$ and $X^3 = 27A_2$. (Note that if φ is one root, say the cube root of A_i, then the other two roots must be $\omega\varphi$ and $\omega^2\varphi$.)

Examining the expressions defining A_1 and A_2, we see that we can find the A_i by solving the quadratic equation

$$A^2 - qA - p^3/27 = 0 \tag{A.5}$$

for A. In this way Lagrange reduced the solution of the cubic to the solution of a quadratic. (He called this associated quadratic equation the *resolvent*.) By first solving the resolvent quadratic and then solving the system (A.4) you can solve the cubic using Lagrange's method.

Lagrange also similarly analyzed the general quartic (degree 4) polynomial equation, reducing it to a cubic resolvent equation. He must have been quite excited at that point, thinking that he had discovered a general scheme to solve all polynomial equations inductively by reducing them to a resolvent equation of degree one less than the given equation. However,

when he looked at the very next case, the general quintic (degree 5) equation, he found that the resolvent equation had degree 6. This was a strong indication that the general quintic might be unsolvable in terms of radicals, and indeed N. H. Abel proved that very result in 1826.

Theorem (Abel 1826)**.** *The general quintic (degree five) polynomial equation cannot be solved in terms of radicals.*

Galois further developed group theory and the theory of polynomial equations. He was able to give necessary and sufficient conditions on the symmetry group of an arbitrary polynomial (of any degree) for it to be solvable in terms of radicals. His theorem is a vast generalization of Abel's theorem, in that it applies to polynomials of any degree. Galois discovered the theorem in 1830 at the age of 18; he was shot and killed in a duel at the age of 20. The theorem remained unpublished until 1846, and it wasn't until the late 1800s that mathematicians generally understood the significance of his results.

Exercises

A.1. Solve the polynomial equation $p(x) = 0$ of A.4 by setting $y = x^2$ and solving the resulting quadratic equation, and then finding x by taking square roots of y. Compare your answer with the roots of $p(x)$ given in A.4. Can you explain?

A.2. One strategy for solving a polynomial equation $p(x) = 0$ is to guess a root z_1 of the equation. The guess can be checked by substitution in $p(x)$. If $p(z_1) = 0$ then z_1 is a root. Once you have found a root z_1, you can use long division of polynomials to divide $p(x)$ by $x - z_1$. Since roots correspond to linear factors, the fact that z_1 is a root means that when you divide you will obtain a quotient polynomial $q(x)$ such that $p(x) = (x - z_1)q(x)$. Now you have reduced the original problem, of solving $p(x) = 0$, to the smaller problem of solving $q(x) = 0$. By repeating the method, you can eventually factor $p(x)$ completely into linear factors, thus solving the polynomial equation. The problem with this strategy is that it is not always possible to guess a solution. Thus, the method may never get started, or it may stall somewhere along the way. However, it works in a surprising number of cases. Use this method to solve the following cubic equations, using the given guess:

(a) $x^3 + 1 = 0$; guess $z_1 = -1$.
(b) $x^3 - 3x^2 + 3x - 1 = 0$; guess $z_1 = 1$.
(c) $x^3 + 2x + 3 = 0$; find your own guess.

A.3. If a polynomial $p(x)$ with *integer* coefficients has a rational root (a root of the form $\frac{r}{s}$ where r, s are integers), then we can always find it using the

rational roots theorem.

Theorem. *Suppose that* $p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$, *where* $a_0 \neq 0$ *and* $a_k \in \mathbb{Z}$ *for all* $k = 0, 1, \ldots, n$. *If* $z_1 = \pm\frac{r}{s}$ *is a rational root of* $p(x)$ *then* $s \mid a_0$ *(*s *divides* a_0*) and* $r \mid a_n$ *(*r *divides* a_n*).*

Use the rational roots theorem to find all rational roots (if any) of the following polynomials:

(a) $3x^3 + 4x - 7$.
(b) $x^3 + x^2 + x + 2$.
(c) $x^3 + 8$.

A.4. Find all the roots of the polynomials in parts (a), (c) of the previous problem.

A.5. Show that the symmetry group G of Example A.4 is isomorphic with the Klein 4-group. (The Klein 4-group was introduced in Section 7.)

A.6. Use Cardano's formula or Lagrange resolvents to solve the following cubic equations:

(a) $x^3 - 3x + 2 = 0$.
(b) $x^3 - 9x^2 + 24x - 16 = 0$.
(c) $x^3 + 3x^2 + 6x + 2 = 0$.
(d) $x^3 + 3x - 4 = 0$.

Show the steps of your calculations. Notice that by inspection 1 is a root of the first and last equations; did your calculations in part (d) reveal that fact? Do you think something is wrong?

A.7. Let $\omega = e^{2\pi i/n} = \cos(2\pi/n) + i\sin(2\pi/n)$ be a primitive nth root of unity. Prove that the roots of $x^n - 1 = 0$ are the complex numbers $z_k = \omega^k$ for $k = 0, 1, \ldots, n-1$. (This is a complete list of n distinct roots.)

(a) Compute ω^n.
(b) Why is the set G of roots of the polynomial a group? What type of group is it?
(c) If you plot the roots in the complex plane, they form the set of vertices of what geometric figure?

Chapter 3

Modular Arithmetic

11 Modular arithmetic

Modular arithmetic is a novel finite system of arithmetic used in public-key cryptography in order to provide security for internet transactions. It is also used in algebraic coding theory, the mathematical theory underlying the encoding of information on DVDs, satellite communications, etc. And it provides new examples of groups. We begin with some elementary number theory.

11.1 Lemma (division algorithm)**.** *Let n be a positive integer. Let m be any integer. There exist unique integers q, r such that*

$$m = qn + r \quad \textit{and} \quad 0 \leq r < n.$$

The integers q, r are called the quotient *and* remainder*, respectively.*

Proof. Recall the *well-ordering principle* of natural numbers, which states that any nonempty subset of $\mathbb{N} = \{0, 1, 2, \dots\}$ must have a least element. Let

$$S = \{m - kn \mid k \in \mathbb{Z} \text{ and } m - kq \geq 0\}.$$

By construction, S is a subset of $\mathbb{N}$. We need to show it is nonempty. If $m \geq 0$ then $m = m - 0 \cdot n \in S$. If $m < 0$ then $m - mn = m(1 - n) \geq 0$ because $1 - n \leq 0$, hence $m - mn$ is in S. In either case S is not empty. By the well ordering principle, the set S has a least element. Let r be the least element of S and let q be the corresponding value of k. Then $m - qn = r$, so $m = qn + r$. Moreover, $r \geq 0$ since $r \in S$. Finally, $r < n$ since otherwise

$r - n$ would be in S, contradicting the fact that r is the least element of S. □

When m is positive, the usual procedure by which we find the quotient q and remainder r is called long division, as you undoubtedly recall.

Given a real number x, there is a unique integer k such that $k \leq x < k+1$. The integer k is called the *floor* of x, written as $\lfloor x \rfloor = k$. The floor function is also known as the *greatest integer* function. Then for a given pair of integers m, n we have:

$$q = \lfloor m/n \rfloor, \quad r = m - qn.$$

Note that when $m < 0$ we need to pay attention. For instance, we have $20 = 2 \cdot 7 + 6$ for $20 \div 7$, but $-20 = -3 \cdot 7 + 1$ for $-20 \div 7$.

11.2 Definition. Let n be a positive integer greater than 1. We call n the *modulus*. The set $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ is called the set of *residues* modulo n. This is the set of remainders for long division by n, so *residue* is another word for *remainder*.

11.3 Definition (residue function)**.** Let a modulus $1 < n \in \mathbb{Z}$ be given. Given an integer m, let q, r be the unique integers such that $m = qn + r$ and $0 \leq r < n$. Then set $\mathrm{res}_n(m) = r$. The resulting function $\mathrm{res}_n : \mathbb{Z} \to \mathbb{Z}_n$ is a surjection from the set $\mathbb{Z}$ of integers onto the set $\mathbb{Z}_n$ of residues modulo n.

We should read $\mathrm{res}_n(m)$ as "the residue of m modulo n." It is just the remainder of dividing m by n, so we can always compute it by long division.

Now we define addition and multiplication on the set $\mathbb{Z}_n$.

11.4 Definition. Let $n > 1$ be a fixed integer modulus. Given residues $a, b \in \mathbb{Z}_n$, we define

$$a \oplus b = \mathrm{res}_n(a + b), \qquad a \odot b = \mathrm{res}_n(ab).$$

We call the binary operations $\oplus, \odot$ *residue addition* and *residue multiplication.*

The residue addition and multiplication rules just defined can be remembered as a two-step procedure: first, add or multiply in $\mathbb{Z}$ as usual, then take the residue of the result modulo n.

11.5 Remark. We use new symbols $\oplus, \odot$ for residue addition and multiplication, in order to distinguish these new and novel operations from the

usual ones. Eventually, we will drop the extra circle around the symbol and use the usual addition and multiplication symbols. For now, it is useful to use different notation in order to avoid confusion.

11.6 Example. The addition and multiplication tables for $\mathbb{Z}_2$ are given below.

$\oplus$	0	1
0	0	1
1	1	0

$\odot$	0	1
0	0	0
1	0	1

We will see later that the above addition table defines a group. This group appears in computer science as the table defining the behavior of a binary half-adder circuit, which is a basic circuit used in all digital computers.

11.7 Example. The addition and multiplication tables for $\mathbb{Z}_4$ are compiled below.

$\oplus$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\odot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

This is more interesting. Notice that the tables are "closed" systems, in which the combination of two elements produces another one in the set $\mathbb{Z}_4$. Also notice the novel equation $2 \odot 2 = 0$. In residue arithmetic, two nonzero elements can multiply to produce 0, which is somewhat strange at first sight.

Residue arithmetic is also called *modular arithmetic*. Looking at the addition tables (for the operation $\oplus$) in the above examples, we suspect that they might be groups, because the tables "look" similar to multiplication tables for other groups we have seen, and that is indeed the case, as we will see later. But not so for the multiplication tables (for the operation $\odot$). They are *never* examples of groups, because there is no inverse of 0 in the tables. Nevertheless, there is always at least one group inside the multiplication table, if you know how to look for it. We will return to this question later, once we have developed a better definition of group. For now, the focus is on understanding modular arithmetic.

It is notable that we can make sense of subtraction in $\mathbb{Z}_n$. Here are the appropriate definitions, which follow the same pattern as the definitions of negatives and subtraction in the integers.

11.8 Definition. The *negative* $\ominus a$ of a residue $a \in \mathbb{Z}_n$ is defined to be

$$\ominus a = \begin{cases} 0 & \text{if } a = 0 \\ n - a & \text{if } a \neq 0. \end{cases}$$

Once we have negatives, we can define subtraction in terms of adding the negative, which is how subtraction is defined for integers and real numbers.

11.9 Definition. For any $a, b \in \mathbb{Z}_n$, we define $a \ominus b = a \oplus (\ominus b)$.

Notice that $a \oplus (\ominus a) = 0$ and $(\ominus a) \oplus a = 0$, for any $a \in \mathbb{Z}_n$.

11.10 Theorem. *For any $a, b \in \mathbb{Z}_n$, we have $a \ominus b = \mathrm{res}_n(a - b)$.*

The proof is a straightforward exercise.

The theorem gives us another way to compute $a \ominus b$. Depending on the situation, one way might be more convenient than the other, so it can be useful to have both.

11.11 Remark (notational convention)**.** It soon becomes tedious to always write $\oplus, \odot, \ominus$ for the residue operations. The convention is to replace these symbols by their corresponding uncircled symbols $+, \cdot, -$ from ordinary arithmetic. This convention results in equations such as:

$$5 + 5 = 0, \quad 6 + 7 = 3, \quad 6 \cdot 7 = 2, \quad 4 \cdot 5 = 0, \quad -6 = 4, \quad 5 - 6 = 9$$

all of which are valid equations in $\mathbb{Z}_{10}$. This is confusing only if you forget that $+$ really means $\oplus$, $\cdot$ means $\odot$, and $-$ means $\ominus$.

Furthermore, it is conventional to omit the multiplication symbol $\cdot$ in cases where doing so will not cause confusion. So, if a, b are variables representing values in some residue system $\mathbb{Z}_n$, we usually interpret ab as $a \cdot b$, just as we do in ordinary algebra.

Modular arithmetic is actually similar to arithmetic everyone already carries out with clocks. In a standard clock, there are 12 numbers arranged in a circle, and we all understand that 3 hours after 11 o'clock is 2 o'clock. This is the same as the equation $11 + 3 = 2$ in $\mathbb{Z}_{12}$. The only difference between addition in $\mathbb{Z}_{12}$ and clock addition is that in $\mathbb{Z}_{12}$ we have renamed the modulus 12 to 0.

We can visualize modular arithmetic modulo n similarly, by thinking of a clock with the residues $0, 1, \ldots, n - 1$ equally spaced around its face. Whenever we reach n hours, the clock resets to 0.

Exercises

11.1. The *floor* of a real number x (written $\lfloor x \rfloor$) is the greatest integer k such that $k \leq x$. Given integers m, n with $n > 0$, show that $q = \lfloor m/n \rfloor$ is the integer quotient such that $m = qn + r$ with $0 \leq r < n$.

11.2. A pocket calculator (or computer) that produces decimal approximations for quotients can often be used to carry out the division algorithm in the set of integers. Given integers m, n with $n > 0$, to find the integer quotient q such that $m = qn + r$ with $0 \leq r < n$, simply calculate $m \div n$ on the device and then take q to be the integer part (the floor) of the result. Once you have found q, of course $r = m - qn$. Use this to find the integer quotient q and remainder r for each of the following division problems:

(a) $1000 \div 31$.
(b) $-1000 \div 31$.
(c) $16075 \div 652$.
(d) $-16075 \div 652$.

This approach is slightly dangerous, because if the numbers m, n are big enough, machine calculation can be subject to roundoff error. If you are using a pocket calculator, try to find a case where it produces an incorrect integer quotient.

11.3. Compute the following in the modular system $\mathbb{Z}_{12}$.

(a) $7 + 5$, $5 + 5$, $11 + 11$, $-1 + (-1)$.
(b) $3 \cdot 3$, $7 \cdot 5$, $5 \cdot 5$, $11 \cdot 11$, $(-1) \cdot (-1)$.
(c) $3 - 3$, $5 - 7$, $-2 - 10$, $5 - (-3)$, $5 - (-11)$, $-2 - (-10)$.

11.4. Make addition and multiplication tables for $\mathbb{Z}_3$.

11.5. Make addition and multiplication tables for $\mathbb{Z}_5$.

11.6. Make addition and multiplication tables for $\mathbb{Z}_6$.

11.7. Make addition and multiplication tables for $\mathbb{Z}_7$.

11.8. Make addition and multiplication tables for $\mathbb{Z}_8$.

11.9. Compute the following in the modular system $\mathbb{Z}_{1201}$.

(a) $75 + 500$, $800 + 500$, $-1 + (-1)$.
(b) $30 \cdot 30$, $800 \cdot 500$, 9^2, $(-1) \cdot (-1)$, 1200^2.
(c) $500 - 800$, $-500 - 800$, $-500 - (-800)$.

11.10. Prove Theorem 11.10.

11.11. The open source `Python` programming language provides built-in commands to compute the integer quotient and remainder for integers of any size. If `a`, `n` are integers, then the useful commands and their effects are:

`a % n`	returns $\text{res}_n(a)$
`a // n`	returns the integer quotient of $a \div n$
`divmod(a,n)`	returns the pair (q, r) such that $a = qn + r$ and $0 \leq r < n$.

Note that the `divmod` command combines the effect of the other two. `Python` is installed by default on all Mac and Linux computers, but it must be downloaded (from `python.org`) and installed on a Windows computer. Once `Python` is installed on your computer, open up a command-line terminal and type `python` to start a `Python` interpreter session. Then you can type commands in order to do computations, similar to typing commands on a calculator. Hit Enter at the end of each command to get results.

Use `Python` to compute the integer quotient q and remainder r for the following problems:

(a) $1234567890987654321 \div 6090609$.
(b) $-1234567890987654321 \div 6090609$.
(c) $12345678909876543211234567890 \div 1234567890987654321$.

11.12. After reading Exercise 11.11, go to a computer and fire up a `Python` session in order to compute the following:

(a) $123456789 - 987654321$ in $\mathbb{Z}_{999750750}$.
(b) 123456789^2 in $\mathbb{Z}_{999750750}$.
(c) 123456789^3 in $\mathbb{Z}_{999750750}$.

Note: In `Python`, the symbol `**` is used instead of `^` for computing powers; i.e., to compute a^b you must type `a ** b` instead of `a ^ b`.

12 Commutative rings

Our main goal in this section is to prove that the modular system $\mathbb{Z}_n$ (along with its operations of addition and multiplication) forms a commutative ring.

We begin with the definition of commutative ring, which is based on the properties of the system $\mathbb{Z}$ of integers.

12.1 Definition. A *commutative ring* is a set R with two binary operations, addition $+$ and multiplication $\cdot$, such that for all a, b, c in the set R:

(a) additive associativity: $a + (b + c) = (a + b) + c$.
(b) additive commutativity: $a + b = b + a$.
(c) additive identity: There is some element $0 \in R$ such that $a + 0 = a = 0 + a$.
(d) additive inverse: for every $a \in R$, there exists $-a \in R$ such that $a + (-a) = 0 = (-a) + a$.
(e) multiplicative associativity: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
(f) multiplicative commutativity: $a \cdot b = b \cdot a$.
(g) multiplicative identity: There is an element $1 \in R$ such that $a \cdot 1 = a = 1 \cdot a$.
(h) distributivity: $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$.

The element 0 is called the *additive identity*, while 1 is the *multiplicative identity*. The element $-a$ is the *additive inverse* of a, or the *negative* of a.

If we remove axiom (f) from the definition then we get the definition of a *ring*. In general, multiplication in a ring is not required to be commutative. For now, all of our rings will be commutative.

The set of integers $\mathbb{Z}$ is a commutative ring under ordinary addition and multiplication. So are the sets of rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$, and complex numbers $\mathbb{C}$. Indeed, the ring axioms are modeled on the fundamental properties of the ordinary number systems.

12.2 Remark. We can subtract in any ring, because additive inverses exist, so we can define $a - b = a + (-b)$. Thus, we can view a commutative ring as a set of elements along with a way to add, subtract, and multiply them, subject to the familiar properties of number systems. But note that division may not always be possible in a ring. In $\mathbb{Z}$, we cannot divide 1 by 2, as the fraction $1/2$ is not an integer.

The set $\mathbb{N}$ of natural numbers under ordinary addition and multiplication is *not* a ring, because most elements of $\mathbb{N}$ do not have an additive inverse; i.e.,

axiom 1(d) fails. In other words, the set $\mathbb{N}$ is not a ring because subtraction doesn't make sense in $\mathbb{N}$.

Here are some formal consequences of the definition of ring. These properties hold in all rings, commutative or not.

12.3 Theorem. *Let R be a ring. For any $a, b, c \in R$ we have:*

(a) $a \cdot 0 = 0 = 0 \cdot a$.
(b) $a(-b) = -(ab) = (-a)b$.
(c) $(-a)(-b) = ab$.
(d) $a(b - c) = ab - ac$.
(e) $(-1)a = -a$.

Notice that we have omitted the symbol $\cdot$ in the products in parts (b)–(e) above. It is standard to omit the multiplication symbol in situations where it leads to no confusion.

The proof of these basic facts is an exercise. For instance, to prove (a) one would begin with the equality $0 + 0 = 0$, which comes from axiom (h) of Definition 12.1 by taking $a = 0$ there, then multiply both sides by a, and so on.

Theorem 12.10 below says that the set $\mathbb{Z}_n$ is a commutative ring. So modular arithmetic provides examples of finite commutative rings. The main goal of this section is to prove this important result.

To accomplish our goal we need to study another, more sophisticated, construction of $\mathbb{Z}_n$ in terms of equivalence classes. This approach gives more powerful information about the properties of $\mathbb{Z}_n$.

We start by saying a bit more about equivalence classes. The following important definition from set theory will be used again in the future.

Definition. Let A be a set. An *equivalence relation* on A is a relation $\sim$ on A which is reflexive ($a \sim a$, for all $a \in A$), symmetric ($a \sim b$ implies $b \sim a$, for all $a, b \in A$), and transitive ($a \sim b$ and $b \sim c$ implies $a \sim c$, for all $a, b, c \in A$).

It is a general fact in set theory that an equivalence relation $\sim$ on a set A always induces a partition of the set into disjoint equivalence classes:

$$A = \bigcup_{a \in A} [a]$$

where $\overline{a} = [a] = \{b \in A \mid a \sim b\}$ is the equivalence class of a. (We write $\overline{a}$, or $[a]$, for the equivalence class containing a. That class is the set of all

elements that are equivalent to a.) Another general fact is that

$$[a] = [b] \iff a \sim b.$$

Thus, there can be many different ways to write the same equivalence class. The element used to write a class is called a *representative* of the class.

We need the following general concept, which will be used again later.

Definition (quotient set)**.** Let $\sim$ be an equivalence relation on a set A, and write $[a]$ (or $\overline{a}$) for the equivalence class containing a. The set $A/\sim = \{[a] : a \in A\}$, the set of all equivalence classes, is called the *quotient* of A by $\sim$.

Now we return to the task of proving that $\mathbb{Z}_n$ is a ring. We need to recall some basic properties about the set $\mathbb{Z}$ of integers.

Recall that if a, b are integers then a *divides* b (written as $a \mid b$) if and only if there is some $k \in \mathbb{Z}$ such that $b = ak$. If a divides b we also say that b *is divisible by* a.

12.4 Definition (Gauss)**.** Let n be a fixed positive integer, and $a, b \in \mathbb{Z}$. We say that a is *congruent* to b modulo n if and only if $n \mid (a-b)$. We write $a \equiv b \pmod{n}$ to mean that a is congruent to b modulo n.

Here are the basic properties of congruences, all of which are routine to prove.

12.5 Theorem. *Let n be a fixed positive integer. Let a, b, c be integers. Then:*

(a) $a \equiv a \pmod{n}$.
(b) $a \equiv b \pmod{n}$ *implies* $b \equiv a \pmod{n}$.
(c) $a \equiv b \pmod{n}$ *and* $b \equiv c \pmod{n}$ *implies* $a \equiv c \pmod{n}$.

Theorem 12.5 says that congruence modulo n is an equivalence relation on the set $\mathbb{Z}$. More precisely, suppose we fix n and write $a \sim b$ if and only if $a \equiv b \pmod{n}$. Then $\sim$ is an equivalence relation on the set $\mathbb{Z}$.

The equivalence relation $\sim$ on the set $\mathbb{Z}$ of integers induces a partition of $\mathbb{Z}$ into disjoint equivalence classes:

$$\mathbb{Z} = \bigcup_{a \in \{0,1,\ldots,n-1\}} [a]$$

where the equivalence class $[a]$ is defined by

$$[a] = \{b \in \mathbb{Z} \mid a \sim b\} = \{b \in \mathbb{Z} \mid a \equiv b \pmod{n}\}.$$

The equivalence class $[a]$ of a is the collection of all integers which are congruent mod n to a. We say that a is a *representative* of its equivalence class $[a]$. Note that a given equivalence class has infinitely many representatives, because

$$[a] = [b] \iff a \sim b.$$

For example, if $n = 12$ then $[2] = [14] = [-10]$ because $2 \equiv 14 \equiv -10 \pmod{12}$; i.e., $2 \sim 14 \sim -10$.

12.6 Definition. Equivalence classes for the equivalence relation $\sim$ defined by congruence modulo n are called *congruence classes*.

The following gives a nice characterization of the numbers in a congruence class. The proof is an exercise.

12.7 Theorem. *Let $a \in \mathbb{Z}$ and let $r = \mathrm{res}_n(a)$. Then the congruence class $[a]$ is the collection of all integers of the form $kn + r$ where $k \in \mathbb{Z}$.*

12.8 Definition. We define addition and multiplication of congruence classes (for some fixed modulus n) by the rules:

$$[a] + [b] = [a + b], \qquad [a] \cdot [b] = [ab].$$

The following easy to prove result means that the definition makes sense; i.e., addition and multiplication of congruence classes are well-defined.

12.9 Theorem. *Let $a, b, c, d \in \mathbb{Z}$. Suppose that $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$; i.e., $a \sim b$ and $c \sim d$ Then:*

(a) $a + c \equiv b + d \pmod{n}$; i.e., $a + c \sim b + d$.

(b) $ac \equiv bd \pmod{n}$; i.e., $ac \sim bd$.

Applying the quotient construction to the set $\mathbb{Z}$ of integers with the equivalence relation $\sim$ defined by congruence modulo n, we obtain:

$$\mathbb{Z}/\sim \; = \{[a] : a \in \mathbb{Z}\} = \{[a] : a = 0, 1, \ldots, n-1\}.$$

Note that the distinct classes in $\mathbb{Z}/\sim$ are those listed in the rightmost set above, because of the aforementioned fact that $[a] = [b] \iff a \sim b$.

Thus there is a bijection from the set $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ onto the set $\mathbb{Z}/\sim$, defined by $a \mapsto [a]$. Furthermore, it can be shown that this bijection preserves addition and multiplication, in the sense that:

$$a \oplus b \mapsto [a] + [b]; \qquad a \odot b \mapsto [a] \cdot [b]$$

for all a, b. (Here we have momentarily reverted to the circle notation, in order to avoid confusion.) This proves that

$$\mathbb{Z}_n \cong \mathbb{Z}/\!\sim$$

i.e., the structure $\mathbb{Z}_n$ is isomorphic to the quotient $\mathbb{Z}/\!\sim$, where $\sim$ is the equivalence relation defined by congruence modulo n.

The quotient construction makes it easier to prove the following result.

12.10 Theorem. *Fix some integer $n > 1$, and let $\sim$ be the equivalence relation defined by congruence modulo n. Arithmetic in the quotient set*

$$\mathbb{Z}/\!\sim \;= \{[a] : a = 0, 1, \ldots, n-1\}$$

satisfies the following properties, holding for all $a, b, c \in \mathbb{Z}$:

(a) $[a] + ([b + [c]]) = ([a] + [b]) + [c]$.
(b) $[a] + [b] = [b] + [a]$.
(c) $[a] + [0] = [a] = [0] + [a]$.
(d) $[a] + [-a] = [0] = [-a] + [a]$.
(e) $[a] \cdot ([b \cdot [c]]) = ([a] \cdot [b]) \cdot [c]$.
(f) $[a] \cdot [b] = [b] \cdot [a]$.
(g) $[a] \cdot [1] = [a] = [1] \cdot [a]$.
(h) $[a] \cdot ([b] + [c]) = [a] \cdot [b] + [a] \cdot [c]$ *and* $([b] + [c]) \cdot [a] = [b] \cdot [a] + [c] \cdot [a]$.

In other words, $\mathbb{Z}/\!\sim$ is a commutative ring (and so is its isomorphic cousin $\mathbb{Z}_n$).

Proof. The proof uses the fact that all of these properties correspond to similar properties of ordinary integers. For instance, to prove (b) just use the fact that addition in $\mathbb{Z}$ is commutative. Then

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

where the middle equality comes from the fact that $a + b = b + a$ for integers a, b. This proves (b). All the other properties are proved similarly. □

12.11 Remarks. 1. The easy proof of the above theorem is a direct consequence of the power of equivalence classes. Working with equivalence classes takes some getting used to, but the effort pays dividends, in that it makes proofs easier.

2. In the construction of $\mathbb{Z}_n$, the elements are the residues $0, 1, \ldots, n-1$ and the operations $\oplus, \odot$ are defined in a special way. In the construction

of $\mathbb{Z}/\sim$, the elements are the congruence classes $[0], [1], \ldots, [n-1]$ and the operations $+, \cdot$ are defined in terms of the usual addition and multiplication of integers. The two constructions turn out to be equivalent, thus providing two perspectives on $\mathbb{Z}_n$.

3. In both constructions of $\mathbb{Z}_n$, we have chosen the set $\{0, 1, \ldots, n-1\}$ of residues modulo n as a complete set of representatives of the quotient set. But it is often useful to work instead with a different complete set of representatives:

$$\mathbb{Z}_n = \begin{cases} \{0, \pm 1, \ldots, \pm(N-1), N\}, & \text{if } n = 2N \\ \{0, \pm 1, \ldots, \pm N\}, & \text{if } n = 2N+1 \end{cases}$$

where N is an integer. For example, we can write $\mathbb{Z}_7 = \{0, \pm 1, \pm 2, \pm 3\}$.

Exercises

12.1. Prove that $a \equiv b \pmod{n}$ if and only if a and b have the same integer residue (remainder) when divided by n.

12.2. Prove Theorem 12.3.

12.3. Define a relation $\approx$ on the set P of all living people by $a \approx b$ if the age of a (in years) is equal to the age of b.

(a) Prove that $\approx$ is an equivalence relation on P.
(b) If Sally is age 18 years, then describe the class [Sally] in words.
(c) Describe the quotient set $P/\approx$ in words.
(d) Is $P/\approx$ finite or infinite?

12.4. Let $A = \{a, b, c, \ldots, x, y, z\}$ be the usual English alphabet. For the purpose of this problem, we define a *word* over A to be any string of one or more letters from the alphabet. For example, *bab*, *abba*, and *dogfrog* are words according to our definition. Define a relation $\approx$ on the set W of words by $\alpha \approx \beta$ if and only if the first letter of α equals the first letter of β (where $\alpha, \beta \in W$).

(a) Prove that $\approx$ is an equivalence relation on W.
(b) Describe the quotient set $W/\approx$ in words. Find a nice set of representatives of the classes in $W/\approx$.
(c) Compute $|W/\approx|$.

12.5. Define a relation $\approx$ on the set $\mathbb{R}$ of real numbers by $a \approx b \iff a - b \in \mathbb{Z}$.

(a) Prove that $\approx$ is an equivalence relation on $\mathbb{R}$.
(b) Compute the class $[1/2]$ of $1/2$. What is its cardinality?
(c) Compute the class $[3/2]$ of $3/2$. How is $[3/2]$ related to $[1/2]$?
(d) Describe the quotient set $\mathbb{R}/\approx$ and find a complete set of representatives for the quotient set elements.

(e) Is $\mathbb{R}/\approx$ finite or infinite?

12.6. Prove Theorem 12.5.

12.7. Prove Theorem 12.7.

12.8. Prove Theorem 12.9.

12.9. Prove parts (a), (h) of Theorem 12.10.

13 Fields

The main goal of this section is to show that the ring $\mathbb{Z}_n$ is a field if and only if the modulus n is a prime number.

13.1 Definition. A *multiplicative inverse* of a is an element a^{-1} such that $aa^{-1} = 1 = a^{-1}a$. An element a in a ring R is said to be a *unit* (or *invertible element*) if there is a multiplicative inverse of a in the ring.

If a multiplicative inverse exists in a ring then it is unique. Every nonzero element is a unit in the rings $\mathbb{Q}, \mathbb{R}$, and $\mathbb{C}$. The only units in the ring $\mathbb{Z}$ are ± 1.

13.2 Definition. A *field* is a commutative ring in which $1 \neq 0$ and every nonzero element is a unit.

Thus, $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields. But $\mathbb{Z}$ is *not* a field.

In a field, the existence of multiplicative inverse means that we can define division, by:

$$b/a = b \cdot a^{-1} \quad (a \neq 0).$$

So we can think of a field as a system in which the usual operations of addition, multiplication, subtraction, and division makes sense and obey the usual laws of algebra.

The central question for this section is: for which values of n is the ring $\mathbb{Z}_n$ a field? To study this question, we look at a more general question: what is the set of units in $\mathbb{Z}_n$?

13.3 Definition. The set of all units in a ring R will be denoted by $R^\times$ (or R^*). This is called the *multiplicative group of units* in the ring.

The set $R^\times$ provides another example of a group. In terms of this notation, the definition of a field can be rephrased as: a commutative ring R is a field if and only if $R^\times = R - \{0\}$.

Recall that the notation $\gcd(a, b)$ stands for the *greatest common divisor* of integers a, b. Furthermore, recall that $\gcd(a, b)$ can be calculated using the *Euclidean algorithm*. Finally, recall that the *extended* Euclidean algorithm produces integers x, y such that

$$ax + by = \gcd(a, b).$$

It is often said that this equation expresses the gcd as a *linear combination* of a, b. Using these properties of the gcd, we can prove the following important result.

13.4 Theorem. *Let n be a positive integer. A congruence class $[a] \in \mathbb{Z}_n$ is a unit if and only if* $\gcd(a, n) = 1$. *Hence, the set $\mathbb{Z}_n^\times$ of units in $\mathbb{Z}_n$ is equal to* $\{[a] \in \mathbb{Z}_n : \gcd(a, n) = 1\}$.

Proof. Suppose that $\gcd(a, n) = 1$. Use the Euclidean algorithm to find integers x, y such that $ax + ny = 1$. This can be done because $\gcd(a, n) = 1$ by hypothesis. Then $ax - 1 = ny$, so $n \mid (ax - 1)$ and thus $ax \equiv 1 \pmod{n}$. This means that $[a] \cdot [x] = [1]$ in $\mathbb{Z}_n$. So $[x] = [a]^{-1}$ and $[a]$ is a unit.

On the other hand, if $\gcd(a, n) = g > 1$, then I claim that $[a]$ is not a unit. To see this, assume it is a unit. Then $[a]^{-1}$ exists in $\mathbb{Z}_n$ and $[a] \cdot [a]^{-1} = [1]$. But the condition $\gcd(a, n) = g > 1$ means that $g \mid a$ and $g \mid n$, so $a = a_1 g$ and $n = n_1 g$ for some integers a_1, n_1. Now

$$[a] \cdot [n_1] = [an_1] = [a_1 g n_1] = [a_1 n] = [0].$$

If we multiply the equation $[a] \cdot [n_1] = [0]$ by $[a]^{-1}$ we obtain $[a]^{-1} \cdot [a] \cdot [n_1] = [0]$, i.e., $[n_1] = [0]$. But this is a contradiction, since the equation $n = n_1 g$ implies that $0 < n_1 < n$. Thus our assumption must be incorrect, and $[a]$ is not a unit. □

13.5 Corollary. *Let n be a positive integer. Then $\mathbb{Z}_n$ is a field if and only if n is a prime number.*

Proof. Suppose that n is prime. Then for all $[a] \neq [0]$ we have $\gcd(a, n) = 1$. Each such $[a]$ is a unit in $\mathbb{Z}_n$ by the theorem. So $\mathbb{Z}_n$ is a field.

Suppose that n is not prime. Then $n = ab$ for some integers $1 < a, b < n$. Then $\gcd(a, n) = a$, so $[a]$ is not a unit in $\mathbb{Z}_n$. This is a nonzero element of $\mathbb{Z}_n$ which is not a unit, so $\mathbb{Z}_n$ is not a field. □

13.6 Definition (notation)**.** From now on, we write $\mathbb{F}_p = \mathbb{Z}_p$ for the field of p elements, when p is a prime number. An alternative notation for $\mathbb{F}_p$ is $\mathrm{GF}(p)$. In this context, GF stands for *Galois field*. The finite fields $\mathbb{F}_p$ of p elements are called Galois fields.

It should be noted that there are other finite fields besides the ones we have constructed. We leave that for later.

13.7 Examples. $\mathbb{Z}_4^\times = \{[1], [3]\}$, $\mathbb{Z}_7^\times = \{[1], [2], [3], [4], [5], [6]\}$, and $\mathbb{Z}_{15}^\times = \{[1], [2], [4], [7], [8], [11], [13], [14]\}$.

Exercises

13.1. Prove that a commutative ring R is a field if and only if $R^\times = R - \{0\}$.

13.2. There is one and only one pair (a, b) of integers for which $\gcd(a, b)$ is not defined. What pair is it?

13.3. Compute the set $\mathbb{Z}_n^\times$ of units for the following cases:

(a) $n = 2$. (c) $n = 5$. (e) $n = 8$.
(b) $n = 3$. (d) $n = 6$. (f) $n = 9$.

13.4. Use guess-and-check (trial and error) to solve the equation $3x = 4$ in $\mathbb{Z}_{10}$, if possible. (When the modulus is small, there are not many cases to try.)

13.5. Use guess-and-check (trial and error) to solve the equation $4x = 5$ in $\mathbb{Z}_{10}$, if possible.

13.6. For which of the following values of n is the ring $\mathbb{Z}_n$ a field:

(a) $n = 2$. (c) $n = 5$. (e) $n = 8$.
(b) $n = 3$. (d) $n = 6$. (f) $n = 9$.

13.7. Use guess-and-check (trial and error) to calculate the following inverses:

(a) $[2]^{-1}$ in $\mathbb{Z}_9$. (b) $[3]^{-1}$ in $\mathbb{Z}_{10}$. (c) $[4]^{-1}$ in $\mathbb{Z}_{15}$.

13.8. Prove that the equation $ax = b$ is solvable in $\mathbb{Z}_n$ whenever $a \in \mathbb{Z}_n$ is a unit. Explain exactly *how* to solve the equation.

13.9. The Euclidean algorithm, which was described in the last book of Euclid's *Elements*, works as follows, for any given pair of integers a, b such that not both are zero:

1) If $b < 0$ replace b by $|b|$. Do the same for a.
2) If $b = 0$ then the gcd is a, so return a and stop.
3) If $b \neq 0$, compute the unique integers q, r such that $a = qb + r$ and $0 \leq r < b$. Then replace the pair (a, b) by the pair (b, r). Go to Step 2.

Use the Euclidean algorithm to compute the following:

(a) $\gcd(87031, 4750)$. (b) $\gcd(48157656, 541541)$.

13.10. Refer to the previous problem for a description of the Euclidean algorithm.

(a) Prove that whenever the pair (a, b) of integers is not $(0, 0)$ then the Euclidean algorithm must stop after finitely many steps.
(b) Explain what happens in the first stage of the algorithm if $0 < a < b$. What are the next values of (a, b)?

13.11. The extended Euclidean algorithm returns a triple $\text{xgcd}(a, b) = (g, x, y)$ of integers such that $g = \gcd(a, b)$ and $g = ax + by$. We assume that both a, b are given positive integers. Then the algorithm works as follows:

1) Set $(x_0, y_0) = (1, 0)$. Set $(x, y) = (0, 1)$.
2) If $b = 0$, return (a, x_0, y_0).
3) If $b > 0$, let q, r be the unique integers such that $a = qb + r$ and $0 \leq r < b$. Replace (a, b) by (b, r).
4) Replace (x, y) by $(x_0 - qx, y_0 - qy)$. Replace (x_0, y_0) by (x, y). Go to Step 2.

Use the extended Euclidean algorithm to compute the following:

(a) $\text{xgcd}(23, 301)$. (b) $\text{xgcd}(87031, 4750)$. (c) $\text{xgcd}(48157656, 541541)$.

13.12. Use the results of the previous problem to calculate the following inverses:

(a) $[23]^{-1}$ in $\mathbb{Z}_{301}$.
(b) $[4750]^{-1}$ in $\mathbb{Z}_{87031}$.
(c) $[541541]^{-1}$ in $\mathbb{Z}_{48157656}$.

Chapter 4

Linear Groups

14 Matrix groups

Now we investigate groups formed by sets of matrices. These groups are often infinite sets, so we will now be considering infinite groups, in contrast to the finite groups we have seen so far. Initially, all of our matrices will have real number entries.

14.1 Definition. A *group of matrices* (or *matrix group* for short) is any nonempty set G of $n \times n$ nonsingular matrices which is:

(a) closed under products: for all $A, B \in G$, the product $AB \in G$.
(b) closed under inverses: for all $A \in G$, the inverse $A^{-1} \in G$.

We only consider square $(n \times n)$ matrices. We have to restrict to nonsingular matrices because we want our matrices to have inverses. Recall that a basic theorem of linear algebra states that *a matrix is invertible if and only if it is nonsingular.* Recall also that *a matrix is nonsingular if and only if its determinant is nonzero.*

14.2 Examples. Here are some important examples of matrix groups.

1. The *general linear group* $\mathrm{GL}(n)$ is the group consisting of all $n \times n$ nonsingular matrices. In symbols,

$$\mathrm{GL}(n) = \{n \times n \text{ matrices } A : \det A \neq 0\}.$$

The group $\mathrm{GL}(n)$ is infinite (the cardinality of the set is infinite). Note that $\mathrm{GL}(1)$ can be identified with the set $\mathbb{R}^\times$ of units in $\mathbb{R}$, because the invertible 1×1 matrices are all of the form $[a]$ for $a \neq 0$.

2. The *special linear group* $\mathrm{SL}(n)$ is the group consisting of all $n \times n$ matrices of determinant equal to 1. In symbols,

$$\mathrm{SL}(n) = \{A \in \mathrm{GL}(n) : \det A = 1\}.$$

By definition, we have an inclusion $\mathrm{SL}(n) \subset \mathrm{GL}(n)$. It is an exercise to verify that this is a matrix group. Note that $\mathrm{SL}(1)$ is actually a finite group, even though $\mathrm{SL}(n)$ is infinite, for all $n > 1$.

3. The *orthogonal group* $\mathrm{O}(n)$ is the group consisting of all $n \times n$ orthogonal matrices. A matrix A is said to be an *orthogonal* matrix if its inverse is equal to its transpose: $A^{-1} = A^{\mathsf{T}}$. So, in symbols:

$$\mathrm{O}(n) = \{A \in \mathrm{GL}(n) : A^{-1} = A^{\mathsf{T}}\}.$$

It is an exercise to verify that this is a matrix group. Note that $|\mathrm{O}(1)| = 2$. For all $n \geq 2$, the group $\mathrm{O}(n)$ is infinite.

4. The *special orthogonal group* $\mathrm{SO}(n)$ is the group of all $n \times n$ orthogonal matrices of determinant equal to 1. In symbols,

$$\mathrm{SO}(n) = \{A \in \mathrm{O}(n) : \det A = 1\}.$$

An orthogonal matrix of determinant 1 is also known as a *proper* orthogonal matrix. So we can rephrase the definition to say: $\mathrm{SO}(n)$ is the group of all proper orthogonal $n \times n$ matrices. Note that $\mathrm{SO}(1) = \mathrm{SL}(1)$ has a single element, but $\mathrm{SO}(n)$ is infinite for all $n \geq 2$.

14.3 Remark. Matrix groups are also called *linear groups*. This is due to the fact that square matrices represent linear operators. If A is an $n \times n$ matrix, then the function $\alpha : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\alpha(X) = AX$ is a linear operator. So every matrix group is isomorphic to a group of linear operators.

We will use some basic linear algebra to find a nice set of generators for the group $\mathrm{GL}(n)$. Recall that *elementary row operations* are used to solve systems of linear equations (by Gaussian elimination). The elementary row operations on a matrix are:

(i) add t times row j to row i, for any $i \neq j$, any $t \in \mathbb{R}$,
(ii) multiply row i by a scalar $t \in \mathbb{R}$, for any i, any $t \neq 0$,
(iii) swap row i and row j, for any $i \neq j$.

If A is a matrix, then the elementary row operations on A are equivalent to left multiplication of A by the appropriate corresponding elementary matrix. The elementary matrices are denoted by:

(i) $E_{ij}(t)$, for any $i \neq j$, any $t \in \mathbb{R}$,
(ii) $M_i(t)$, for any i, any $t \neq 0$,
(iii) P_{ij}, for any $i \neq j$.

By definition, the elementary matrices are precisely those matrices that are obtained from an identity matrix by performing a single elementary row operation of the corresponding type. This rule defines a bijection from elementary row operations onto elementary matrices.

14.4 Theorem. *If A is any $n \times n$ matrix and B is the matrix resulting from performing a single elementary row operation to A then $B = UA$, where U is the corresponding elementary matrix of the same type.*

This simple result is proved in most linear algebra textbooks. The proof is an easy calculation. The theorem has the following pleasant consequence.

14.5 Corollary. *Any nonsingular $n \times n$ matrix A can be expressed as a product of elementary matrices.*

Proof. The proof is constructive: it tells us not only that the desired factorization exists, it also gives an algorithm that we may use to find one.

Let A be a nonsingular $n \times n$ matrix. Then the reduced echelon form of A is I, so by applying Gaussian elimination to A we can row reduce A to I. This means that there is a finite sequence of elementary row operations that transform A to I. Let $U_1, U_2, \ldots, U_k$ be the elementary matrices corresponding to the elementary row operations, in order. Then by the theorem, we have

$$I = (U_k U_{k-1} \cdots U_2 U_1)A.$$

It follows by matrix algebra that $A = U_1^{-1} U_2^{-1} \cdots U_{k-1}^{-1} U_k^{-1}$. Since the inverse of any elementary matrix is another elementary matrix of the same type, we are finished. □

The corollary gives us a nice set of generators for the general linear group $\mathrm{GL}(n)$. Before we formulate the result, we make a formal definition.

14.6 Definition. In general, we say that a group G is *generated* by a set $S \subset G$ of its elements if every element of G is expressible as a product of elements of S and inverses of elements of S.

Now here is the result.

14.7 Theorem. *The group $\mathrm{GL}(n)$ is generated by the set of $n \times n$ elementary matrices.*

Proof. This follows immediately from the preceding corollary, which states that any matrix in $\mathrm{GL}(n)$ is expressible as a product of elementary matrices. □

14.8 Remark. In general, to *understand* a group, it is desirable to find a nice subset of generators. Finding a set of generators reduces many questions about the group to questions about its generators.

We have previously proved that the group $\mathbb{S}_n$ is generated by its transpositions. This is a case in point: many questions about permutations reduce to a question about transpositions. So too for the matrix group $\mathrm{GL}(n)$: many questions about nonsingular matrices reduce to a question about elementary matrices.

It should be emphasized that *generating sets are not unique*. There are many different sets generating $\mathrm{GL}(n)$; the same is true of $\mathbb{S}_n$.

Matrix groups are symmetry groups

The groups introduced in this section are also symmetry groups. To see this, recall that an $n \times n$ matrix A represents the linear operator $\alpha : \mathbb{R}^n \to \mathbb{R}^n$ defined by the rule $\alpha(X) = AX$. Nonsingular matrices represent linear *automorphisms*, the bijective linear operators. So

$$\mathrm{GL}(n) = \text{ the group of linear automorphisms of } \mathbb{R}^n.$$

This is the symmetry group of the vector space $\mathbb{R}^n$, because the linear automorphisms preserve the vector space structure.

The special linear group $\mathrm{SL}(n)$ is the group of linear automorphisms of $\mathbb{R}^n$ preserving volume and orientation, in an appropriate sense. To understand this, recall the following fact from multivariable calculus: given three column vectors $P = (p_1, p_2, p_3)$, $Q = (q_1, q_2, q_3)$, $R = (r_1, r_2, r_3) \in \mathbb{R}^3$, the absolute value of the determinant of the 3×3 matrix $M = [P|Q|R]$ they form gives the volume of the parallelopiped they generate, and the sign of the determinant determines its orientation in some appropriate sense. Then a linear operator $\alpha : \mathbb{R}^n \to \mathbb{R}^n$, given by the rule $\alpha(X) = AX$ with A a 3×3 matrix, is volume and orientation preserving if and only if the determinant of M remains unchanged when we replace the vectors P, Q, R by $\alpha(P), \alpha(Q), \alpha(R)$. Equivalently, α preserves volume and orientation if and only if $\det M = \det AM$, where A represents α. This is equivalent to $\det A = 1$. That explains the idea for $\mathbb{R}^3$, and it turns out that this can be (with a fair amount of work) extended to $\mathbb{R}^n$ for any n.

The orthogonal group $\mathrm{O}(n)$ is the group of linear automorphisms of $\mathbb{R}^n$ preserving the usual dot product. Since dot product determines length and angle, we can also say that $\mathrm{O}(n)$ is the group of linear automorphisms of $\mathbb{R}^n$ preserving length and angle, but it is simpler to focus on the dot product. A linear operator $\alpha : \mathbb{R}^n \to \mathbb{R}^n$ is given by the rule $\alpha(X) = AX$, where A is a matrix. If for any pair X, Y of vectors in $\mathbb{R}^n$ we have

$$\alpha(X) \cdot \alpha(Y) = X \cdot Y$$

then we say that the operator α preserves the dot product. Recall that the dot product of two column vectors may also be written as a matrix product: $X \cdot Y = X^{\mathsf{T}}Y$. Thus, the above dot product equality is equivalent to the following:

$$(AX)^{\mathsf{T}}(AY) = X^{\mathsf{T}}Y \iff X^{\mathsf{T}}A^{\mathsf{T}}AY = X^{\mathsf{T}}Y.$$

Since this condition has to hold for all $X, Y \in \mathbb{R}^n$, it follows that it holds if and only if $A^{\mathsf{T}}A = I$, which is equivalent to the condition $A^{-1} = A^{\mathsf{T}}$. So α preserves dot product if and only if $A^{-1} = A^{\mathsf{T}}$.

From the last paragraph it is not hard to see that $\mathrm{SO}(n)$ is the group of linear automorphisms of $\mathbb{R}^n$ which preserve volume, orientation, and dot product (distance and angle).

We can only hint at the many beautiful and profound connections between matrix groups and geometry. There are many books devoted entirely to one or more aspects of this, so it is impossible to be comprehensive here.

Exercises

14.1. Recall that the determinant of a product of matrices equals the product of their determinants. Use this to prove that:
   (a) $\mathrm{GL}(n)$ is a matrix group.
   (b) $\mathrm{SL}(n)$ is a matrix group.

14.2. Prove that any matrix group G must contain the identity matrix I.

14.3. Describe all the elements of $\mathrm{SL}(1)$. What is the order $|\mathrm{SL}(n)|$ of $\mathrm{SL}(n)$? Justify your answers.

14.4. Recall that the inverse of a product of two matrices is equal to the product of their inverses in reverse order: $(AB)^{-1} = B^{-1}A^{-1}$. The same is true of the transpose: $(AB)^{\mathsf{T}} = B^{\mathsf{T}}A^{\mathsf{T}}$. Use these facts to prove that $\mathrm{O}(n)$ is a matrix group.

14.5. Prove that $|\mathrm{O}(1)| = 2$; i.e., $\mathrm{O}(1)$ is a group of order 2.

14.6. Prove that $\mathrm{O}(n) = \{A \in \mathrm{GL}(n) : AA^{\mathsf{T}} = I\}$. (Don't make the error of assuming that matrix multiplication is commutative.)

14.7. Suppose that an $n \times n$ orthogonal matrix $A = [A_1|A_2|\cdots|A_n]$ is regarded as a matrix of column vectors.

(a) Show that $A_i^{\mathsf{T}} A_j = \delta_{ij}$. (Here δ_{ij} is the *Kronecker delta* symbol, defined to be 1 if $i = j$ and 0 otherwise.)

(b) Show that the dot product $A_i \cdot A_j = \delta_{ij}$.

(c) Deduce that if $i \neq j$ then $A_i \perp A_j$.

(d) Deduce that $|A_i| = 1$, i.e., A_i is a unit vector for all i.

(e) Deduce that the columns of an $n \times n$ orthogonal matrix form an orthonormal basis of $\mathbb{R}^n$.

14.8. Prove that $\mathrm{SO}(n)$ is a matrix group.

14.9. Prove that $\mathrm{SO}(1) = \mathrm{SL}(1)$ is a group of one element.

14.10. Prove that $\mathrm{SO}(n) = \mathrm{SL}(n) \cap \mathrm{O}(n)$. [Hint: Show each side is a contained in the other.]

14.11. Prove that if $G, H \subset \mathrm{GL}(n)$ are any two matrix groups (consisting each of $n \times n$ matrices) then $G \cap H$ is another matrix group.

14.12. (a) Prove that if $A \in \mathrm{O}(n)$ then $\det A = \pm 1$.

(b) The matrices $A \in \mathrm{O}(n)$ of determinant -1 are called *improper* orthogonal matrices. Is the set of improper orthogonal matrices a matrix group? Prove your answer.

14.13. Let $D(n)$ be the set of all diagonal matrices in $\mathrm{GL}(n)$. Show that $D(n)$ is a matrix group.

14.14. (a) Prove that $E_{ij}(t) \in \mathrm{SL}(n)$, for any $i \neq j$, any $t \in \mathbb{R}$.

(b) Let $D_1(n)$ be the set of diagonal matrices in $\mathrm{SL}(n)$. Show that $D_1(n)$ is a matrix group.

(c) $(*)$ Prove that $\mathrm{SL}(n)$ is generated by the set $S = D_1(n) \cup \{E_{ij}(t) : i \neq j,\ t \in \mathbb{R}\}$. [Hint: Argue that it is possible to row reduce any matrix $A \in \mathrm{SL}(n)$ to a diagonal matrix only using type (i) elementary row operations.]

15 The group of rotations of the plane

In this section we will look more closely at a special example, namely the group SO(2). This is the group of all 2×2 orthogonal matrices of determinant 1.

What can be said about the matrices in SO(2)? Let's figure out the answer, which involves a calculation. Suppose that

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SO}(2).$$

Then we know that $A^{-1} = A^{\mathsf{T}}$ and $\det A = 1$. The first condition implies that $I = A^{\mathsf{T}}A$, which was obtained by right multiplication by A. So we know that

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

Equivalently, this says that

$$\begin{bmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

In other words, the column vectors $A_1 = \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$, $A_2 = \left[\begin{smallmatrix} b \\ d \end{smallmatrix}\right]$ in the columns of A must have unit length, so they lie somewhere on the unit circle. Also, the dot product $A_1 \cdot A_2 = ab + cd = 0$, so the vectors A_1, A_2 are perpendicular. Since $A_1 = \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ is on the unit circle $x^2 + y^2 = 1$, there must be some angle θ such that $A_1 = \left[\begin{smallmatrix} \cos\theta \\ \sin\theta \end{smallmatrix}\right]$. Since $A_1 \perp A_2$, the angle between A_1 and A_2 is $\pi/2$, so $A_2 = \left[\begin{smallmatrix} \cos(\theta+\pi/2) \\ \sin(\theta+\pi/2) \end{smallmatrix}\right] = \left[\begin{smallmatrix} -\sin\theta \\ \cos\theta \end{smallmatrix}\right]$. This proves that $A \in \mathrm{SO}(2)$ must be of the form

$$A = R_\theta := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

for some real number θ. Conversely, it is easy to check that any matrix of the above form is in SO(2). This proves the following result.

15.1 Theorem. *The group* SO(2) *of all proper orthogonal* 2×2 *matrices is the group of all matrices of the form* R_θ*, as defined above, where* $\theta \in \mathbb{R}$*.*

Because the trigonometric functions are periodic, with period 2π, it follows that $R_\theta = R_{\theta+2\pi}$. So in fact the set SO(2) can be written more compactly as

$$\mathrm{SO}(2) = \{R_\theta : \theta \in [0, 2\pi)\}.$$

Regarded as linear operators on $\mathbb{R}^2$, a matrix $R_\theta \in \mathrm{SO}(2)$ defines the function $\rho_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ given by the rule $\rho_\theta(X) = R_\theta X$.

15.2 Theorem. *The linear operator ρ_θ defined by $\rho_\theta(X) = R_\theta X$ is a rotation of $\mathbb{R}^2$ through θ radians, in the sense that if $X \in \mathbb{R}^2$ is regarded as a vector, then $Y = \rho_\theta(X)$ is the vector obtained by rotating X through an angle of θ radians.*

Proof. Observe that the operation rot_θ of rotating the plane by θ radians is a *linear* operator on $\mathbb{R}^2$: $\mathrm{rot}_\theta(X_1+X_2) = \mathrm{rot}_\theta(X_1)+\mathrm{rot}_\theta(X_2)$ and $\mathrm{rot}_\theta(cX) = c\,\mathrm{rot}_\theta(X)$ for all $c \in \mathbb{R}$, $X, X_1, X_2 \in \mathbb{R}^2$. Of course, the function ρ_θ is also a linear operator. In general, to prove equality of functions f, g, one must show that $f(x) = g(x)$ for all x. But for linear transformations, showing equality is easier, because it suffices to prove they agree on a basis of the domain. So we only need to check that

$$\rho_\theta(\hat{\imath}) = \mathrm{rot}_\theta(\hat{\imath}), \quad \rho_\theta(\hat{\jmath}) = \mathrm{rot}_\theta(\hat{\jmath})$$

where $\{\hat{\imath}, \hat{\jmath}\} = \{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\}$ is the standard basis of $\mathbb{R}^2$. This verification is an exercise, and it completes the proof. □

15.3 Corollary. *We have $\rho_{\theta_1}\rho_{\theta_2} = \rho_{\theta_1+\theta_2}$ and $(\rho_\theta)^{-1} = \rho_{-\theta}$. In particular, rotations commute: $\rho_{\theta_1}\rho_{\theta_2} = \rho_{\theta_2}\rho_{\theta_1}$ for all $\theta_1, \theta_2 \in \mathbb{R}$.*

Proof. This is clear from the fact that ρ_θ is a rotation. So rotating by θ_2 radians followed by rotating by θ_1 radians is the same as rotating by $\theta_1 + \theta_2$ radians, etc. □

This corollary immediately implies that we have similar relations on the rotation matrices:

$$R_{\theta_1}R_{\theta_2} = R_{\theta_1+\theta_2} \quad \text{and} \quad (R_\theta)^{-1} = R_{-\theta}.$$

In particular, rotation matrices commute: $R_{\theta_1}R_{\theta_2} = R_{\theta_2}R_{\theta_1}$ for all $\theta_1, \theta_2 \in \mathbb{R}$.

15.4 Corollary. *The matrix group* $\mathrm{SO}(2)$ *is isomorphic to the group $\{\rho_\theta : \theta \in \mathbb{R}\}$ of rotations of the euclidean plane $\mathbb{R}^2$.*

Proof. The isomorphism is given by $f(R_\theta) = \rho_\theta$. □

Now we consider *improper* orthogonal 2×2 matrices. By definition, these are the matrices $A \in \mathrm{O}(2)$ of determinant equal to -1. One such matrix is the matrix

$$H_0 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The corresponding linear operator on $\mathbb{R}^2$ is defined by $X \mapsto H_0X$; i.e., $(x, y) \mapsto (x, -y)$. Geometrically, this is the operator of reflection across the horizontal axis. Now we can describe all the improper orthogonal matrices in O(2); it turns out that they are all reflections.

15.5 Theorem. *(a) For any improper orthogonal 2×2 matrix H, the matrix product H_0H is a proper orthogonal matrix, i.e., $H_0H \in \mathrm{SO}(2)$. The same holds for HH_0.*

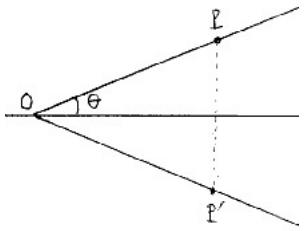
(b) Reflection refl_θ across the line through the origin at angle θ with the horizontal axis of $\mathbb{R}^2$ is a linear operator on $\mathbb{R}^2$, and its matrix H_θ is an improper orthogonal matrix.

(c) $H_\theta = H_0R_{-2\theta}$.

Proof. (a) Recall that for square matrices, the determinant of a product equals the product of the determinants. Thus $\det(H_0H) = \det(H_0)\,\det(H) = (-1)(-1) = 1$. So $H_0H \in \mathrm{SO}(2)$, as required. The other case is similar.

(b) It is easy to check that reflection is a linear operator. So we can represent refl_θ by some 2×2 matrix H_θ with respect to the standard basis of $\mathbb{R}^2$, so that $\mathrm{refl}_\theta(X) = H_\theta X$, for all $X \in \mathbb{R}^2$. Since the operator refl_θ preserves length of, and angles between, vectors, it is clear that $H_\theta \in \mathrm{O}(2)$. Furthermore, $\det(H_\theta) = -1$ because H_θ cannot be a rotation. So H_θ is an improper orthogonal matrix. By part (a), it follows that $H_0H_\theta \in \mathrm{SO}(2)$; i.e., it is equal to some rotation matrix.

(c) It suffices to check that $H_0H_\theta = R_{-2\theta}$, because once we have that equation we can left multiply by H_0 (using $H_0^2 = I$) to obtain the result. Since we know that H_0H_θ is some rotation matrix, it is enough to look at what it does to a single chosen (nonzero) vector. Choose a point $P \neq 0$ on the line fixed by refl_θ. Then $H_\theta P = P$ and $H_0H_\theta P = H_0P = P'$, where P' is the reflection of P across the horizontal axis. In the picture,



the angle $\angle POP'$ (where O is the origin) is an angle of -2θ radians, so the effect of left multiplying by H_0H_θ is the same on P as the effect of rotating P by -2θ radians. So $H_0H_\theta = R_{-2\theta}$, and we are done. □

In summary, we now have a complete understanding of the full orthogonal group $\mathrm{O}(2)$.

15.6 Corollary. $\mathrm{O}(2) = \{R_\theta : \theta \in \mathbb{R}\} \cup \{H_\theta : \theta \in \mathbb{R}\}$.

Another corollary of the theorem is that $\mathrm{O}(2)$ is generated by $\mathrm{SO}(2)$ and H_0. (This is still true if H_0 is replaced by any reflection.)

Now that we understand $\mathrm{SO}(2)$ and $\mathrm{O}(2)$, we return to the dihedral groups $\mathbb{D}_n$ defined in a previous section. Recall that

$$\mathbb{D}_n = \{i, r, r^2, \ldots, r^{n-1}\} \cup \{d, dr, dr^2, \ldots, dr^{n-1}\}$$

is the symmetry group of a regular n-gon. In the displayed decomposition, the first set consists of rotations of the n-gon, and the second consists of reflections of it. The reflection d is the one that fixes vertex n of the n-gon.

Note that all rotations and reflections of the n-gon must fix the center point (centroid) of the n-gon. If we place the n-gon on the euclidean plane $\mathbb{R}^2$ so that its centroid lies at the origin then the rotations and reflections in $\mathbb{D}_n$ extend to rotations and reflections of $\mathbb{R}^2$. (This requires further work to prove rigorously, but it is intuitively clear.)

This means that the elements of the symmetry group $\mathbb{D}_n$ may be represented by elements of the orthogonal group $\mathrm{O}(2)$; i.e., they can be represented by 2×2 orthogonal matrices. We choose to place the n-gon in such a way that its nth vertex lies on the positive x-axis. Then the extension of the symmetry d is the reflection H_0. So the representation $f : \mathbb{D}_n \to \mathrm{O}(2)$ is defined by

$$r \mapsto R_{2\pi/n}, \qquad d \mapsto H_0.$$

The representation preserves products, in the sense that $f(ab) = f(a)f(b)$, so the images displayed above determine the values of f on every element of $\mathbb{D}_n$. Since f is injective, it defines an isomorphism of $\mathbb{D}_n$ onto the subgroup $\langle R_{2\pi/n}, H_0\rangle$ of $\mathrm{O}(2)$ generated by $R_{2\pi/n}, H_0$.

So we can "understand" the dihedral group $\mathbb{D}_n$ by working with matrices.

What happens when we take the limit as n approaches ∞? Well, the regular n-gon approach a circle. The number of rotations, and thus also the number of reflections, increases to infinity. So in some sense it is fair to say that $\lim_{n\to\infty} \mathbb{D}_n = \mathrm{O}(2)$. This enables another way to think of $\mathrm{O}(2)$ as a symmetry group: $\mathrm{O}(2)$ is the symmetry group of a circle. For this reason, it is sometimes said that the group $\mathrm{O}(2)$ is the "infinite dihedral group."

Exercises

15.1. Show that any matrix of the form R_θ must belong to SO(2), for any $\theta \in \mathbb{R}$.

15.2. Show that the origin is a fixed point for any rotation operator ρ_θ.

15.3. Prove that the following identities follow from those in Corollary 15.3.

(a) $R_{\theta_1} R_{\theta_2} = R_{\theta_1+\theta_2}$ and $(R_\theta)^{-1} = R_{-\theta}$, for all $\theta, \theta_1, \theta_2 \in \mathbb{R}$.

(b) $R_{\theta_1} R_{\theta_2} = R_{\theta_2} R_{\theta_1}$ for all $\theta_1, \theta_2 \in \mathbb{R}$.

15.4. Use the results of the preceding exercise to give a conceptual derivation of the addition formulas for sine and cosine.

15.5. Give a different proof of Theorem 15.2, by showing that the angle between X and $\rho_\theta(X)$ is equal to θ, for any $0 \neq X \in \mathbb{R}^2$. [Hint: Recall that the angle between two vectors in $\mathbb{R}^2$ is determined by dot products.]

15.6. Show that the product of any two reflection matrices in O(2) must be a rotation matrix.

15.7. (a) Show that $H_\theta = R_{2\theta} H_0$.

(b) Show that $R_{2\theta} H_0 R_{2\theta} = H_0$. What formula holding for dihedral groups is this similar to?

15.8. (a) Argue that the composite of a reflection and a rotation (in either order) must be a reflection.

(b) Show that $H_{\theta_1} R_{\theta_2}$ must be some reflection matrix, i.e., some H_{θ_3}. Figure out what reflection matrix it is; i.e., figure out how to express θ_3 in terms of θ_1, θ_2. Justify your answer.

16 Matrix groups over other fields

Most of the basic theory of matrices, and indeed all of linear algebra, which is usually developed initially over the field $\mathbb{R}$ of real numbers, generalizes to any field F. In this generalization, the vector space $\mathbb{R}^n$ of n-tuples of real numbers is replaced by the vector space F^n of n-tuples over F, and matrices with entries from $\mathbb{R}$ are replaced by matrices with entries from F.

16.1 Examples. Here are some important examples of matrix groups over an arbitrary field F. The field F could be $\mathbb{Q}, \mathbb{C}$, or even a finite Galois field $\mathbb{F}_p$.

1. The *general linear group* $\mathrm{GL}(n, F) = \mathrm{GL}_n(F)$ is the group consisting of all $n \times n$ nonsingular matrices with entries from the field F. In symbols,

$$\mathrm{GL}_n(F) = \{n \times n \text{ matrices } A : \det A \neq 0\}.$$

The group $\mathrm{GL}_n(F)$ is finite if the field F is finite.

2. The *special linear group* $\mathrm{SL}(n, F) = \mathrm{SL}_n(F)$ is the group consisting of all $n \times n$ matrices of determinant equal to 1. In symbols,

$$\mathrm{SL}_n(F) = \{A \in \mathrm{GL}_n(F) : \det A = 1\}.$$

By definition, we have an inclusion $\mathrm{SL}_n(F) \subset \mathrm{GL}_n(F)$. Again, this is a finite group if the field F is finite.

It is also possible to define orthogonal groups over fields other than $\mathbb{R}$, but there are technicalities that we do not want to face at the moment.

When one allows the field F to be the field of complex numbers, we of course have $\mathrm{GL}_n(\mathbb{C})$ and $\mathrm{SL}_n(\mathbb{C})$ as above, but two important extra examples appear, as follows.

16.2 Examples. 1. The *unitary group* $\mathrm{U}(n) = \mathrm{U}_n(\mathbb{C})$ is the group consisting of all $n \times n$ unitary matrices. A square matrix A with complex entries is *unitary* if $A^{-1} = A^*$, where $A^* = \overline{A}^{\mathsf{T}}$. The matrix A^* is called the *conjugate transpose* of A. It is obtained by first taking the complex conjugate of each entry of A to get the matrix $\overline{A}$, and then taking the transpose.

2. The *special unitary group* $\mathrm{SU}(n) = \mathrm{SU}_n(\mathbb{C})$ is the group consisting of all $n \times n$ special unitary matrices. A square matrix A with complex entries is *special unitary* if it is unitary and has determinant equal to 1.

Unitary groups play a fundamental role in mathematical physics.

16.3 Example. Many basic properties of matrices work for matrices with entries from an arbitrary ring R. This observation leads to many more new examples of matrix groups. For instance, the group

$$\mathrm{SL}(n,\mathbb{Z}) = \mathrm{SL}_n(\mathbb{Z}) = \{n \times n \text{ matrices } A \text{ with integer entries} \mid \det A = 1\}$$

makes sense and has been studied extensively. It is an example of an *arithmetic group*. Arithmetic groups have connections to number theory and lattice theory.

Lattice theory has recently been applied to invent new public-key cryptosystems.

Exercises

16.1. (a) List all the matrices in $\mathrm{GL}_2(\mathbb{F}_2)$.
(b) List all the matrices in $\mathrm{SL}_2(\mathbb{F}_2)$.

16.2. Use counting principles to compute $|\mathrm{GL}_2(\mathbb{F}_p)|$.

16.3. Use counting principles to compute $|\mathrm{SL}_2(\mathbb{F}_p)|$.

16.4. Let F be any field. Let G be the set of all matrices of the form $E(t) = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$, where $t \in F$.
(a) Show that $E(s)E(t) = E(s+t)$ for any $s, t \in F$.
(b) Show that $E(t)^{-1} = E(-t)$ for any $t \in F$.
(c) Prove that the set G is a matrix group. (You have to show it is a nonempty set of matrices that is closed under products and inverses.)

16.5. Let G be the set of all block matrices of the form $\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$ with A, B, C all 2×2 matrices over a field F such that $\det(AC) \neq 0$. Verify that G is a matrix group.

16.6. Show that $\mathrm{SU}(1) = \mathrm{SU}_1(\mathbb{C})$ is isomorphic to the multiplicative group $\{e^{i\theta} \mid \theta \in \mathbb{R}\}$ of points on the unit circle, regarded as a subgroup of $\mathbb{C}^\times$.

Chapter 5

Abstract Groups

17 Abstract groups

The axiomatic definition of abstract group is based on special classes of examples of groups, such as permutation groups and matrix groups. Those examples share some common features: closure under products and inverses, associativity, and identity. In the definition we are about to give, closure under products is implicit in the definition of binary operation, while closure under inverses is an axiom.

We start with the concept of a binary operation on a set. Intuitively, a binary operation is a *law of combination* which combines two elements of a set to produce another element of the set. Ordinary addition and multiplication are canonical examples.

17.1 Definition. Let S be any given set. Any function from $S \times S$ to S is called a *binary operation* or *law of combination* on the set S. If f is a binary operation then tradition demands that we write $x\, f\, y$ for the value[1] at the input pair (x, y) instead of the usual $f(x, y)$.

In this context the word *binary* refers to the fact that the function depends on two input variables. By the same token, a *unary operation* on the set S would be a function from S to itself. For example, the function that sends each integer to its negative is a unary operation on the set $\mathbb{Z}$ of integers.

17.2 Examples. 1. Addition $(+)$ is a binary operation on any of the usual

[1] Writing a function between its arguments is called *infix* notation in computer science.

number sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$. Multiplication $(\cdot)$ is another binary operation on any of the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$.

2. Matrix multiplication is a binary operation on the set $\mathrm{Mat}_n(F)$ of all $n \times n$ matrices with entries in a given field F.

3. Composition of functions is a binary operation on the set $\mathbb{S}_n$ of permutations of $\underline{\mathbf{n}} = \{1, \ldots, n\}$.

4. More generally, composition of functions is a binary operation on the set S^S of all self-maps $S \to S$ of *any* set S.

5. Here is a binary operation $\#$ on the finite set $S = \{a, b, c, d\}$ which is defined by means of a "multiplication" table as follows:

$\#$	a	b	c	d
a	a	a	d	d
b	b	b	c	c
c	a	d	b	c
d	b	c	a	d

This table defines a law of combination for pairs of elements of S. For instance, it says that $a\#c = d$ and $c\#b = d$. For finite sets S, we can always define a binary operation (law of combination) on S by a table.

It is important to realize that closure under products is built in to the definition of a binary operation: to say that $*$ is a binary operation on S *means* that S is closed under $*$, since all the values $x * y$ must fall again within S, for any $x, y \in S$. This is just another way of saying that $*$ is a function mapping $S \times S \to S$.

17.3 Definition. A *group* is a set G along with a given binary operation $*$ on G, such that the following three axioms hold:

(G1) The operation $*$ is *associative*: $(a * b) * c = a * (b * c)$ for all $a, b, c \in G$.

(G2) There is an *identity* element $e \in G$ satisfying: $e * a = a = a * e$ for all $a \in G$.

(G3) Each $a \in G$ has an *inverse* in G: given $a \in G$ there exists some $a' \in G$ such that $a * a' = a' * a = e$.

Note that closure under $*$ is implicit in this definition, since $*$ is a binary operation on G. Moreover, closure under inverses is the content of axiom (G3).

When describing a group we should specify not only the set G but also the binary operation $*$ on G, since a given set can have many different binary

operations defined on it. People often use a notation such as $(G, *)$ to denote a group. If the binary operation $*$ is implied by context, then we often just write G for simplicity.

17.4 Definition. An *abelian*[2] group is any group $(G, *)$ in which the commutative law holds: $a * b = b * a$, for all $a, b \in G$.

As we have seen, matrix and permutation groups are usually not abelian, because matrix multiplication and functional composition are usually not commutative.

17.5 Theorem (Basic properties)**.** *Let $(G, *)$ be any group, with binary operation $*$, and let $a, d, x, y \in G$.*

(a) The identity element $e \in G$ in axiom (G2) *is unique.*

(b) The inverse of any $a \in G$ in axiom (G3) *is unique.*

(c) The equation $a * x = a * y$ implies that $x = y$. (This is called left cancellation.*)*

(d) The equation $x * a = y * a$ implies that $x = y$. (This is called right cancellation.*)*

(e) Each of the equations $a * x = b$, $x * a = b$ $(a, b \in G)$ has a unique solution $x \in G$.

(f) The inverse of a product is the product of the inverses in reverse order.

Proof. (a) Suppose that e, f are identity elements of G. Then by axiom (G2) we have $e * a = a$ and $a = a * f$ for all $a \in G$. In particular, taking $a = f$ in the first equality and $a = e$ in the second, we get $e * f = f$ and $e = e * f$. Hence $e = f$. This proves uniqueness of identity.

(b) Suppose that b, c are both inverses of a given $a \in G$. Then $a * b = e = b * a$ and $a * c = e = c * a$ by axiom (G3). By the associative law (G1) we have $c * (a * b) = (c * a) * b$, so $c * e = e * b$, so $c = b$ by (G2). This proves uniqueness of inverses.

(c) Suppose $a * x = a * y$. Then $a' * (a * x) = a' * (a * y)$ where a' is the inverse of a. By (G1) this implies that $(a' * a) * x = (a' * a) * y$, so by (G3) we have $e * x = e * y$, which implies by (G2) that $x = y$.

(d) This is proved similarly to (c), except we multiply by the inverse a' on the right instead of on the left.

(e) Suppose that $a * x = b$. Then by left multiplication by the inverse a' of a we have $a' * (a * x) = a' * b$, so by (G1) we have $(a' * a) * x = a' * b$.

[2] In honor of Niels Henrik Abel (1802–1829).

Thus by (G3) we have $e * x = a' * b$, so by (G2) we obtain $x = a' * b$. This is the unique solution. This proves the first claim. The other claim is proved similarly, using right multiplication instead of left multiplication.

(f) Given two elements a, b of G let their respective inverses be a', b'. Then by axiom (G3) we have $a * a' = e$, $b * b' = e$. Thus

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

where we used generalized associativity for the first equality. Let z be the inverse of $a * b$. Then $(a * b) * z = e$ by (G3). So $(a * b) * z = (a * b) * (b' * a')$. By left cancellation we obtain $z = b' * a'$, i.e., the inverse of $a * b$ is $b' * a'$. This proves the statement for products of length two, and it easily extends to products of more than two elements, by induction on the length of the product. □

Additive versus multiplicative notation

If the binary operation $*$ is written as addition $(+)$ or multiplication $(\cdot)$ then the group is known as an *additive* group or a *multiplicative* group, respectively.

If $(G, +)$ is an additive group, it is customary to denote the identity element by the symbol 0 and the inverse of a by the symbol $-a$. In this case the group axioms take the following form:

(G1) $(a + b) + c = a + (b + c)$;

(G2) $0 + a = a = a + 0$;

(G3) $a + (-a) = 0 = (-a) + a$.

It is customary to use the additive notation for a group *only for abelian groups* and we shall follow that convention in this course.

If $(G, \cdot)$ is a multiplicative group, it is customary to abbreviate products $a \cdot b$ by ab. In this case we usually denote the identity element by the symbol 1 and the inverse of a by the symbol a^{-1}. Then the group axioms take the form:

(G1) $(ab)c = a(bc)$;

(G2) $1a = a = a1$;

(G3) $aa^{-1} = 1 = a^{-1}a$.

The default notation for the group operation is multiplicative notation, but additive groups appear frequently as well.

17.6 Examples. (a) Any permutation group is a group. Any matrix group is a group.

(b) The dihedral group $\mathbb{D}_n$ is a group. All symmetry groups are groups.

(c) The *abstract cyclic group* is the multiplicative group C_n generated by a symbol x subject to the relation $x^n = 1$. As a set, $C_n = \{1, x, x^2 \ldots, x^{n-1}\}$, so $|C_n| = n$.

(d) Any vector space V gives an additive abelian group $(V, +)$ under vector addition. The identity element is the zero vector $\mathbf{0}$ in V and the additive inverse of a vector $\mathbf{v} \in V$ is the vector $-\mathbf{v}$. In particular, $(\mathbb{R}^n, +)$ is an example of such a group, and more generally we have the group $(F^n, +)$ where F is any field.

(e) Any ring R (commutative or not) contains *two* groups. One is the additive abelian group $(R, +)$, in which 0 is the additive identity and the inverse of a is written as $-a$. In particular, $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$ are all additive abelian groups. Also, $(\mathbb{Z}_n, +)$ is an additive abelian group, for any positive integer n.

(f) Recall that $R^\times = R^*$ is the set of units[3] in a given ring R. The second group in the ring R is the *multiplicative group of units* in R, i.e., the group $(R^\times, \cdot)$ or just $R^\times$ for short. In particular, for any positive integer n, we have the multiplicative group $\mathbb{Z}_n^\times$ of units in the ring $\mathbb{Z}_n$. The abelian group $\mathbb{Z}_n^\times$ is of extreme importance for modern cyptography.

(g) Let F be any field. Then by definition $F^\times = F - \{0\}$, the set of all nonzero elements of F, since every nonzero element of a field is invertible. So the previous example gives in this case the group $(F^\times, \cdot) = (F - \{0\}, \cdot)$, which is called the *multiplicative group of the field* F.

(h) The *trivial group* is the set $\{e\}$ consisting of only one element, with $e * e = e$.

The number of elements of a group is called its *order*.

17.7 Definition. Let $(G, *)$ be a group. The *order* of the group is the cardinality $|G|$ of the set G. If the set G is an infinite set then we often write $|G| = \infty$ and call G an infinite group, otherwise G is a finite group.

The word *order* is also used in group theory in another way, as follows, when speaking about an element of a group.

17.8 Definition. Let $a \in G$ where $(G, *)$ is a group. The *order* of a is the least positive integer r such that a combined with itself r times yields the identity element e. If no such r exists, then the order is defined to be ∞.

[3] A *unit* is an invertible element.

In any group, the order of the identity element is always 1. Since the word *order* has two different meanings within group theory, we always have to determine its usage from the context.

17.9 Lemma. *If a is an element of order r in a group $(G, *)$, then the inverse of a is equal to $a * \cdots * a$ ($r-1$ factors) obtained by combining a with itself $r-1$ times.*

Proof. Let b be equal to $a * \cdots * a$ ($r-1$ factors). Then it is clear that $a * b = e = b * a$. It follows from uniqueness of inverses that b equals the inverse of a. □

Now we discuss *laws of exponents* in groups. We need to distinguish between multiplicative and additive groups, which use different notation. If a is an element of a multiplicative group, then we define a^n to be $aa\cdots a$ (n times repeated) for any positive integer n, we define $a^0 = 1$, and we define $a^{-n} = (a^{-1})^n$. Note that $(a^n)^{-1} = a^{-n}$. Moreover, we have

$$a^m\, a^n = a^{m+n}, \qquad (a^m)^n = a^{mn}$$

for any $m, n \in \mathbb{Z}$.

In an additive group we have to use a different notation. In this case we define na to be $a + a + \cdots + a$ (n summands) for any positive integer n, we define $0a = 0$, and we define $(-n)a = n(-a)$. Note that $-(na) = (-n)a$. Moreover,

$$ma + na = (m+n)a, \qquad n(ma) = (nm)a$$

for any $m, n \in \mathbb{Z}$. In an additive group, "powers" are written as multiples, because in ordinary arithmetic repeated addition is written as a multiple while repeated multiplication is written as a power.

17.10 Remark. We can rephrase Definition 17.8 in terms these notations as follows. In a multiplicative group the order of a is the least positive integer r such that $a^r = 1$. In the additive case the order of a is the least positive integer r such that $ra = 0$.

Now we define the important notion of isomorphism of groups.

17.11 Definition. Let $(G, *)$ and $(H, \#)$ be given groups. An *isomorphism* of G onto H is any bijection $f\colon G \to H$ such that $f(a * b) = f(a)\#f(b)$ for all $a, b \in G$. Whenever such an f exists then we say that G is *isomorphic* to H, and write $G \cong H$ or $G \simeq H$ (interchangeably).

So an isomorphism is a bijective mapping from one group to the other which matches up products in the two groups. If one group is an additive group and the other a multiplicative group then this means that sums get matched with products.

If two groups are isomorphic then *they are essentially the same group*, except for the form of their elements. In particular, isomorphic groups must have the same structural properties (i.e, they have same order, the same number of subgroups, etc). The following is easy to check.

17.12 Theorem. *Isomorphism of groups is an equivalence relation on the class of groups: it is reflexive, symmetric, and transitive.*

17.13 Example. It is easy to check that the set $(\mathbb{R}^+, \cdot)$ of all positive real numbers is a group under multiplication. We claim that the multiplicative group $(\mathbb{R}^+, \cdot)$ is isomorphic to the additive group $(\mathbb{R}, +)$ of real numbers. The isomorphism is given by the natural logarithm function $x \mapsto \ln x$. We know this function is invertible (its inverse is the exponential function $x \mapsto e^x$) so it is a bijection of $\mathbb{R}^+$ onto $\mathbb{R}$. Furthermore, the equation $\ln(ab) = \ln(a) + \ln(b)$ says that products in $\mathbb{R}^+$ match up with sums in $\mathbb{R}$, so the function ln is indeed a group isomorphism, as claimed.

If a group $(G, *)$ is finite then it may be described by giving its complete multiplication table. (Replace multiplcation by addition if it is an additive group.) For instance, the addition table of $(\mathbb{Z}_4, +)$ and the multiplication table of the cyclic group $G = \langle \alpha \rangle$ generated by a 4-cycle α are displayed in the tables below

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	1	α	α^2	α^3
1	1	α	α^2	α^3
α	α	α^2	α^3	1
α^2	α^2	α^3	1	α
α^3	α^3	1	α	α^2

where we write the numbers $0, 1, 2, 3$ as a shorthand for the corresponding residue classes $[0], [1], [2], [3]$ in $\mathbb{Z}_4$. Note that the correspondence

$$[0] \to 1 = \alpha^0, \quad [1] \to \alpha = \alpha^1, \quad [2] \to \alpha^2, \quad [3] \to \alpha^3$$

defines an isomorphism between the two groups. In a more succinct notation, the isomorphism is defined by $f([x]) = \alpha^x$ for $x = 0, 1, 2, 3$.

The multiplication table of a finite group has the property that the elements in any row of the table form a permutation of the elements of

any other row. The same is true of the columns of the table. In particular, no element appears twice in any row or any column. You should be able to show how these claims follow from the group axioms.

Consider a group of order 2. Assume that its binary operation is written multiplicatively, so $G = \{1, a\}$ as a set, where 1 is the identity element and $a \neq 1$ (else the group would have order 1). Then necessarily $a^2 = 1$, because $a^2 = a$ implies $a = 1$. So the group multiplication table must look like the one on the left below:

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

$+$	0	1
0	0	1
1	1	0

The table on the right is the addition table for the additive group $(\mathbb{Z}_2, +)$. It should be clear that the two groups are isomorphic. This analysis shows that any group of two elements must be isomorphic to $\mathbb{Z}_2$. This argument can be extended to prove that any group of three elements must be isomorphic to $\mathbb{Z}_3$.

Exercises

17.1. Explain your reasoning for:
   (a) Is $\mathbb{N} = \{1, 2, 3, \dots\}$ a group under addition? If we include 0 is it a group?
   (b) Is $\mathbb{N} = \{1, 2, 3, \dots\}$ a group under multiplication?
   (c) Is the set $\mathbb{Z}$ of integers a group under addition?
   (d) Is $\mathbb{Z}$ a group under multiplication?
   (e) Is $\mathbb{Z} - \{0\}$ a group under multiplication?

17.2. What is wrong with writing $\frac{a}{b}$ for ab^{-1} in a (nonabelian) multiplicative group? If you think there is nothing wrong with it, then how will you write $b^{-1}a$ when $ab^{-1} \neq b^{-1}a$?

17.3. Prove that any group of three elements must be isomorphic to the additive group $\mathbb{Z}_3$ by analyzing its multiplication table.

17.4. Suppose that $G = \{1, a, b, c\}$ is a multiplicative group of four elements in which 1 is the identity element. By analyzing the possible multiplication tables, prove that G is isomorphic to either $(Z_4, +)$ or to a group in which $a^2 = b^2 = c^2 = 1$. (The latter group is called the Klein 4-group.)

17.5. List the elements in the following multiplicative groups:
   (a) $(\mathbb{Z}^\times, \cdot)$, (b) $(\mathbb{Z}_6^\times, \cdot)$, (c) $(\mathbb{Z}_8^\times, \cdot)$, (d) $(\mathbb{Z}_{15}^\times, \cdot)$.

17.6. Give multiplication tables for the groups in the previous problem.

17.7. Prove that $|\mathbb{Z}_n^\times| = \varphi(n)$, where $\varphi(n)$ is *Euler's phi-function* from number theory.

17.8. Prove that the elements in any row of the group multiplication table of a finite group G form a permutation of the elements of the first row. Then do the same for columns.

17.9. In this problem, we write $\mathbb{Z}_n$ for the additive group $(\mathbb{Z}_n, +)$. Find the order of:

(a) 1 in $\mathbb{Z}_7$, (b) 2 in $\mathbb{Z}_7$, (c) 1 in $\mathbb{Z}_{10}$, (d) 2 in $\mathbb{Z}_{10}$, (e) 3 in $\mathbb{Z}_{10}$.

17.10. In this problem, we write $\mathbb{Z}_n$ for the additive group $(\mathbb{Z}_n, +)$. Find the order of any $a \in \mathbb{Z}_n$ and prove your answer.

17.11. In this problem, we write $\mathbb{Z}_n^\times$ for the multiplicative group $(\mathbb{Z}_n^\times, \cdot)$ of units. Find the orders of:

(a) $1, 2, 3, 4, 5, 6$ in $\mathbb{Z}_7^\times$, (b) $1, 2, 4, 5, 7, 8$ in $\mathbb{Z}_9^\times$, (c) $1, 3, 7, 9$ in $\mathbb{Z}_{10}^\times$.

17.12. (The circle group) Let S^1 be the set of all points on the usual unit circle in the plane $\mathbb{R}^2$. Show that S^1 is a group under the law of combination given by

$$(\cos\theta, \sin\theta) * (\cos\theta', \sin\theta') = (\cos(\theta + \theta'), \sin(\theta + \theta')).$$

Be sure to give a formula for the inverse of elements of this group, and prove that they really are inverses.

17.13. Show that the circle group of the previous problem is isomorphic to the matrix group $\mathrm{SO}(2)$.

17.14. Find an isomorphism of the multiplicative group $\mathbb{Z}^\times$ onto the additive group $\mathbb{Z}_2$.

17.15. Consider the set G of all 2×2 real matrices of the form $\left[\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right]$.

(a) Show that G is a group under ordinary matrix addition, and find an isomorphism from the additive group $(\mathbb{C}, +)$ onto G.

(b) Now let G' be the subset of G consisting of all elements of G except the zero matrix. Show that G' is a group under ordinary matrix multiplication.

(c) Find an isomorphism from the multiplicative group $(\mathbb{C}^\times, \cdot)$ onto G'.

17.16. Prove part (f) of Theorem 17.5 using right cancellation instead of left cancellation.

17.17. Prove that isomorphism of groups is an equivalence relation on the class of groups (Theorem 17.12).

17.18. Prove that if G is a group in which every element (except the identity) has order 2 then G must be abelian.

17.19. (Monoids) A *monoid* is a set M along with a binary operation $* : M \times M \to M$ such that $*$ is associative and there is an identity element $e \in M$. Show that $(\mathbb{N}, +)$ and $(\mathbb{Z}, \cdot)$ are monoids but not groups.

17.20. Show that a set R with two binary operations $+, \cdot$ is a ring if and only if the

following three properties hold:

(a) $(R, +)$ is an additive abelian group. Denote its identity element by 0.
(b) $(R, \cdot)$ is a multiplicative monoid (see Problem 17.19 for the definition of monoid). Denote its identity element by 1.
(c) Addition and multiplication are connected by the distributive laws: $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$, for all $a, b, c \in R$.

Thus, one could take the three properties as the definition of ring.

17.21. Show that a set F with two binary operations $+, \cdot$ is a field if and only if the following four properties hold:

(a) $(F, +)$ is an additive abelian group. Denote its identity element by 0.
(b) $(F - \{0\}, \cdot)$ is a multiplicative abelian group. Denote its identity element by 1.
(c) $1 \neq 0$.
(d) Addition and multiplication are connected by the distributive law: $a(b + c) = ab + ac$, for all $a, b, c \in F$.

Thus, one could take the four properties as the definition of field.

18 Subgroups

Finding subgroups inside known groups is an important way of finding new examples of groups.

18.1 Definition. Let $(G, *)$ be a group. A subset H of G is called a *subgroup* of G if $(H, *)$ is a group in its own right. We write $H < G$ (or $H \leq G$ interchangeably) to denote that H is a subgroup of G. A subgroup H is a *proper* subgroup of G (written as $H \lneq G$) if $H \neq G$.

By definition, permutation groups are subgroups of some $\mathbb{S}_n$ and matrix groups are subgroups of some $\mathrm{GL}_n(F)$, where F is a field. So we have already seen many examples of subgroups.

Note that every group is regarded as a subgroup of itself. (So in the above notation, writing $G < G$ is perfectly valid.) In any group the subset consisting solely of the identity element is always a subgroup; this subgroup is called the *trivial* subgroup.

In order for a given subset H of a group G to be a subgroup, it is clearly necessary that the binary operation $* : G \times G \to G$ restricts to a binary operation $* : H \times H \to H$. This is just another way of saying that H must be closed under products.

The following theorem covers both the multiplicative and additive cases together. In the latter case, you should read "sum" for "product" in the theorem, because $a * b = a + b$ when the operation $*$ is equal to $+$.

18.2 Theorem (The subgroup criterion)**.** *Let H be a nonempty subset of a given group $(G, *)$. Then H is a subgroup of G if and only if H is closed under products and inverses. (Closure under products means that $a * b \in H$ whenever $a, b \in H$, and closure under inverses means that H contains the inverse of each of its elements.)*

Proof. ($\Longrightarrow$) Suppose that H is a subgroup of G. Then the fact that H is a group in its own right means that when we restrict the operation $*\colon G \times G \to G$ to the subset $H \times H$, it induces a binary operation $*\colon H \times H \to H$. This is equivalent to saying that H is closed under products. Also, the fact that axiom (G3) holds for H means that each $a \in H$ has an inverse a' in H; that must also be its inverse in G since inverses in G are unique. This implies that H is closed under inverses.

($\Longleftarrow$) Suppose that H is a nonempty subset of G which is closed under products and inverses. Then the restriction of $*$ to $H \times H$ maps into H, and thus defines a binary operation on H. Axiom (G1) is automatic in H

since it holds in the bigger set G. Axiom (G3) for H is just closure under inverses, which is true by assumption. Finally, closure under products and inverses implies that the identity e is in the set H, since there must be some $a \in H$ (because H is nonempty) and then its inverse a' is in H and thus $a*a' = e \in H$ by closure under products. This proves axiom (G2) for H. □

The following result is a slightly simplified version of the subgroup criterion.

18.3 Theorem (Simplified subgroup criterion)**.** *Let H be a nonempty subset of a given group G, and write b' for the inverse of $b \in G$. Then H is a subgroup of G if and only if $a * b' \in H$ for all $a, b \in H$.*

Proof. ($\Longrightarrow$) Suppose $H < G$. Then by the subgroup criterion, for any $a, b \in H$ it follows that $b' \in H$ and hence $a * b' \in H$.

($\Longleftarrow$) For the converse, suppose that H is a nonempty subset and $a*b' \in H$ for all $a, b \in H$. Since H is non-empty there is at least one element $c \in H$. Hence $c * c' \in H$, so the identity $e \in H$. Hence $b' = e * b' \in H$ for every $b \in H$, proving that H is closed under inverses. Finally, if a, b are any elements of H, then $b' \in H$ as we have just shown. Note that $(b')' = b$, so $b = d'$ where $d = b' \in H$. Hence the product $a * b = a * d'$ must be an element of H. This shows that H is closed under products. So $H < G$ by the subgroup criterion. □

The criterion for finding subgroups of a *finite* group is even simpler: we only have to check closure under products.

18.4 Corollary (Subgroup criterion for finite groups)**.** *Let H be a nonempty subset of a given finite group G. Then H is a subgroup of G if and only if $a * b \in H$ for all $a, b \in H$.*

Proof. Every element of a finite group must have finite order, so closure under products implies also closure under inverses. (By Lemma 17.9, if a has order r then the inverse of a is obtained by combining a with itself $r - 1$ times.) □

18.5 Example. We compute the subgroups of $\mathbb{S}_3$, the symmetric group on 3 letters, using the finite subgroup criterion. We have (in the cycle notation)

$$\mathbb{S}_3 = \{(1), (1,2), (2,3), (1,3), (1,2,3), (3,2,1)\}.$$

Here we use (1) for the identity permutation. The smallest subgroup of $\mathbb{S}_3$ is the trivial subgroup $\{(1)\}$. Next we have the two element subgroups

$\{(1),(1,2)\}$, $\{(1),(2,3)\}$, and $\{(1),(1,3)\}$. The subgroup $\{(1),(1,2,3),(3,2,1)\}$ is of order 3. Finally, we have $\mathbb{S}_3$ itself, a subgroup of order 6. It is easy to check that these are the only subgroups of $\mathbb{S}_3$.

18.6 Theorem. *The intersection of any number of subgroups of a given group G is always a subgroup of G.*

Proof. This is an application of the subgroup criterion. Suppose that I is some indexing set and $H_i \leq G$ for each $i \in I$. Then we need to show that $K = \bigcap_{i\in I} H_i$ is a subgroup of G. Note that K is nonempty since the identity element belongs to each subgroup H_i and hence belongs to the intersection K. Suppose that $x, y \in K$. Then $x, y \in H_i$ for all $i \in I$. Since H_i is a subgroup, this means that both $x * y$ and the inverse of x are in H_i, for all $i \in I$, so $x * y \in K$ and the inverse of x is in K. By the subgroup criterion, K is a subgroup of G. □

In contrast, unions of subgroups are usually *not* subgroups.

18.7 Definition. If S is any set of group elements in some group G then $\langle S\rangle$ is the smallest subgroup of G containing the elements of S. The subgroup $\langle S\rangle$ is called the subgroup *generated by* the set S. In particular, if $a \in G$ is a group element, then we write $\langle a\rangle$ short for $\langle\{a\}\rangle$; this is called the *cyclic subgroup generated by a.*

18.8 Examples. 1. If $a \in G$ has finite order r, then $\langle a\rangle = \{1, a, a^2, \ldots, a^{r-1}\}$ if G is a multiplicative group, and $\langle a\rangle = \{0, a, 2a, \ldots, (r-1)a\}$ if G is an additive group. In either situation, $\langle a\rangle$ is isomorphic to the abstract cyclic group C_r of order r.

2. In the additive group $\mathbb{Z}$ of integers, we have $\langle 1\rangle = \mathbb{Z}$ and $\langle -1\rangle = \mathbb{Z}$. Hence $\mathbb{Z}$ is a cyclic group under addition). Furthermore, $\langle 0\rangle = \{0\}$ (the trivial group) and $\langle 2\rangle = \langle -2\rangle = 2\mathbb{Z}$ (the subgroup of even integers).

3. In the multiplicative group $\mathbb{R}^\times$ of nonzero real numbers, the subgroup $\langle\pi\rangle = \{\pi^k \mid k \in \mathbb{Z}\}$. This group is a proper subgroup of $\mathbb{R}^\times$, and it is isomorphic to the additive group $\mathbb{Z}$. More generally, for *any* chosen element $a \in \mathbb{R}^\times$, it can be seen that $\langle a\rangle = \{a^k \mid k \in \mathbb{Z}\}$.

4. Even more generally, suppose that $a \in R^\times$ is an element of infinite order in the multiplicative group of units in a ring R. Then $\langle a\rangle = \{a^k \mid k \in \mathbb{Z}\}$ is an infinite cyclic group isomorphic to the additive group $\mathbb{Z}$.

18.9 Definition. If S is a set of elements of a group G, we say that G is *generated by* S if $G = \langle S\rangle$. If there is a *finite* set S with this property then

we say that G is *finitely generated*. A group is called *cyclic* if it is generated by a single element: i.e., if $G = \langle a \rangle$ for some $a \in G$.

If G is generated by a set S, then every element of G can be expressed as a product of elements of S and their inverses.

18.10 Examples. 1. The symmetric group $\mathbb{S}_n$ is generated by the set of transpositions it contains.

2. The general linear group $\mathrm{GL}(n)$ is generated by the set of elementary matrices.

3. The dihedral group $\mathbb{D}_n$ is generated by the two elements r, d defined earlier, so $\mathbb{D}_n = \langle r, d \rangle$. (Recall that r is a basic rotation and d a reflection.)

4. The additive group $(\mathbb{Z}, +)$ of integers is cyclic, because $\mathbb{Z} = \langle 1 \rangle$. So is the additive group $(\mathbb{Z}_n, +)$ of integers modulo n, because $\mathbb{Z}_n = \langle [1] \rangle$.

5. Every finite group is finitely generated. So is every cyclic group (including any infinite cyclic group).

6. The additive group $\mathbb{R}$ of real numbers is not finitely generated.

7. The matrix group $\mathrm{GL}(n)$ is not finitely generated. (We have proved that the set S of all elementary matrices generates $\mathrm{GL}(n)$, but S is an infinite set. It turns out that no finite generating set exists, but it isn't so easy to prove.)

Next we investigate another way to find subgroups from subsets of elements of a given group.

18.11 Definition. Suppose that S is a set of elements of a group $(G, *)$. The *centralizer* of S in G is the subgroup $Z_G(S) = \{x \in G \mid x*s = s*x \text{ for all } s \in S\}$. The *center* of G is $Z(G) = Z_G(G) = \{x \in G \mid x*g = g*x \text{ for all } g \in G\}$, the centralizer of G in itself. If $a \in G$ then we write $Z_G(a)$ short for $Z_G(\{a\})$.

It is an exercise to verify that centralizers really are subgroups. In particular, this implies that the center $Z(G)$ of a group G is always a subgroup. The center is, by definition, the set of elements that commute with all the elements of the group.

Exercises

18.1. Show that the set $\{\pm 1, \pm i\}$ is a subgroup of the multiplicative group $\mathbb{C}^\times$. Is it a cyclic group?

18.2. Show by example that a union of two subgroups need not be a subgroup.

18.3. Show that the set $2\mathbb{Z}$ of all even integers is a subgroup of the additive group $\mathbb{Z}$, but the set $2\mathbb{Z}+1$ of all odd integers is not a subgroup.

18.4. Show that for any integer n,

(a) the set $n\mathbb{Z}$ of all multiples of n is a subgroup of the additive group $\mathbb{Z}$.

(b) the subgroup $n\mathbb{Z}$ is isomorphic to $\mathbb{Z}$ itself.

(c) these are the only subgroups of $\mathbb{Z}$; i.e., every subgroup of $\mathbb{Z}$ is of the form $n\mathbb{Z}$ for some integer n.

18.5. If G is any subgroup of $\mathrm{GL}(n)$, let $H = \{A \in G \mid \det A = \pm 1\}$. Prove that $H < G$.

18.6. Let F be a field. If G is any subgroup of $\mathrm{GL}_n(F)$, let $H = \{A \in G \mid \det A = \pm 1\}$. Prove that $H < G$.

18.7. (Abstract Klein 4-group) The abstract *Klein 4-group* K may be defined as the (unique) group $\{1, a, b, c\}$ of four elements such that 1 is the identity and a, b, c all have order 2.

(a) Show that this description determines the group K uniquely, by writing out its only possible multiplication table.

(b) Find all the subgroups of K.

18.8. Prove that if α is an element of order n in a permutation group G then the subgroup $\langle\alpha\rangle$ generated by α is isomorphic to the additive group $(\mathbb{Z}_n, +)$.

18.9. (Classification of cyclic groups) Prove that if $(G, *)$ is a cyclic group then it is isomorphic with either the additive group $\mathbb{Z}_n$ for some n or with the additive group $\mathbb{Z}$ of all integers.

18.10. Apply the previous exercise to deduce that the additive group $(\mathbb{R}, +)$ is not cyclic.

18.11. Show by contradiction that the additive group $(\mathbb{Q}, +)$ is not cyclic.

18.12. Show that every subgroup of a cyclic group must be cyclic. [Hint: Use the result of Exercise 18.9.]

18.13. Show that the group $(\mathbb{Z}_n, +)$ is generated by $[a] \in \mathbb{Z}_n$ if and only if a, n are relatively prime. Use this to deduce that a cyclic group of order n has exactly $\varphi(n)$ generators. [Hint: Use the result of Exercise 18.9 for the second part.]

18.14. If H is a subgroup of a group $(G, *)$ and $a \in G$, let $a * H * a' = \{a * h * a' \mid h \in H\}$, where a' is the inverse of a.

(a) Show that $a * H * a'$ is a subgroup of G.

(b) If H is finite, say $|H| = n$, then what is $|a * H * a'|$?

18.15. Show that the dihedral group $\mathbb{D}_n$ ($n \geq 3$) is not cyclic.

18.16. Show that if G is a group of order n then G is cyclic if and only if it has an element of order n.

18.17. Write $\mathbb{Z}_n$ for the additive group $(\mathbb{Z}_n, +)$. Show that $\mathbb{Z}_n = \langle a\rangle$ for $a \in \mathbb{Z}_n$ if and only if $\gcd(a, n) = 1$.

18.18. Show that the multiplicative group $\mathbb{F}_7^\times$ is cyclic by finding a generator. Do

the same for $\mathbb{F}_{13}^{\times}$.

18.19. Is the multiplicative group $\mathbb{Z}_8^{\times}$ cyclic? Same question for $\mathbb{Z}_{10}^{\times}$. Justify your answers.

18.20. Find a minimal generating set for the Klein 4-group (see Exercise 18.7).

18.21. Show that the matrix group $\mathrm{O}(2)$ is generated by the set $\mathrm{SO}(2) \cup \{A\}$, where $A \in \mathrm{O}(2)$ is any improper orthogonal matrix.

18.22. Show that if a, b are elements of some multiplicative group G then $\langle a \rangle < \langle b \rangle$ if and only if $a = b^k$ for some integer k.

18.23. (The quaternion group) The quaternion group is the group $Q = \{\pm 1, \pm i, \pm j, \pm k\}$ of order 8 defined by the following multiplication table:

$\cdot$	1	-1	i	$-i$	j	$-j$	k	$-k$
1	1	-1	i	$-i$	j	$-j$	k	$-k$
-1	-1	1	$-i$	i	$-j$	j	$-k$	k
i	i	$-i$	-1	1	k	$-k$	$-j$	j
$-i$	$-i$	i	1	-1	$-k$	k	j	$-j$
j	j	$-j$	$-k$	k	-1	1	i	$-i$
$-j$	$-j$	j	k	$-k$	1	-1	$-i$	i
k	k	$-k$	j	$-j$	$-i$	i	-1	1
$-k$	$-k$	k	$-j$	j	i	$-i$	1	-1

in which 1 is the identity element, i, j, k all behave like imaginary units in that $i^2 = j^2 = k^2 = -1$, and products of any pair chosen from i, j, k behave like cross products of the standard unit vectors in $\mathbb{R}^3$.

(a) Find the cyclic subgroups $\langle -1 \rangle$, $\langle i \rangle$, $\langle j \rangle$, and $\langle k \rangle$.
(b) Show that Q is not cyclic.
(c) Find a minimal set of generators of Q, and justify your answer.
(d) Compute the center $Z(Q)$.

Note: The quaternion group Q is related to Hamilton's *quaternions*, which puts a division ring structure on Euclidean four dimensional space.

18.24. Let $C = \{z \in \mathbb{C} \colon |z| = 1\}$ be the set of all complex numbers of unit norm, where as usual the *norm* (length) of a complex number $z = x + iy$ is defined to be $|z| = \sqrt{x^2 + y^2}$. By Euler's identity $e^{i\theta} = \cos\theta + i \sin\theta$ (valid for all $\theta \in \mathbb{R}$) it follows that $C = \{e^{i\theta} \mid \theta \in \mathbb{R}\}$.

(a) Prove that C is a subgroup of the multiplicative group $\mathbb{C}^{\times}$.
(b) Find an isomorphism from S^1 onto C, where S^1 is the circle group defined in a previous exercise.

18.25. If $G = \mathbb{D}_n$ find $Z_G(r)$ where r is the basic rotation. Then find $Z_G(f)$ where f is any reflection.

18.26. Prove that if $a \in G$ then $\langle a \rangle < Z_G(a)$.

18.27. Prove that $Z_G(S) < G$ for any set S of elements in a group G. Why does this also prove that the center $Z(G) < G$?

18.28. Compute the center of $\mathbb{D}_4$ and justify your answer.

18.29. Show that $Z(\mathbb{S}_n)$ for $n \geq 3$ is the trivial group. What is $Z(\mathbb{S}_2)$?

18.30. Prove that $Z(G)$ is always abelian.

18.31. Show that $Z(\mathbb{D}_n)$ has order 1 or 2 depending whether n is odd or even, respectively.

18.32. Prove that $Z(G) = G$ if and only if G is abelian.

18.33. This problem is about permutations, written in terms of the cycle notation.

(a) Show that $(1,3) = (2,3)(1,2)(2,3)$.

(b) Show that $(1,4) = (3,4)(2,3)(1,2)(2,3)(3,4)$.

(c) Prove that for $j > 1$ we have $(1,j) =$

$$(j-1,j)(j-2,j-1)\cdots(1,2)\cdots(j-2,j-1)(j-1,j).$$

(d) Prove that for $i < j$ we have $(i,j) =$

$$(j-1,j)(j-2,j-1)\cdots(i,i+1)\cdots(j-2,j-1)(j-1,j).$$

This shows that it is possible to write any transposition as a product of *adjacent* ones; i.e., ones of the form $(k, k+1)$.

18.34. Prove that $\mathbb{S}_n$ is generated by the set $\{(1,2),(2,3),\ldots,(n-1,n)\}$ of *adjacent* transpositions. [Hint: Use Problem 18.33.]

18.35. (a) Show that if $\alpha = (1,2)$, $\beta = (1,2,\ldots,n)$ are permutations written in the cycle notation then for any $1 < i < n$ we have $(i,i+1) = \beta^{i-1}\alpha(\beta^{i-1})^{-1} = \beta^{i-1}\alpha\beta^{n-i+1}$.

(b) Prove that $\mathbb{S}_n$ is generated by the set $S = \{(1,2),(1,2,3,\ldots,n)\}$. [Hint: Use part (a) and the result of the preceding exercise.]

19 Cyclic groups

Cyclic groups are the simplest groups to understand, and they appear as subgroups of any group. We collect their main properties here in one place, for ease of reference.

Recall that a group is called *cyclic* if it is generated by a single element. In multiplicative notation, if x is a generator, then the cyclic group generated by x is the set

$$\langle x \rangle = \{x^k : k \in \mathbb{Z}\}$$

of all integer powers of x. This could be a cyclic subgroup in some larger group, or an abstract cyclic group. There are two cases to be analyzed: either the generator x has infinite order, or not.

If the generator x has infinite order (i.e., $x^k \neq 1$ for all positive integers k) then all the integer powers of x must be distinct, because if $x^j = x^k$ for $j \neq k$ then $x^{j-k} = x^{k-j} = 1$, contradicting the assumption that x has infinite order. Thus the group $\langle x \rangle$ is infinite. In that case, we claim that it is isomorphic to the additive group $\mathbb{Z}$ of integers. An isomorphism is defined by the rule $f(k) = x^k$. This is a bijection, with inverse g defined by $g(x^k) = k$; you can easily check that $f(g(x^k)) = x^k$ and $g(f(k)) = k$ for all k. Since f goes from an additive group to a multiplicative one, we have to check that $f(j + k) = f(j)f(k)$, which is true since $x^{j+k} = x^j x^k$. Since f is a bijection and $f(j + k) = f(j)f(k)$ for all $j, k \in \mathbb{Z}$, it follows that f is an isomorphism, as claimed.

The remaining possibility is that x has finite order, say x has order n for some positive integer n. Then the set of powers

$$\langle x \rangle = \{x^k : k \in \mathbb{Z}\} = \{x^k : k = 0, 1, \ldots, n-1\}$$

collapses to a *finite* set since $x^n = 1$. It is customary to denote this finite group by C_n. People often write $C_n = \langle x : x^n = 1 \rangle$ to indicate that C_n is generated by an element x satisfying the relation $x^n = 1$. The relation $x^n = 1$ implies that $x^j = x^k$ in C_n if and only if $j \equiv k \pmod{n}$. Thus, powers of x are multiplied in the group C_n by adding their exponents modulo n. That is, we have

$$x^a x^b = x^c \text{ in } C_n \text{ where } c = \text{res}_n(a + b).$$

Recall that $[a]+[b] = [a+b]$ in the additive group $\mathbb{Z}_n$. Equivalently, $[a]+[b] = [c]$ in $\mathbb{Z}_n$ where $c = \text{res}_n(a + b)$ is the residue modulo n of $a + b$. So the bijection $f : \mathbb{Z}_n \to C_n$ defined by the rule $f([k]) = x^k$ is an isomorphism, because $f([j] + [k]) = f([j])f([k])$ for all $k = 0, 1, \ldots, n-1$.

To summarize, we have proved the following important result.

19.1 Theorem. *Any infinite cyclic group is isomorphic to the additive group $\mathbb{Z}$ of integers. Any finite cyclic group is isomorphic to the additive group $\mathbb{Z}_n$ of integers modulo n, for some positive integer n.*

Since isomorphism is transitive, this means also that all infinite cyclic groups are isomorphic, and all finite cyclic groups of the same order are isomorphic. So we now understand all cyclic groups, up to isomorphism.

The above theorem gives important information about all groups, because if G is any group and $x \in G$ then $H = \langle x \rangle$ is a cyclic subgroup of G, hence is isomorphic to $\mathbb{Z}$ or to some $\mathbb{Z}_n$. Furthermore, any result proved about cyclic groups applies equally well to the cyclic subgroups found in any group.

19.2 Theorem. *Let x be an element of a group G. If x has order n then x^k has order n/g, where $g = \gcd(n, k)$. If x has infinite order then x^k has infinite order.*

Proof. Suppose $|x| = n$. Because $\langle x \rangle$ is isomorphic to the additive group $\mathbb{Z}_n$, with x^k corresponding to $[k]$, it suffices to show that the order of $[k]$ in $\mathbb{Z}_n$ is n/g. By definition of order, the order of $[k]$ is the least positive integer m such that $m[k] = [0]$; i.e., the least positive integer m such that $[mk] = [0]$. If $k > 0$ then mk must be the least common multiple of n, k: $mk = \operatorname{lcm}(n, k)$. Hence $m = \operatorname{lcm}(n, k)/k$. Now a theorem from basic number theory says that $nk = \gcd(n, k) \operatorname{lcm}(n, k)$, so $m = \operatorname{lcm}(n, k)/k = n/\gcd(n, k)$, and the proof is finished in case $k > 0$. If $k = 0$ then the result is trivial since $g = n$ and the identity has order 1. If $k < 0$ then we can use the fact that the order of x is the same as the order of x^{-1}. This implies that the order of x^k is the same as the order of x^{-k}. Since $\gcd(n, k) = \gcd(n, -k)$, the stated formula works in the negative case as well.

Finally, if x has infinite order then so does x^k, because assuming that x^k has finite order leads immediately to a contradiction. □

Recall that two integers k, n are said to be *relatively prime* if $\gcd(n, k) = 1$. The Euler phi function $\varphi(n)$ is defined to be the number of integers k in the range $1 \leq k \leq n - 1$ such that k is relatively prime to n. It is easily proved that:

(i) If m, n are relatively prime then $\varphi(mn) = \varphi(m)\varphi(n)$.
(ii) If p is prime then $\varphi(p^t) = p^t - p^{t-1}$.

These two properties can be used to calculate $\varphi(n)$ whenever we can find the prime factorization of n.

The next result follows easily from the previous theorem.

19.3 Corollary. *(a) The order of any element of C_n divides n.*

(b) If $x \in C_n$ has order n then $\langle x^j \rangle = \langle x^k \rangle \iff \gcd(n,j) = \gcd(n,k)$ and $|x^j| = |x^k| \iff \gcd(n,j) = \gcd(n,k)$.

(c) If $x \in C_n$ has order n then x^k generates C_n if and only if $\gcd(n,k) = 1$. So C_n has $\varphi(n)$ distinct generators.

(d) $[k]$ in $\mathbb{Z}_n$ generates $\mathbb{Z}_n$ if and only if $\gcd(n,k) = 1$. So the additive group $\mathbb{Z}_n$ has $\varphi(n)$ distinct generators.

Proof. This is left to you as an exercise. □

We also get information about the multiplicative groups $\mathbb{Z}_n^\times$ of units, whenever they are cyclic.

19.4 Corollary. *If the multiplicative group $\mathbb{Z}_n^\times$ is cyclic, then it is isomorphic to the additive group $\mathbb{Z}_{\varphi(n)}$ and it has $\varphi(\varphi(n))$ generators.*

Proof. We already proved that $\mathbb{Z}_n^\times = \{[k] : \gcd(n,k) = 1\}$, so $|\mathbb{Z}_n^\times| = \varphi(n)$. If it is cyclic then it must be isomorphic to $\mathbb{Z}_{\varphi(n)}$ by the previous theorem. Part (c) of the preceding corollary says that there are $\varphi(\varphi(n))$ generators. □

Of course, this result raises the question: for which values of n is the multiplicative group $\mathbb{Z}_n^\times$ cyclic? Note that $\mathbb{Z}_n^\times$ is cyclic if and only if an element of order $\varphi(n)$ exists in the group. Such elements are called primitive roots.

19.5 Definition. A congruence class $[a] \in \mathbb{Z}_n^\times$ is a *primitive root* in $\mathbb{Z}_n^\times$ if it has order $\varphi(n)$; i.e., if it generates the group. Furthermore, an integer a is called a *primitive root modulo n* if its residue class $[a]$ is a primitive root in $\mathbb{Z}_n^\times$.

Wikipedia has a nice article on primitive roots in modular arithmetic, for those who wish to know more. The answer to our question is provided by the following theorem from classical number theory.

19.6 Theorem (primitive roots theorem)**.** *There is a primitive root in the multiplicative group $\mathbb{Z}_n^\times$ if and only if $n = 2$, 4, p^t or $2p^t$ where p is an odd prime.*

We leave the proof, which is elementary but somewhat time-consuming, to the number theory textbooks. Primitive roots are used in cryptography, so the theorem has practical applications.

We now return to the study of cyclic groups in general.

19.7 Theorem. *Every subgroup of an infinite cyclic group is infinite cyclic. Every subgroup of a finite cyclic group is finite cyclic.*

Proof. It suffices to prove the first statement for the additive cyclic group $\mathbb{Z}$, since any infinite cyclic group is isomorphic to $\mathbb{Z}$. Let G be any subgroup of $\mathbb{Z}$. If G is the trivial subgroup $\{0\}$ then we are done, as $\{0\} = \langle 0 \rangle$ is cyclic. Otherwise, G must have at least one positive element, so by the well-ordering principle of natural numbers, the set of positive elements of G has a least member, say k. Then we claim that $G = \langle k \rangle$. Clearly $G \supset \langle k \rangle$ by closure, so it suffices to prove the reverse inclusion. Let $m \in G$. By the division algorithm, there are unique integers q, r such that $m = qk + r$ and $0 \leq r < k$. Then $r = m - qk \in G$ by closure, since $m, k \in G$. Since k is the *least* positive integer in G, it follows that $r = 0$. Hence $m = qk$ and thus $m \in \langle k \rangle$. This proves the reverse inclusion that establishes the equality $G = \langle k \rangle$, which implies that $G = k\mathbb{Z}$ is the subgroup consisting of all multiples of k. This is infinite cyclic.

It suffices to prove the second claim for the additive group $\mathbb{Z}_n$, since any cyclic group of order n is isomorphic to $\mathbb{Z}_n$. We can use exactly the same argument as above to see that if G is any subgroup of $\mathbb{Z}_n$ then G is either the trivial subgroup or $G = \langle [k]] \rangle$, where k is the least positive element of G, where we represent elements of $\mathbb{Z}_n$ by their residues $0, 1, 2, \ldots, n-1$. □

19.8 Corollary. *Let G be a finite cyclic group of order n. Then the order of any subgroup must divide n. Furthermore, G has precisely one subgroup of order k for every divisor k of n.*

Proof. If H is a subgroup of G then H is cyclic by the previous theorem. Thus H is generated by some power x^k where x is a generator of G. We proved in Theorem 19.2 that the order of x^k is n/g where $g = \gcd(n, k)$, so $|x^k|$ is a divisor of n. Since the order of x^k is the same as the order of the subgroup it generates, the order of H is a divisor of n.

If $k \mid n$ then $\langle x^{n/k} \rangle$ is a subgroup of order k, since $|x^{n/k}| = k$. Furthermore, this is the only subgroup of order k. □

19.9 Corollary. *For any positive divisor d of n, the number of elements of order d in C_n or $\mathbb{Z}_n$ is $\varphi(d)$.*

Proof. Since $C_n \cong Z_n$, it suffices to prove this for C_n. By the previous corollary, there is precisely one subgroup of order d, say $\langle y \rangle \cong C_d$ for some $y \in C_n$, where y has order d. Since this is the unique subgroup of order d, it must contain every element of C_n of order d. By Corollary 19.3(c), there are precisely $\phi(d)$ generators of C_d, and they are the elements of order d in C_n. □

19.10 Example. The number of elements of order 20 in the cyclic group $\mathbb{Z}_{900}$ is $\varphi(20) = \varphi(4 \cdot 5) = (4-2)(5-1) = 8$. The unique subgroup of order 20 is the subgroup $\langle [900/20] \rangle = \langle [45] \rangle$.

Exercises

19.1. Compute the following:
   (a) The number of generators of $\mathbb{Z}_{20}$, $\mathbb{Z}_{100}$, and $\mathbb{Z}_{1000}$.
   (b) The number of generators of C_{20}, C_{100}, and C_{1000}.
   (c) The order of $[235]$ in $\mathbb{Z}_{1000}$ and the order of x^{235} in the abstract cyclic group $C_{1000} = \langle x : x^{1000} = 1 \rangle$.
   (d) The number of elements of $\mathbb{Z}_{1000}$ or C_{1000} of order 40.

19.2. Compute the following:
   (a) The order of the multiplicative group $\mathbb{Z}_{250}^\times$.
   (b) The number of generators of the multiplicative group $\mathbb{Z}_{250}^\times$.
   (c) The number of elements of $\mathbb{Z}_{250}^\times$ of order 25.

19.3. Prove Corollary 19.3.

19.4. Compute the following:
   (a) The order of the multiplicative group $\mathbb{F}_{499}^\times$.
   (b) The number of generators of the multiplicative group $\mathbb{F}_{499}^\times$.
   (c) The number of elements of $\mathbb{F}_{499}^\times$ of order 41.

19.5. The fact that the multiplicative group $\mathbb{F}_p^\times$ of units in a finite field of p elements (where p is a prime) is always cyclic is of great importance in public-key cryptography. But actually finding a generator is sometimes difficult. Try to find a generator of the group $\mathbb{F}_{499}^\times$. You may wish to use a computer to aid your search.

19.6. The Elgamal cryptosystem works as follows, in order to setup secure communication from Bob (or anyone) to Alice.[4]
   (a) First Alice chooses a very large prime p and finds a generator $[g]$ of the cyclic group $\mathbb{F}_p^\times$. She chooses an integer $1 \le k \le p-1$ at random and computes $[h] = [g]^k$ in the multiplicative group $\mathbb{F}_p^\times$. She publishes

[4] Cryptographic tradition demands that the two parties are named Alice and Bob. Also by tradition, the evil attacker trying to decrypt the secret messages is named Eve.

the data $K_A = (p, g, h)$ as her *public-key* and keeps the *private-key* k secret.[5]

(b) To encrypt a secret message $[x] \in \mathbb{F}_p^\times$ to send to Alice, Bob chooses a random integer $1 \leq m \leq p-1$ and computes $c_1 = [g]^m$ and $c_2 = [x] \cdot [h]^m$ in the group $\mathbb{F}_p^\times$. The encrypted message that he sends to Alice is the pair (c_1, c_2).

(c) When Alice receives the encrypted ciphertext message (c_1, c_2), she computes the product $(c_1^k)^{-1} \cdot c_2$ in the group $\mathbb{F}_p^\times$, using her secret key k. Note that Alice knows about the extended Euclidean algorithm, so computing a modular inverse is no problem for her.

Prove that this cryptosystem works; that is, prove that $(c_1^k)^{-1} \cdot c_2 = [x]$ in the multiplicative group $\mathbb{F}_p^\times$.

19.7. In order for an evil attacker Eve to break an Elgamal cryptosystem, she needs to solve the *discrete logarithm problem*, which is the problem of finding an exponent k such that $g^k = h$ in a cyclic group C_n, where g is a generator of the group. Write $k = \log_g h$ to mean that $g^k = h$ in C_n. Note that the value $k = \log_g h$ is only defined modulo n. Put on your evil attacker hat, and find the following discrete logarithms:

(a) $\log_x x^{27}$ in $C_{10} = \langle x : x^{10} = 1 \rangle$.

(b) $\log_{x^9} x^{271}$ in $C_{50} = \langle x : x^{50} = 1 \rangle$.

(c) $\log_{[2]}[9]$ in $\mathbb{Z}_{11}^\times$. Use trial and error.

(d) $\log_{10} 37$ in $\mathbb{Z}_{47}^\times$. You may want to seek help from a computer.

Remark. It is truly remarkable that cyclic groups can be used to construct a cryptosystem sufficiently secure that it is used worldwide for secure internet transmissions. It is believed that solving the discrete logarithm problem in the cyclic group $\mathbb{F}_p^\times$ is so hard a problem that even a supercomputer would take billions of years to finish, assuming that the prime p is sufficiently[6] large. Unfortunately, this belief remains unproven.

[5] This is a one-way system, in the sense that it can be used only by Bob (or anyone) to send secret messages to Alice. If Bob wishes to receive secret messages, then he must setup his own public-key K_B by following the same steps as Alice. After Bob publishes his public-key, Alice (or anyone) can use it to send secret messages to Bob.

[6] On the order of 1500 decimal digits for current technology.

Chapter 6

Quotients and Homomorphisms

20 Cosets

We now introduce cosets, which will be used to prove Lagrange's theorem and to construct quotient groups. Cosets are a fundamental concept in group theory.

20.1 Definition. If H is a subgroup of a group $(G, *)$ and $a \in G$ then we write $a * H = \{a * x : x \in H\}$ and $H * a = \{x * a : x \in H\}$. These sets are called *left* and *right cosets* of H in G, respectively.

If $(G, \cdot)$ is a multiplicative group then we write aH and Ha for the left and right cosets of H, whereas if $(G, +)$ is a additive group then we write them as $a + H$ and $H + a$ instead.

20.2 Examples. 1. Let $H = \{1, r, \cdots, r^{n-1}\}$ be the rotation subgroup of the dihedral group $\mathbb{D}_n$. Then $dH = \{d, dr, \ldots, dr^{n-1}\}$, where $d \in \mathbb{D}_n$ is any reflection, and $Hd = \{d, rd, \ldots, r^{n-1}d\}$. So $dH = Hd$. Furthermore, if $a \in H$ is any rotation, then $aH = H$ and $Ha = H$.

2. Let $H = 2\mathbb{Z} = \{2k : k \in \mathbb{Z}\}$ be the subgroup of even integers in the additive group $(\mathbb{Z}, +)$. Then $1 + H = 1 + 2\mathbb{Z} = \{2k + 1 : k \in \mathbb{Z}\} = H + 1$ is the set of all odd integers. Furthermore, $m + H = 1 + H$ for any odd integer m, and $m + H = 0 + H = H$ for any even integer m.

3. Let $H = n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$ be the subgroup of multiples of n in the additive group $(\mathbb{Z}, +)$. Then $a + H = a + n\mathbb{Z} = a + \{a + nk : k \in \mathbb{Z}\}$ is the set of all integers which are congruent to a modulo n. Note that $a + H = b + H$

if and only if $a \equiv b \pmod{n}$.

4. Let $H = \{[0], [2], [4]\}$ in the additive group $G = \mathbb{Z}_6$. It is easy to check that H is a subgroup of G. Then $H = [0] + H = [2] + H = [4] + H$. Also, $[1] + H = [3] + H = [5] + H = \{[1], [3], [5]\}$.

5. Consider $G = \mathbb{Z}_9^\times = \{[1], [2], [4], [5], [7], [8]\}$, the multiplicative group of units in the ring $\mathbb{Z}_9$. Let $H = \{[1], [8]\}$ in Then H is a subgroup of G, and $H = [1]H = [8]H$, $[2]H = [7[H] = \{[2], [7]\}$, and $[4]H = [5]H = \{[4], [5]\}$. These are all the left cosets.

It is annoying to always have to distinguish between the multiplicative and additive notation, so from now on we adopt the convention that all groups will be multiplicative groups, unless stated otherwise. We leave it to reader to make the necessary adjustments in notation for additive groups.

20.3 Definition (equivalence relations induced by a subgroup)**.** Let H be a subgroup of a given group G. Define a relation $\sim_L$ on G by: $a \sim_L b$ whenever $a^{-1}b \in H$. Define another relation $\sim_R$ on G by: $a \sim_R b$ whenever $ba^{-1} \in H$. The relations $\sim_L, \sim_R$ are called *left and right equivalence*.

Note that the relations $\sim_L$ and $\sim_R$ depend on both the group G and the chosen subgroup H.

20.4 Lemma. *Both relations $\sim_L$ and $\sim_R$ are equivalence relations on G. The equivalence classes for $\sim_L$ are the left cosets of H and the equivalence classes for $\sim_R$ are the right cosets of H.*

Proof. The proof that $\sim_L$ and $\sim_R$ are equivalence relations is an easy exercise.

We prove the claim about left cosets. Let $a, b \in G$. We have $a \sim_L b$ if and only if $a^{-1}b \in H$ if and only if $b \in aH$. Thus $[a] = \{b \in G : a \sim_L b\} = aH$. This proves that the equivalence class of a is equal to the left coset aH. The proof of the claim about right cosets is similar. □

In general cosets (left or right) are just subsets of the group G, and are not necessarily subgroups. From now on we choose to work only with *left* cosets for the sake of having a definite choice, but it should be understood that everything we prove about left cosets applies equally well to right cosets.

20.5 Lemma (properties of left cosets)**.** *Let G be a group and H a subgroup of G. Let $a, b \in G$. Then:*

(a) $a \in aH$.

(b) $aH = H$ *if and only if* $a \in H$.
(c) $aH = bH$ *if and only if* $a \sim_L b$.
(d) $aH = bH$ *if and only if* $b = ax$ *for some* $x \in H$.
(e) *Any pair of left cosets of* H *are either disjoint or coincide.*

Proof. Exercise. □

20.6 Definition. Let G be a group and H any subgroup of G. We write $G/H = \{aH \mid a \in G\}$ for the quotient set $G/\!\sim_L$ of all left cosets of H. This is called the *quotient* of G by H. We read the notation G/H as "G mod H."

By definition, G/H is a set of sets. The elements of G/H are the left cosets of H, which by definition are certain subsets of G. Since $\sim_L$ is an equivalence relation on the set G, it follows from the fundamental theorem of equivalence relations that *G can be expressed as the disjoint union of its distinct left cosets.* Those distinct left cosets are the elements of the quotient set G/H.

20.7 Definition. The number of distinct elements of the set G/H (i.e., its cardinality as a set), which is the same as the number of distinct left cosets, is denoted by either $|G/H|$ or $[G : H]$, and is called the *index* of H in G. It can be infinite, but it must be a finite number if G is a finite group.

Note that as a varies over G, there will in general be a lot of repetition in the left cosets aH. When computing the index, you count just the distinct cosets.

20.8 Examples. 1. Let $G = \mathbb{S}_n$ and let $H = \mathbb{A}_n$. Then there are just two left cosets: $G/H = \mathbb{S}_n/\mathbb{A}_n = \{\mathbb{A}_n, \alpha\mathbb{A}_n\}$, where α is any odd permutation. This is just the splitting of all permutations into the even ones ($\mathbb{A}_n$) and the odd ones ($\alpha\mathbb{A}_n$). So $[\mathbb{S}_n : \mathbb{A}_n] = 2$.

2. Take $G = \mathbb{Z}$ (under addition) and $H = n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$. The left cosets have the form $a + H = a + n\mathbb{Z} = \{a + nk : k \in \mathbb{Z}\}$ for various $a \in \mathbb{Z}$. Moreover, $a + H = b + H$ if and only if $-a + b \in H$, i.e., if and only if $a \equiv b \pmod{n}$. The distinct cosets are the $a + H$ where $0 \le a \le n - 1$. Note that $a + n\mathbb{Z} = [a]$, the congruence class determined by a. So the set G/H of left cosets is

$$G/H = \mathbb{Z}/n\mathbb{Z} = \{n\mathbb{Z}, 1 + n\mathbb{Z}, \ldots, n - 1 + n\mathbb{Z}\} = \{[0], [1], \ldots, [n - 1]\} = \mathbb{Z}_n.$$

So the index is $[\mathbb{Z} : n\mathbb{Z}] = n$. Note that we have reconstructed $\mathbb{Z}_n$ in terms of cosets. To say it another way, the coset construction is a vast generalization of the construction of $\mathbb{Z}_n$ given previously.

3. Let $G = \mathrm{O}(n)$ for $n \geq 2$ and let $H = \mathrm{SO}(n)$. Recall that the determinant of any orthogonal matrix is ± 1. By definition, elements of $\mathrm{SO}(n)$ are proper orthogonal matrices (of determinant 1). So we have $G/H = \mathrm{O}(n)/\mathrm{SO}(n) = \{\mathrm{SO}(n), A \cdot \mathrm{SO}(n)\}$, where A is any improper orthogonal matrix. This reflects the fact that the improper orthogonal matrices are those of determinant -1. So the index is $[\mathrm{O}(n) : \mathrm{SO}(n)] = 2$.

4. Consider $G = \mathbb{Z}_9^\times = \{[1], [2], [4], [5], [7], [8]\}$, the multiplicative group of units in the ring $\mathbb{Z}_9$. Then $H = \{[1], [8]\}$ is a subgroup of G. The left cosets of H in G are $H = [1]H$, $[2]H = \{[2], [7]\}$, and $[4]H = \{[4], [5]\}$. So the index is $[G : H] = 3$. Notice that in this case $|G|/|H| = 6/2 = 3$.

The next result is a fundamental result about finite groups, with numerous applications. In particular, it is heavily used in the design of public-key cryptosystems.

20.9 Theorem (Lagrange's Theorem)**.** *Let G be a finite group and H a subgroup of G. Then $|G| = [G : H] \cdot |H|$. In words, the order of G is the order of the subgroup H times its index in G.*

Proof. Since G is a finite set the left cosets must be finite sets as well. Moreover, *all the left cosets must have the same cardinality.* This is because H is in bijective correspondence with aH, for any $a \in G$. The correspondence is given by the map $x \mapsto ax$ ($x \in H$). So all the left cosets have the same cardinality as the subgroup H and therefore the same cardinality as one another. (Note that $H = 1H$ is also a left coset.) There are precisely $[G : H]$ distinct left cosets. And G is the disjoint union of the left cosets, since cosets are equivalence classes. So if $m = [G : H]$ then we have the disjoint union

$$G = a_1 H \cup a_2 H \cup \cdots \cup a_m H$$

where these are the *distinct* left cosets. Thus $|G| = m|H|$, as required. □

If a is an element of a group G, recall that its order $|a|$ is the least positive integer k such that $a^k = 1$. Infinite groups can have elements of infinite order, but if G is a finite group then every element has finite order. Lagrange's theorem tells us for example that if G is finite then the order of all its subgroups and all its elements must divide the group order $|G|$.

20.10 Corollary. *Suppose G is a finite group.*

(a) *If H is a subgroup of G then $|H|$ must divide $|G|$.*
(b) *If $a \in G$ then its order $|a|$ must divide $|G|$.*
(c) *Any group of order p, where p is prime, must be cyclic.*
(d) *If H is a subgroup of G then $[G : H] = |G|/|H|$.*

Proof. (a) and (d) are obvious consequences of Lagrange's Theorem.

(b) Let $H = \langle a \rangle$. Then $|H| = |a|$ (the cardinality of H is equal to the order of a). The statement in part (b) now follows from part (a).

(c) Let G be a group of order p, where p is prime. Since $p > 1$ we know that G must contain at least one element x which is different from the identity. Then $|x|$ must divide p by part (b). Moreover, $|x| > 1$ since the only element of order 1 in G is the identity element. The only divisors of p are p and 1, so it follows that $|x| = p$. Hence $G = \langle x \rangle$ and G is cyclic. □

Lagrange's theorem can also be applied to number theory, to give easy proofs of both Euler's theorem and Fermat's little theorem. That both of these famous number theoretic results follow so easily from Lagrange's theorem illustrates the power of the abstract approach.

Recall that Euler's phi-function $\varphi(n)$ is defined to be the number of integers a such that $1 \leq a \leq n-1$ and $\gcd(a, n) = 1$. Thus the multiplicative group $\mathbb{Z}_n^\times$ of units has order $\varphi(n)$. Recall also that the ring $\mathbb{Z}_n = \{[a] : a \in \mathbb{Z}\} = \{[0], [1], \ldots, [n-1]\}$, where $[a]$ is the congruence class of a, given by $[a] = \{b \in \mathbb{Z} : a \equiv b \pmod{n}\}$.

20.11 Corollary. *Let a be any integer.*
(a) *(Euler's theorem) For any positive integer n such that a, n are relatively prime, we have $a^{\varphi(n)} \equiv 1 \pmod{n}$.*
(b) *(Fermat's little theorem) For any prime p such that p does not divide a, we have $a^{p-1} \equiv 1 \pmod{p}$.*

Proof. (a) Let $G = \mathbb{Z}_n^\times$ be the multiplicative group of units in the finite ring $\mathbb{Z}_n$. We have $|G| = \varphi(n)$, by definition of $\varphi(n)$. If a, n are relatively prime then $\gcd(a, n) = 1$ and thus $[a] \in \mathbb{Z}_n^\times$. By the first corollary to Lagrange's theorem, the order $|[a]|$ must divide $|G| = \varphi(n)$. Suppose $|[a]| = k$. Then $[a]^k = [1]$ holds in $G = \mathbb{Z}_n^\times$, where k divides $\varphi(n)$. There is some $m \in \mathbb{Z}$ such that $\varphi(n) = km$. Now the equality $[a]^k = [1]$ implies $([a]^k)^m = [1]^m = [1]$. In other words, $[a]^{km} = [a]^{\varphi(n)} = [1]$ holds in $G = \mathbb{Z}_n^\times$. This implies (by definition of congruence class multiplication) that the equality $[a^{\varphi(n)}] = [1]$ holds in $G = \mathbb{Z}_n^\times$. To finish, recall that this equality in $\mathbb{Z}_n$ is equivalent to the desired congruence, so the proof is complete.

(b) We can repeat the argument with p in place of n, noting that $\varphi(p) = p-1$. Note that p does not divide a if and only if $\gcd(a,p) = 1$. Alternatively, we can just note that the result in (b) is just a special case of that in (a). $\square$

20.12 Definition. Let G be a group and H a subgroup of G. Any set S of elements of G such that:

(a) $G/H = \{aH : a \in S\}$,

(b) for all $a, b \in S$, $a \neq b$ implies $aH \neq bH$

is called a set of left coset *representatives* of the quotient set G/H.

Picking a set of coset representatives is the same as choosing an element from each distinct coset. If S is such a set, then we can write G as a *disjoint* union: $G = \bigsqcup_{a \in S} aH$. If the index $[G : H] = n$ is finite, then we can write this disjoint union as: $G = a_1H \sqcup a_2H \sqcup \cdots \sqcup a_nH$, where $S = \{a_1, a_2, \ldots, a_n\}$.

One of the left cosets (say a_1H) is always the same as the subgroup H itself, and we may choose the identity element 1 as its coset representative. If we have made that choice, then the above becomes $G = H \sqcup a_2H \sqcup \cdots \sqcup a_mH$, where $S = \{1, a_2, \ldots, a_n\}$.

In general coset representatives are far from unique. Thus, whenever we define functions in terms of a set of coset representatives, then we must pause to verify that our function is well-defined (independent of the choice of coset representatives). We will see examples later.

Exercises

20.1. Prove that the relation $\sim_L$ defined by a subgroup H (see 20.1) is an equivalence relation on the group G.

20.2. Prove Lemma 20.5.

20.3. Let G be the additive group $\mathbb{Z}_{2n}$. Let $H = \{[2k] \in \mathbb{Z}_{2n} : 0 \leq k < n\}$. Show that $H < G$, and compute the index $[G : H]$.

20.4. Compute the index $[G : H]$ for the following cases:

(a) $H = \{[0], [3]\}$ in $G = (\mathbb{Z}_6, +)$.

(b) $H = \{[0], [10]\}$ in $G = (\mathbb{Z}_{90}, +)$.

(c) $H = \{[1], [4], [13], [16]\}$ in $G = (\mathbb{Z}_{17}^\times, \cdot)$.

(d) $H = \{[1], [7]\}$ in $G = (\mathbb{Z}_{48}^\times, \cdot)$.

20.5. List the left cosets in G/H in part (a) and (c) of the preceding problem.

20.6. Let $G = C_n = \{1, x, x^2, \ldots, x^{n-1}\}$ be the abstract cyclic group of order n, generated by an element x of order n. Suppose that k divides n and let $H = \langle x^k \rangle$ be the subgroup generated by x^k. Describe the left cosets in G/H and compute the index $[G : H]$.

20.7. Describe the distinct left cosets of the additive subgroup $(\mathbb{Z}, +)$ in the additive group $(\mathbb{R}, +)$. In other words, describe the quotient set $\mathbb{R}/\mathbb{Z}$.

20.8. Describe the distinct left cosets of the additive subgroup $(\mathbb{R}, +)$ in the additive group $(\mathbb{C}, +)$. In other words, describe the quotient set $\mathbb{C}/\mathbb{R}$. What is $[\mathbb{C} : \mathbb{R}]$?

20.9. Let $\mathbb{R}^+$ be the set of positive real numbers. Describe the distinct left cosets of the multiplicative subgroup $(\mathbb{R}^+, \cdot)$ in the multiplicative group $(\mathbb{R}^\times, \cdot)$. In other words, describe the quotient set $\mathbb{R}^\times/\mathbb{R}^+$. What is $[\mathbb{R}^\times : \mathbb{R}^+]$?

20.10. Describe the distinct left cosets of $\mathrm{SL}(2)$ in $\mathrm{GL}(2)$.

20.11. If p is a prime, describe the distinct left cosets of $\mathrm{SL}_2(\mathbb{F}_p)$ in $\mathrm{GL}_2(\mathbb{F}_p)$, and compute the index $[\mathrm{GL}_2(\mathbb{F}_p) : \mathrm{SL}_2(\mathbb{F}_p)]$.

20.12. Suppose that H is a subgroup of a group G. Show that if $aH = Ha$ and $bH = Hb$ then $abH = Hab$.

20.13. Suppose that H is a subgroup of a group G. Show that if $aH = Ha$ then $a^{-1}H = Ha^{-1}$.

20.14. Explain why a group G of order 20 has no subgroups or elements of order 3, 7, or 9.

20.15. Show that if G is a group of order n then $x^n = 1$ for every $x \in G$.

20.16. Prove that a group of prime order has no subgroups other than itself and the trivial subgroup.

20.17. Prove that every element except the identity has order p in a group of prime order p.

20.18. In the multiplicative group $\mathbb{F}_{11}^\times$ we have $[2]^2 = [4]$, $[2]^4 = [4]^2 = [5]$, and $[2]^5 = [2]^4 \cdot [2] = [5] \cdot [2] = [10]$. Use one of the corollaries to Lagrange's theorem to explain why this immediately implies that $[2]$ must have order 10 in the group.

20.19. Suppose that H, K are finite subgroups of a group G, and let $m = |H|$, $n = |K|$.

(a) Show that the order $|H \cap K|$ of $H \cap K$ must be a common divisor of m, n.

(b) Show that if m, n are relatively prime then $H \cap K = \{1\}$ is the trivial subgroup.

20.20. Suppose that G is a group of order pq where p, q are distinct primes. Show that if G is not cyclic then every element in G except the identity must have order p or q.

21 Quotient groups

From now on we use multiplicative notation unless indicated otherwise. The results apply equally well to additive groups, but the notation needs to be translated accordingly.

If H is a subgroup of a group G, recall that $G/H = \{aH : a \in G\}$ is the set of all left cosets of H in G. The key question for this section is: when is the set G/H of left cosets a group? Recall that we have previously defined the product of two subsets S, T in a group by $ST = \{xy \mid x \in S, y \in T\}$. In particular, this means that $aH = \{ax \mid x \in H\}$, $Ha = \{xa \mid x \in H\}$, and $(aH)(bH) = \{axby \mid x, y \in H\}$. We want the product $(aH)(bH)$ of two left cosets to always equal another left coset.

21.1 Theorem. *[coset multiplication] Suppose that H is a subgroup of a group G. Then the following are equivalent:*

(a) For any $a, b \in G$, there is some $c \in G$ such that $(aH)(bH) = cH$.
(b) For any $a, b \in G$, $(aH)(bH) = (ab)H$.
(c) $Hb = bH$, for all $b \in G$.

Proof. (a) $\implies$ (b): Since $ab = a1b1 \in (aH)(bH)$, it follows from the given equality $(aH)(bH) = cH$ that $ab \in cH$. Hence $(ab)H = cH$.

(b) $\implies$ (c): Let $b \in G$. Then $(bH)(b^{-1}H) = (bb^{-1})H = 1H = H$; i.e., $bHb^{-1}H = H$. This implies that $bHb^{-1} = bHb^{-1}1 \subset H$, so $bH \subset Hb$. Similarly, $(b^{-1}H)(bH) = (b^{-1}b)H = H$ implies that $b^{-1}Hb \subset H$, so $Hb \subset bH$. We have shown that $bH \subset Hb$ and $Hb \subset bH$, so $Hb = bH$.

(c) $\implies$ (a): Suppose that $Hb = bH$. Then by associativity we have $(aH)(bH) = a(Hb)H = a(bH)H = (ab)(HH) = (ab)H$. □

So multiplication of left cosets always produces another left coset precisely when condition (c) of the previous theorem holds. This leads to the next definition.

21.2 Definition. Let G be a group and H a subgroup of G. We say that H is a *normal subgroup* of G if $Ha = aH$ for every $a \in G$. We will write $H \triangleleft G$ (or $G \triangleright H$) to indicate that H is a normal subgroup of G.

The trivial subgroup $\{1\}$ and the entire group G are always normal subgroups of a group G. The definition says that H is a normal subgroup if and only if every left coset is also a right coset. Thus, *every* subgroup of an abelian group is normal.

21.3 Theorem. *Suppose that H is a subgroup of a group G. The following are equivalent:*

(a) $H \triangleleft G$.

(b) $aHa^{-1} = H$, *for all* $a \in G$.

(c) $aHa^{-1} \subset H$, *for all* $a \in G$.

Proof. (a) $\implies$ (b): If H is normal in G then by definition $aH = Ha$ for all $a \in G$. Multiplying this equality by a^{-1} on the right gives the equality $aHa^{-1} = Haa^{-1}$. But $Haa^{-1} = H1 = H$.

(b) $\implies$ (c): This is clear.

(c) $\implies$ (a): Suppose $aHa^{-1} \subset H$, for all $a \in G$. Right multiply by a to get $aH \subset Ha$. Replacing a by its inverse in the inclusion $aHa^{-1} \subset H$, we get $a^{-1}Ha \subset H$, so by left multiplication by a we get $Ha \subset aH$. We have shown that both $aH \subset Ha$ and $Ha \subset aH$, so $aH = Ha$. This proves (a). □

21.4 Remarks. 1. Note that if H is a subgroup of a group G then aHa^{-1} is also a subgroup of G, for any $a \in G$. This is called a *conjugate* subgroup of H. Note also that $|H| = |aHa^{-1}|$, since the map $x \mapsto axa^{-1}$ defines a bijection from H onto aHa^{-1}. In general, the conjugate subgroup aHa^{-1} can be different from H. Then $H \triangleleft G$ if and only if all the conjugate subgroups of H are equal to H. (We can say that H is *stable under conjugation* when $H \triangleleft G$.)

2. The relation $\triangleleft$ is not transitive. That is, if $H \triangleleft K$ and $K \triangleleft G$ then it is not always true that $H \triangleleft G$.

3. By replacing a by a^{-1} in parts (b), (c) of the preceding theorem, we see that $H \triangleleft G$ is also equivalent to:

(b') $a^{-1}Ha = H$, for all $a \in G$.

(c') $a^{-1}Ha \subset H$, for all $a \in G$.

Condition (c) in the theorem is a closure condition. It says that a subgroup H is normal if and only if it is closed under conjugation. Elements of the form axa^{-1} are called *conjugates* of x.

The following simple observation can often be used to find a normal subgroup of a finite group whose order is an even number. It can also be applied more generally to infinite groups.

21.5 Proposition. *If H is a subgroup of a group G of index 2 then H must be a normal subgroup of G. (In other words, $[G : H] = 2$ implies $H \triangleleft G$).*

Proof. Let $a \in G$. Either $a \in H$ or $a \notin H$, and we consider the two cases separately. Case 1: If $a \in H$ then $aH = H$ and $Ha = H$ by the fact that H is closed under products, so $aH = Ha$. Case 2: If $a \notin H$ then $aH = G - H = \{g \in G \mid g \notin H\}$ since there are just two left cosets, and they are disjoint. Similarly, $Ha = G - H$ for exactly the same reason. Thus $aH = Ha$. Cases 1 and 2 taken together show that $aH = Ha$ for all $a \in G$, so $H \triangleleft G$. □

21.6 Examples. 1. Since $[\mathbb{S}_n : \mathbb{A}_n] = 2$, it follows that the alternating group $\mathbb{A}_n$ is a normal subgroup of the symmetric group $\mathbb{S}_n$, for any n.

2. Since $[\mathrm{O}(2) : \mathrm{SO}(2)] = 2$, it follows that $\mathrm{SO}(2) \triangleleft \mathrm{O}(2)$.

The following is a fundamental result in group theory. It turns out that when $N \triangleleft G$, coset multiplication makes G/N into a group.

21.7 Theorem. *[quotient group] Let N be a normal subgroup of a group G. Then the coset multiplication rule $(aN)(bN) = (ab)N$ (for any $a, b \in G$) is a well-defined binary operation on the set G/N of left cosets. With respect to this multiplication, G/N is a group, with identity element $1N = N$. The inverse of aN is $a^{-1}N$.*

Proof. To show that coset multiplication is well-defined we must show that if $aN = cN$ and $bN = dN$ then $(ab)N = (cd)N$. Clearly $(aN)(bN) = (cN)(dN)$, and by Theorem 21.1, it follows that $(ab)N = (cd)N$. So coset multiplication is well-defined.

Since $(aNbN)cN = (ab)NcN = (ab)cN = a(bc)N = aN(bc)N = aN(bNcN)$, the associative law (G1) holds. Since $(1N)(aN) = aN = (aN)(1N)$ it follows that (G2) holds with $N = 1N$ serving as the identity. Finally, $(aN)(a^{-1}N) = 1N = (a^{-1}N)(aN)$ shows that (G3) holds and $(aN)^{-1} = a^{-1}N$. □

21.8 Definition. When $N \triangleleft G$, the group G/N is called the *quotient group* of G by N. Quotient groups are also called *factor groups*.

21.9 Remark. If G is a finite group then by Lagrange's theorem $|G/N| = |G|/|N|$ (the order of the quotient group G/N equals the cardinality of G divided by the cardinality of N). Of course, by the definition of the index $[G\colon N]$ we also have $|G/N| = [G\colon N]$.

Exercises

21.1. Show that the trivial subgroup is always a normal subgroup of any group G. Also, show that $G \triangleleft G$.

21.2. Show that *every* subgroup of a group G is normal if:

(a) G is abelian.

(b) G is cyclic.

21.3. Prove that the center $Z(G)$ of a group G is a normal subgroup of G.

21.4. Show that if H is a subgroup of order n in a group G and H is the only subgroup of order n, then $H \triangleleft G$.

21.5. Show that a subgroup H of a group G is normal if and only if it satisfies the condition: $ab \in H \iff ba \in H$, for all $a, b \in G$.

21.6. Prove that the intersection of any number of normal subgroups of a group G is a normal subgroup of G.

21.7. Let $\mathbb{D}_4$ be the symmetry group of the square, and let r be the basic rotation. Is the subgroup $H = \{1, r^2\}$ a normal subgroup of D_4? Prove your answer.

21.8. Let $\mathbb{D}_n$ be the dihedral group on n vertices and let R be its rotation subgroup. Prove that $R \triangleleft \mathrm{D}_n$.

21.9. Let $\mathbb{D}_n$ be the dihedral group on n vertices and let $d \in \mathbb{D}_n$ be a reflection. Then $H = \langle d \rangle = \{1, d\}$ is a cyclic subgroup of $\mathbb{D}_n$. Find necessary and sufficient conditions for this subgroup to be normal in $\mathbb{D}_n$, and prove your answer.

21.10. Prove that $\mathbb{A}_n$ is a subgroup of $\mathbb{S}_n$ of index 2 by first proving that there is a bijection from $\mathbb{A}_n$ onto $\alpha\mathbb{A}_n$, where α is any odd permutation.

21.11. Find all normal subgroups of the quaternion group Q. Justify your answer.

21.12. Show that if H is a subgroup of index 2 in a group G then G/H is isomorphic to the additive group $(\mathbb{Z}_2, +)$.

21.13. Show that $\mathrm{SO}(n) \triangleleft \mathrm{O}(n)$ and identify a group that is isomorphic to $\mathrm{O}(n)/\mathrm{SO}(n)$.

21.14. Show that $\mathrm{SL}(n) \triangleleft \mathrm{GL}(n)$ and $\mathrm{GL}(n)/\mathrm{SL}(n)$ is isomorphic to the multiplicative group $\mathbb{R}^\times$ of the field of real numbers.

21.15. Show that the set G of all real 2×2 matrices of the form $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ is a subgroup of $\mathrm{GL}(2)$. Let N be the set of all matrices of the form $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$. Prove that $N \triangleleft G$. (Note that you need to show it is a subgroup as well as prove that it is normal.)

21.16. Show that $n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$ is a normal subgroup of the additive group $\mathbb{Z}$ of integers, and that $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$.

21.17. Prove that if H, K are subgroups of G and one of them is normal in G, then their product HK is a normal subgroup of G.

21.18. Show that if H, K are subgroups of a group G such that $K \triangleleft G$ then $H \cap K \triangleleft H$.

21.19. Give an example to show that there are groups $K < H < G$ such that $K \triangleleft H$ and $H \triangleleft G$ but K is not a normal subgroup of G. In other words, normality is not transitive.

21.20. Is it possible to find a group G in which every subgroup is normal but G is not abelian? If so, exhibit one, otherwise prove that such a group is abelian.

22 Homomorphisms

This is a central concept in group theory that ties everything together. To define it, we momentarily revert to generic notation.

22.1 Definition. Suppose that $(G, *)$ and $(H, \#)$ are groups. A *homomorphism* of groups is a function $f : G \to H$ such that $f(a * b) = f(a)\#f(b)$ for all $a, b \in G$.

In particular, group isomorphisms are bijections with this property, so every isomorphism is a homomorphism. But not conversely, because we do not require that a homomorphism be a bijection.

22.2 Remark. If G and H are additive groups then $f : G \to H$ is a homomorphism if and only if $f(a + b) = f(a) + f(b)$, for all $a, b \in G$. If G is additive and H is multiplicative then $f : G \to H$ is a homomorphism if and only if $f(a + b) = f(a) \cdot f(b)$, for all $a, b \in G$. If G is multiplicative and H is additive then $f : G \to H$ is a homomorphism if and only if $f(a \cdot b) = f(a) + f(b)$, for all $a, b \in G$. Finally, if both G and H are multiplicative then $f : G \to H$ is a homomorphism if and only if $f(a \cdot b) = f(a) \cdot f(b)$, for all $a, b \in G$.

The first two parts of the next result states that a group homomorphism must match up the identity and inverses in the two groups.

22.3 Theorem (properties of homomorphisms)**.** *Let $f : G \to H$ be a homomorphism from a group $(G, *)$ to a group $(H, \#)$. Then:*

(a) $f(e_G) = e_H$; i.e., f maps the identity of G onto the identity of H.

(b) The inverse of $f(a)$ in H is the image of the inverse of a; i.e., if a' is the inverse of a in G then the inverse of $f(a)$ in H is $f(a)' = f(a')$.

(c) If f is a bijection, then $f^{-1} : H \to G$ must also be a homomorphism.

Proof. (a) Let $y = f(e_G)$ be the image of the identity of G. We must show that $y = e_H$, the identity of H. We have $e_H\#f(x) = f(x) = f(e_G * x) = f(e_G)\#f(x) = y\#f(x)$. Thus $e_H\#f(x) = y\#f(x)$ and by right cancellation we conclude that $e_H = y$.

(b) Let $y = f(b)$ where b is the inverse of a in G. Let z be the inverse of $f(a)$ in H. We must show that $y = z$. But $f(a)\#z = e_H$ by definition of inverse, and $f(a)\#y = f(a)\#f(b) = f(a * b) = f(e_G) = e_H$, by part (a) and the defining property of homomorphism. Thus $f(a)\#z = f(a)\#y$. By left cancellation, this implies $z = y$.

(c) If f matches up products and is a bijective correspondence, then its inverse must match up products as well. □

22.4 Remark. By part (a) of the preceding result, a homomorphism between additive groups must take 0 to 0, while a homomorphism between multiplicative groups must take 1 to 1. A homomorphism from an additive group to a multiplicative group must send 0 to 1, while a homomorphism from an multiplicative group to an additive group must send 1 to 0,

Similarly for inverses in part (b). If f is a homomorphism between additive groups then $-f(a) = f(-a)$, whereas if f is a homomorphism between multiplicative groups then $f(a)^{-1} = f(a^{-1})$. If f is a homomorphism from an additive group to a multiplicative group then $f(a)^{-1} = f(-a)$. Finally, if f is a homomorphism from an multiplicative group to an additive group then $-f(a) = f(a^{-1})$.

It quickly becomes tiresome to distinguish between these possibilities. From now on in the general theory, we will always use multiplicative notation for both groups connected by a homomorphism, unless stated otherwise, trusting the reader to make the necessary adjustments in other situations.

22.5 Definition. Let $f : G \to G'$ be a group homomorphism from a group G into a group G'. The *kernel* of f (written as $\ker f$) is the subset of the domain group G defined by $\ker f = \{x \in G \mid f(x) = 1\}$.

Note that if G' is an additive group, then $\ker f = \{x \in G \mid f(x) = 0\}$.

22.6 Theorem (kernels are normal subgroups)**.** *The kernel of a homomorphism $f : G \to G'$ of groups is a normal subgroup of the domain group G.*

The proof is an easy exercise.

22.7 Definition. Let G be a group and K a normal subgroup of G. Then we have a homomorphism $\pi : G \to G/K$ defined by the rule $\pi(a) = aK$. It is surjective. This particular homomorphism is called the *canonical homomorphism* or the *canonical quotient map.*

One can check (exercise) that the canonical homomorphism actually is a homomorphism.

22.8 Theorem (first isomorphism theorem)**.** *Let G, G' be groups and $f : G \to G'$ a group homomorphism. Set $K = \ker f$ and $I = \operatorname{im} f = f(G)$.*

Then $f = \iota \circ \overline{f} \circ \pi$, where $\iota : I \to G'$ is the inclusion map (coming from the inclusion $I \subset G'$), $\pi : G \to G/K$ is the canonical homomorphism, and $\overline{f} : G/K \to I$ is an isomorphism, given by the rule $\overline{f}(aK) = f(a)$ for $a \in G$. (In particular, $G/K \cong I$.)

This theorem is also known as the *fundamental theorem of homomorphisms*. The factorization of the homomorphism f in the theorem may be pictured by the following commutative diagram:

$$\begin{array}{ccc} G & \xrightarrow{\quad f \quad} & G' \\ {\scriptstyle\pi}\downarrow & & \uparrow{\scriptstyle\iota} \\ G/K & \xrightarrow{\quad \overline{f} \quad} & I \end{array}$$

and the diagram is a very convenient way to remember the theorem. In the diagram, there are two routes from G to G'. The theorem says these two routes are the same: $f = \iota \circ \overline{f} \circ \pi$, and that the induced map $\overline{f}$ is an isomorphism. (The latter fact is most often used in practice.)

Proof. The map $\iota : I \to G'$ is defined by the rule $\iota(y) = y$ for any $y \in I$. The map π has previously been defined, and the definition of $\overline{f}$ is given in the statement of the theorem. We must show that $f = \iota \circ \overline{f} \circ \pi$; in other words, we must show that

$$f(x) = (\iota \circ \overline{f} \circ \pi)(x) = \iota(\overline{f}(\pi(x)))$$

for all x in G. Well, let $x \in G$. Then by the definitions of the maps we have $\iota(\overline{f}(\pi(x))) = \iota(\overline{f}(xK)) = \iota(f(x)) = f(x)$. This proves the desired equality.

Let us check that the map $\overline{f}$ is a *well-defined* function; i.e., the value $\overline{f}(aK)$ does not depend on the choice of coset representative. Suppose that $aK = bK$ for two elements a, b of G. We have to show that $\overline{f}(aK) = \overline{f}(bK)$. Well, by definition of the map $\overline{f}$ we have $\overline{f}(aK) = f(a)$ and $\overline{f}(bK) = f(b)$, so we have to prove that $f(a) = f(b)$. This follows from the equality $aK = bK$, which implies that $a^{-1}b \in K$, so there exists some $x \in K$ such that $a^{-1}b = x$, so $b = ax$, so $f(b) = f(ax) = f(a)f(x) = f(a)1 = f(a)$ as desired. This proves that the map $\overline{f}$ is well-defined.

It remains to prove that the induced map $\overline{f} : G/K \to I$ is an isomorphism. First, note that $\overline{f}$ is a homomorphism, since $\overline{f}(aK \cdot bK) = \overline{f}(abK) = f(ab) = f(a)f(b) = \overline{f}(aK)\overline{f}(bK)$, for any $a, b \in G$. It is injective since if aK lies in the kernel of $\overline{f}$, then $\overline{f}(aK) = f(a)$ is the identity element of G', so $a \in K$, so $aK = K$. This proves that the kernel of $\overline{f}$ is just the identity

coset K of G/K, so $\overline{f}$ is injective as claimed. Finally, $\overline{f}$ is surjective onto I by its definition. We have shown that $\overline{f}$ is a bijective homomorphism. Thus it is an isomorphism. The proof is complete. □

22.9 Examples. 1. For $n \geq 2$, there is a homomorphism from the symmetric group $\mathbb{S}_n$ to the multiplicative group $\{1, -1\} = \mathbb{Z}^\times$ given by the rule: $\alpha \mapsto \operatorname{sgn}(\alpha)$. This homomorphism is surjective (i.e., the image is $\{1, -1\}$). Its kernel is the set $\mathbb{A}_n$ of even permutations, since we know that $\operatorname{sgn}(\alpha) = 1$ if and only if α is even. By the theorem, it follows that $\mathbb{S}_n/\mathbb{A}_n \cong \mathbb{Z}^\times$ is cyclic of order 2.

2. Let n be a positive integer. Recall that we write $\mathbb{Z}_n$ for the additive group of integer residues modulo n. Consider the homomorphism (of additive groups) $\mathbb{Z} \to \mathbb{Z}_n$ given by $a \to \overline{a}$, where $\overline{a}$ is the residue of a mod n. This is a homomorphism since $\overline{a + c} = \overline{a} + \overline{c}$ for all $a, c \in \mathbb{Z}$. The kernel of this homomorphism is the subgroup $n\mathbb{Z}$ of all multiples of n. It is clear that the homomorphism is surjective onto $\mathbb{Z}_n$. Thus we have a group isomorphism $\mathbb{Z}/(n\mathbb{Z}) \cong \mathbb{Z}_n$. This shows that the additive group $\mathbb{Z}_n$ is, in fact, a quotient group. (Note: To be sure, $\mathbb{Z}$ and $\mathbb{Z}_n$ are actually rings when we consider the two operations of addition and multiplication together, but in the isomorphism above we are for the moment ignoring the multiplication and just paying attention to the additive group structure.)

The next result states that every normal subgroup arises as the kernel of some homomorphism.

22.10 Theorem (normal subgroups are kernels)**.** *If K is a given normal subgroup of a group G, then there exists a homomorphism $f : G \to G'$ (for some group G') such that $K = \ker f$.*

Proof. We set $G' = G/K$ and we let f be the canonical homomorphism. Then its kernel is K, so the proof is complete. □

The following two results are known as the second and third isomorphism theorems for groups. Along with the first isomorphism theorem (Theorem 22.8) these results describe fundamental properties of quotient groups and homomorphisms.

22.11 Theorem (second isomorphism theorem)**.** *If $K \triangleleft G$ and $H < G$ then:*

(a) $H \cap K \triangleleft H$.
(b) The product set $HK = \{xy \colon x \in H, y \in K\}$ is a subgroup of G.
(c) $H/(H \cap K) \cong HK/K$.

This follows from the first isomorphism theorem. The proof is an exercise.

22.12 Theorem (third isomorphism theorem)**.** *If $H \triangleleft G$, $K \triangleleft G$, and $H \subset K$ then:*

(a) $(K/H) \triangleleft (G/H)$.
(b) $(G/H)/(K/H) \cong G/K$.

This also follows from the first isomorphism theorem. The proof is an exercise.

We finish our discussion of quotients with an easy application of the first isomorphism theorem, to classify all cyclic groups.

22.13 Theorem (classification of cyclic groups)**.** *Let G be a cyclic group. Then either G is isomorphic to the additive group $\mathbb{Z}$ of integers or G is isomorphic to the additive group $\mathbb{Z}_n$ of residues modulo n, for some n.*

Proof. Let $G = \langle a \rangle$. The integers $\mathbb{Z}$ form a group under addition, and the (surjective) map $f : \mathbb{Z} \to G$ defined by the rule $k \to a^k$ is a group homomorphism.

If the order of a is infinite, then the kernel of f is the trivial subgroup $\{0\}$, and so f is injective and hence an isomorphism. So $\mathbb{Z} \cong G$ in case $|a|$ is infinite.

If the order of a is finite, say the order is n, then the kernel of f is the set $n\mathbb{Z}$ of all multiples of n, and by the first isomorphism theorem, $\mathbb{Z}/n\mathbb{Z} \cong G$. But $\mathbb{Z}/n\mathbb{Z}$ is $\mathbb{Z}_n$, by definition. So in this case we have $\mathbb{Z}_n \cong G$. □

Exercises

22.1. Prove that kernels are normal subgroups (see 22.6).
22.2. Prove that an isomorphism is a bijective homomorphism.
22.3. (Injectivity test for homomorphisms) Prove that if $f : G \to H$ is a group homomorphism with kernel K then f is injective if and only if K is the trivial subgroup.
22.4. Prove that the canonical homomorphism (see 22.7) is really a homomorphism; i.e., show that $\pi(ab) = \pi(a)\pi(b)$ for all a, b in G.
22.5. Prove that if $f : G \to H$ and $g : H \to K$ are group homomorphisms then their composite $g \circ f$ is a group homomorphism.
22.6. (Homomorphic images of subgroups are subgroups) Show that if $f : G \to G'$ is a group homomorphism and $H < G$ then $f(H) < G'$, where $f(H) = \{f(x) \mid x \in H\}$.

22.7. (Homomorphic preimages of subgroups are subgroups) Show that if $f : G \to G'$ is a group homomorphism and $H' < G'$ then $f^{-1}(H') < G$, where $f^{-1}(H') = \{x \in G \mid f(x) \in H'\}$. Furthermore, show that $\ker f < f^{-1}(H')$.

22.8. Define a map $f\colon \mathbb{R} \to \mathbb{C}^\times$ by $f(x) = e^{2\pi i x}$ for any $x \in \mathbb{R}$. (Here $\mathbb{R}$ is the additive group of real numbers.)

(a) Prove that f is a group homomorphism and compute its kernel.

(b) Use the first isomorphism theorem to show that the quotient group $\mathbb{R}/\mathbb{Z}$ is isomorphic to the circle group $\{e^{2\pi i x} \mid x \in \mathbb{R}\}$.

22.9. Prove that $\mathbb{R}/\mathbb{Z}$ in the previous problem is isomorphic to the group SO(2) of rotations of the plane.

22.10. Let G be an abelian group. Let $H = \{x^2 \mid x \in G\}$ be the set of squares in G, and let $K = \{x \in G \mid x^2 = 1\}$ be the set of elements of order one or two. Prove that:

(a) The function $f : G \to G$ defined by $f(x) = x^2$ is a homomorphism.

(b) Identify the kernel of f and justify your answer.

(c) Show that $G/K \cong H$.

22.11. Prove that $\mathrm{SL}(n) \triangleleft \mathrm{GL}(n)$ and $\mathrm{GL}(n)/\mathrm{SL}(n) \cong \mathbb{R}^\times$ by applying the first isomorphism theorem.

22.12. Prove that $\mathrm{SO}(n) \triangleleft \mathrm{O}(n)$ and $\mathrm{O}(n)/\mathrm{SO}(n) \cong \mathbb{Z}^\times$ by applying the first isomorphism theorem.

22.13. Prove the *second isomorphism theorem*: If $K \triangleleft G$ and $H < G$ then $H \cap K \triangleleft H$, HK is a subgroup of G, and $H/(H \cap K) \cong HK/K$. (See 22.11.)

22.14. Prove the *third isomorphism theorem*: If $H \triangleleft G$, $K \triangleleft G$, and $H \subset K$ then $(K/H) \triangleleft (G/H)$ and $(G/H)/(K/H) \cong G/K$. (See 22.12.)

22.15. If G is a group, show that the function $f : G \to G$ defined by $f(x) = x^2$ is a homomorphism if and only if G is abelian.

22.16. Use the third isomorphism theorem (see 22.12) to prove that if $m = nk$ for positive integers m, n, k then the additive group $\mathbb{Z}_m$ has a subgroup $\mathbb{Z}'_n$ isomorphic to $\mathbb{Z}_n$, and $\mathbb{Z}_m/\mathbb{Z}'_n \cong \mathbb{Z}_k$. [Hint: Start by setting $G = \mathbb{Z}$, $H = m\mathbb{Z}$, and $K = k\mathbb{Z}$.]

22.17. Show directly that any quotient group of a cyclic group is cyclic. Why does this give a different proof of the result in the preceding problem?

Chapter 7

Simple Groups and Direct Products

23 Simple groups

The notion of a simple group was introduced by Galois. Around 1981, a complete classification of the finite simple groups was obtained. This was the culmination of many decades of effort of many researchers. The classification theorem is regarded as one of the most important achievements of twentieth century mathematics; see Appendix C for a brief history of its proof.

Some of the theorems in this section will be stated without proof. They are given here for the sake of general knowledge of the subject, and will not be used in the sequel.

23.1 Definition (Galois)**.** A group G is called *simple* if it has no normal subgroups other than $\{1\}$ and G itself.

This means that G is simple if and only if the only quotient groups of G are isomorphic to $\{1\}$ and G.

Notice that the definition of a simple group is analogous to the definition of a prime number in the integers. A prime integer is one which has no factors other than the trivial ones, and a simple group is one which has no factor groups other than the trivial ones.

23.2 Examples. 1. Every cyclic group of prime order is a simple group. This is an easy consequence of Lagrange's theorem.

2. Every alternating group $\mathbb{A}_n$ is simple, except for $\mathbb{A}_4$. This was discovered by Galois, and the proof is not easy.

The following result, known as the *correspondence theorem*, is another basic theorem about homomorphisms and quotient groups.

23.3 Theorem (correspondence theorem)**.** *Let K be any normal subgroup of a given group G and let $\pi : G \to G/K$ be the canonical homomorphism. Then the mapping $H \mapsto \pi(H) = H/K$ defines a bijective correspondence between the subgroups of G containing K and the subgroups of G/K. Furthermore, in this correspondence $H \triangleleft G$ if and only if $H/K \triangleleft G/K$.*

Proof. This is an exercise in using the isomorphism theorems. □

As an application of the correspondence theorem, we describe a way of "factoring" a given finite group into a list of simple groups, analogous to the way in which we factor a given positive integer into its prime factors.

Suppose that G is a given finite group. Let N_1 be a proper normal subgroup of G which is as large as possible. Then by the correspondence theorem G/N_1 has no nontrivial proper normal subgroup, so G/N_1 is simple. Next, choose a proper normal subgroup N_2 of N_1 that is as large as possible. Then as before, the quotient N_1/N_2 must be simple. Continue in this way as long as possible. Eventually you will arrive at a subgroup N_{k-1} in which the largest proper normal subgroup is the trivial subgroup $\{1\}$, and the process terminates with $N_k = \{1\}$. We know that the process must terminate since G is finite.

A series of normal subgroups such as the one we just described is called a *composition series* of G. It describes a way in which G can be factored as a series of simple group quotients. This explains the interest in simple groups: in some sense every finite group can be described by a series of simple groups.

These observations lead to the following definition.

23.4 Definition. Let G be a group. A *composition series* of G is a sequence of normal subgroups

$$G = N_0 \triangleright N_1 \triangleright N_2 \triangleright \cdots \triangleright N_{k-1} \triangleright N_k = \{1\}$$

where N_j/N_{j+1} is a simple group for each j. The various simple quotient groups N_j/N_{j+1} are called the *composition factors* of G.

The following important theorem says that the set of composition factors of a finite group are uniquely determined by the group, apart from the order in which they are produced. This is a fundamental property of every finite group.

23.5 Theorem (Jordan–Hölder)**.** *Any finite group has a composition series. Moreover, its composition factors are unique, except for order and isomorphism. In other words, if*

$$G = N_0 \triangleright N_1 \triangleright N_2 \triangleright \cdots \triangleright N_{m-1} \triangleright N_m = \{1\}$$

and

$$G = K_0 \triangleright K_1 \triangleright K_2 \triangleright \cdots \triangleright K_{n-1} \triangleright K_n = \{1\}$$

are any two composition series for G, then $m = n$ and there is a permutation $\alpha \in \mathbb{S}_n$ such that $N_i/N_{i+1} \cong K_{\alpha(i)}/K_{\alpha(i+1)}$ for each i.

Proof. The existence of the composition series was proved in the remarks preceding the theorem. The proof of the uniqueness statement is omitted, but can be easily found by consulting the literature. □

Recall that the fundamental theorem of arithmetic says that every positive integer can be written as a product of primes, and the prime factors are unique apart from their order. The Jordan–Hölder theorem is somewhat analogous, it says that every finite group has a composition series in which the composition factors are simple groups, and the composition factors are unique apart from their order. In this analogy, the simple groups are analogous to prime numbers.

The following definition is based on the fundamental work of Galois on solutions of polynomial equations.

23.6 Definition. A group G is called *solvable* if it has a series of normal subgroups

$$G = N_0 \triangleright N_1 \triangleright N_2 \triangleright \cdots \triangleright N_{k-1} \triangleright N_k = \{1\}$$

such that the quotient group N_j/N_{j+1} is cyclic of prime order, for each j.

In other words, the composition factors of a solvable group are all cyclic groups of prime order. Later we will prove that *any finite abelian group is solvable.*

The reason for the terminology 'solvable' is explained by the following result, known as the *fundamental theorem of Galois theory.* This is the famous result that connects group theory to the ancient problem of solving polynomial equations. The proof is beyond the scope of this course, and will not be discussed here.

23.7 Theorem (Galois)**.** *Let $p(x)$ be an irreducible polynomial with rational coefficients, and let $G = \mathrm{Gal}(p)$ be its Galois group. Then the complex roots of $p(x)$ are expressible in terms of radicals if and only if G is a solvable group.*

The following famous result is usually known as the *odd order theorem*. It was first proved in 1963 by Walter Feit and John Thompson.

23.8 Theorem (the Feit–Thompson odd order theorem)**.** *Every finite group of odd order is solvable.*

This was a landmark result in group theory. The published proof of this theorem appeared in *Pacific Journal of Mathematics*, vol. 13, pp. 775–1029. Yes, that is correct: the proof occupies 255 printed pages! Perhaps we can be forgiven for not giving the proof here.

Exercises

23.1. Use Lagrange's theorem to prove that every cyclic group of prime order is a simple group.

23.2. Compute the composition factors of $\mathbb{S}_2$ and $\mathbb{S}_3$. Are they solvable groups?

23.3. Show that $\mathbb{A}_3$ is simple.

23.4. Compute the composition factors of $\mathbb{S}_4$. Is $\mathbb{S}_4$ a solvable group? Justify your answer.

23.5. Assuming that $\mathbb{A}_5$ is a simple group (this was proved by Galois) show that $\mathbb{S}_5$ is not a solvable group.

23.6. Let p be a prime. Show that the dihedral group $\mathbb{D}_p$ of order $2p$ is a solvable group, by computing its composition factors.

23.7. Find a composition series of $\mathbb{D}_4$, and compute its composition factors. Is $\mathbb{D}_4$ a solvable group?

23.8. Galois proved that the alternating group $\mathbb{A}_5$ is simple. Galois also showed that the symmetry group of the general quintic equation (degree 5 polynomial with arbitrary variable coefficients) is $\mathbb{S}_5$. Assuming these facts, prove that the roots of a general quintic cannot be expressed in terms of radicals.

23.9. Prove the correspondence theorem (Theorem 23.3).

23.10. Show that if G is a group of prime power order p^r for a prime p and $r \geq 1$ then the composition factors of G are all isomorphic to $\mathbb{Z}_p$.

B Appendix: Brief history of simple groups

The notion of a simple group was introduced by Galois in 1831. He knew that the cyclic groups of prime order, and the alternating groups $\mathbb{A}_n$ for $n \geq 5$, were examples of simple groups. Jordan found some more simple groups (they were matrix groups) in 1870, and L.E. Dickson in Chicago found some more in the years between 1892 and 1905. Around 1905 Miller and Cole showed that five groups described by E. Mathieu in 1861 were in fact simple groups; these five, which did not seem to fit in with other known examples, came to be known as *sporadic* groups.

No other examples of finite simple groups were discovered until the 1950s. Then a French mathematician, Claude Chevalley, had an important insight which led him to construct infinitely many new examples of finite simple groups, using matrix groups as a tool in the construction. Other examples quickly followed in the 1960s and 1970s, including new sporadic groups. During those decades, research into finite group theory was intense, and the focus was on simple groups. Most of the new finite simple groups discovered during those years are linear groups or variations thereof.

The study of simple groups has led to a truly colossal theorem: the *classification* of all finite simple groups. This was supposedly achieved around 1981.[1]

B.1 Theorem (the classification theorem)**.** *Every finite simple group is isomorphic to one of the following groups:*

(a) A cyclic group of prime order.

(b) An alternating group of degree at least 5.

(c) A simple group of Lie type, including both

- *the classical Lie groups, namely the simple groups related to the projective special linear, unitary, symplectic, or orthogonal transformations over a finite field;*
- *the exceptional and twisted groups of Lie type (including the Tits group).*

(d) One of the 26 sporadic simple groups.

The proof of this theorem took decades of effort on the part of hundreds

[1] Actually, although the theorem was announced in 1981 as finished, there was a gap in the proof, in that a classification of the so-called quasi-thin groups was never completely written down. This gap was only very recently filled in, by the 2004 publication of a two-volume work of 1,221 pages by Aschbacher and Smith. (Stephen Smith is a professor at the University of Illinois at Chicago.)

of mathematicians. According to the preface of a book[2] on the classification:

> The existing proof of the classification of the finite simple groups runs to somewhere between 10,000 and 15,000 journal pages, spread across some 500 separate articles by more than 100 mathematicians, almost all written between 1950 and the early 1980s.

The proof followed a plan that was outlined by Daniel Gorenstein at a group theory conference at the University of Chicago in 1972. At the time, the experts were skeptical. Some experts were on record predicting that a complete classification of the finite simple groups would take hundreds of years to achieve. But a University of Chicago Ph.D. student in the audience, Michael Aschbacher (now a professor at Cal Tech), went to work on Gorenstein's plan, producing a series of breakthroughs that eventually, with a lot of help from others, led to the classification, following rather precisely the plan outlined by Gorenstein. This took only nine years, and the experts were astonished!

B.2 Examples (and remarks)**.** 1. Every cyclic group of prime order is simple. This is an easy consequence of Lagrange's theorem.

2. The alternating groups $\mathbb{A}_n$ for $n \geq 5$ are simple. This fact was first proved by Galois, and the fact that $\mathbb{A}_5$ is simple is the reason why the quintic equation is unsolvable in terms of radicals.

3. In the classification of finite simple groups, most of the simple groups are linear groups (i.e., groups of matrices) over finite fields.

4. There are 26 sporadic simple groups. The largest of these is known as the *Monster*; it was discovered around 1980 by Robert Greiss at the University of Michigan. It has about 8×10^{53} elements — more than the number of atoms in the universe! The Monster is a group of rotations in a euclidean space of 196883 dimensions. In other words, we can represent its elements by $n \times n$ matrices where $n = 196883$.

5. In 1998 at the International Congress of Mathematicians in Berlin, a Fields Medal[3] was awarded to Richard Borcherds for his work on settling the so-called *Monstrous Moonshine Conjecture* of Conway and Norton, which gives a connection between the Monster group and a certain function which

[2] Gorenstein, Daniel; Lyons, Richard; Solomon, Ronald, *The classification of the finite simple groups.* Mathematical Surveys and Monographs, 40.1. American Mathematical Society, Providence, RI, 1994.

[3] The Fields Medal has generally been regarded as the highest honor a mathematician can achieve, since there is no Nobel prize for mathematics.

appears in conformal field theory in physics. To accomplish this, Borcherds invented a subject known as "vertex operator algebras" and in particular defined a new Lie algebra known as the Monster Lie algebra. It turns out that vertex operator algebras have applications to physics.

Currently there are two teams of mathematicians working on a revision of the proof of the classification theorem. Their stated goal is to reduce the length of the proof to between 3,000 and 4,000 pages.

24 Direct products of groups

Recall the Cartesian product of two sets in set theory, which is used to construct the euclidean plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ as the set of all ordered pairs of real numbers. For any sets A, B we can similarly form the Cartesian product set

$$A \times B = \{(x, y) \mid x \in A, y \in B\}.$$

Given two groups one can make their Cartesian product into a group in a very natural way. This makes it possible to construct many new examples of groups by taking products of ones we already know.

As usual, we use multiplicative notation in this section, leaving it to the reader to make the necessary notational adjustments in other cases.

24.1 Definition. The *direct product* of given groups G, H is the group $G \times H = \{(x, y) \mid x \in G, y \in H\}$ of ordered pairs, with binary operation

$$(x, y) \cdot (u, v) = (xu, yv), \text{ for all } x, u \in G,\ y, v \in H.$$

The identity element of $G \times H$ is the pair $(1_G, 1_H)$. The inverse of a pair (x, y) is the pair (x^{-1}, y^{-1}); that is: $(x, y)^{-1} = (x^{-1}, y^{-1})$.

It must be checked that this actually works; that is, we must verify that the above multiplication rule on $G \times H$ makes $G \times H$ a group. This is left for you as an exercise.

If G, H have finite order then we have $|G \times H| = |G| \cdot |H|$; i.e., the order of the direct product $G \times H$ is the product of the orders of G and H. If one of the groups G, H is infinite then so is their direct product.

24.2 Examples. 1. The group $\mathbb{Z}_2 \times \mathbb{Z}_2$ is a group of order 4. It is not hard to see that it is isomorphic with the Klein 4-group.

2. If m, n are relatively prime then $\mathbb{Z}_m \times \mathbb{Z}_n$ is isomorphic to $\mathbb{Z}_{mn}$. In particular, if p, q are distinct primes then $\mathbb{Z}_p \times \mathbb{Z}_q \cong \mathbb{Z}_{pq}$.

Recall that if H, K are subsets of a group G then $HK = \{hk : h \in H, k \in K\}$.

24.3 Theorem (Internal direct products)**.** *Let G be a group and suppose that H, K are normal subgroups of G. If $HK = G$ and $H \cap K = \{1\}$ then $G \cong H \times K$ and every element of G is uniquely expressible as a product of an element of H by an element of K.*

Proof. We define a map f from $H \times K$ to G by the rule $f((x,y)) = xy$, for any $x \in H$, $y \in K$. Since $HK = G$ the map f is surjective.

Next we claim that *elements of H must commute with elements of K.* For any $x \in H$, $y \in K$ consider the product $(xyx^{-1})y^{-1} = x(yx^{-1}y^{-1})$. Since K is normal, the left hand side is an element of K. Since H is normal, the right hand side is an element of H. Thus the product under consideration lies in the intersection $H \cap K$, so $xyx^{-1}y^{-1} = 1$, so $xy = yx$. This proves the claim.

We can now verify that f is actually a homomorphism: $f(\,(x,y)(x',y')\,) = f((xx',yy')) = xx'yy' = xyx'y' = f((x,y))f((x',y'))$. Finally, we check that f is injective. Let (x,y) ($x \in H$, $y \in K$) be an element of the kernel. Then $xy = 1$, so $y = x^{-1}$ must belong to H, thus to $H \cap K$, and thus $y = 1$. Thus $x = 1$ as well, and $(x,y) = (1,1)$. This shows that the kernel of f must be the trivial subgroup of $H \times K$, so f is injective, as desired. We have shown that f is a bijective homomorphism. Thus f is an isomorphism from $H \times K$ to G, so $H \times K \cong G$.

It remains to show that every element $g \in G$ is uniquely expressible in the form hk, for some $h \in H$, $k \in K$. That g has such an expression is clear from the hypothesis $G = HK$, so we only need to prove uniqueness. Suppose that $g = h_1k_1 = h_2k_2$ where $h_1, h_2 \in H$, $k_1, k_2 \in K$. Then by left multiplying by h_2^{-1} and right multiplying by k_1^{-1} we obtain $h_2^{-1}h_1 = k_2k_1^{-1}$. In this equation, the left hand side is an element of H while the right hand side is an element of K. So both sides of the equation belong to $H \cap K$. But $H \cap K = \{1\}$, so $h_2^{-1}h_1 = 1$ and $k_2k_1^{-1} = 1$; i.e., $h_1 = h_2$ and $k_1 = k_2$. This proves the uniqueness statement. □

The point of the previous theorem is to "factor" the group G as a direct product of two of its subgroups. This is a group-theoretic analogue of factoring integers.

24.4 Definition. Whenever a group G is isomorphic to the direct product of normal subgroups H, K then we say it is the *internal direct product* of those subgroups, and we write $G = H \times K$.

According to the theorem, we can factor G as the internal direct product of normal subgroups H, K if and only if $HK = G$ and $H \cap K = \{1\}$.

24.5 Examples. 1. If m, n are relatively prime then $\mathbb{Z}_{mn}$ has a unique subgroup H of order m, and a unique subgroup K of order n. In fact, $H = \langle [n] \rangle$ and $K = \langle [m] \rangle$ as you can easily check. Then $H + K = \mathbb{Z}_n$ and $H \cap K = \{[0]\}$, so $\mathbb{Z}_{mn}$ is the internal direct product of H, K. Of course

$H \cong \mathbb{Z}_m$ and $K \cong \mathbb{Z}_n$ since all subgroups of a cyclic group are cyclic. Note that we write $H + K$ instead of HK here because the groups are additive groups.

2. The group $\mathbb{S}_3$ has subgroups $H = \langle (1,2) \rangle$, $K = \langle (1,2,3) \rangle$ of order 2 and 3, respectively. It is clear that $HK = \mathbb{S}_3$ and $H \cap K = \{(1)\}$. Furthermore, K is a normal subgroup of $\mathbb{S}_3$ since it is a subgroup of index 2. Alas, the subgroup H is *not* normal. So $\mathbb{S}_3$ is not equal to the direct product of these two subgroups. (In fact, it is impossible to express $\mathbb{S}_3$ as the internal direct product of any two of its subgroups.)

The discussion extends to products of more than two groups.

24.6 Definition. The *direct product* of given groups $G_1, G_2, \ldots, G_n$ is the group $G_1 \times G_2 \times \cdots \times G_n = \{(x_1, x_2, \ldots, x_n) \mid x_k \in G_k \text{ for all } k = 1, \ldots, n\}$, with binary operation

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

The identity element is the tuple $(1, \ldots, 1)$. The inverse of $(x_1, \ldots, x_n)$ is the tuple $(x_1^{-1}, \ldots, x_n^{-1})$ of inverses; i.e. $(x_1, \ldots, x_n)^{-1} = (x_1^{-1}, \ldots, x_n^{-1})$.

Again, it must be verified that this really is a group. The verification is no more difficult than for the case of products of two groups, and thus left to the reader.

If $H_1, H_2, \ldots, H_n$ are subsets of a group G then we extend the product notation in the obvious way: $H_1 H_2 \cdots H_n = \{h_1 h_2 \cdots h_n : h_k \in H_k, \text{ for all } k = 1, 2, \ldots, n\}$.

24.7 Theorem (Internal direct products)**.** *Let G be a group and suppose that $H_1, H_2, \ldots, H_n$ are normal subgroups of a group G. If $H_1 H_2 \cdots H_n = G$ and*

$$H_k \cap (H_1 \cdots H_{k-1} H_{k+1} \cdots H_n) = \{1\}, \text{ for all } k = 1, \ldots, n$$

then $G \cong H_1 \times H_2 \times \cdots \times H_n$ and every element of G is uniquely expressible as a product of the form $h_1 h_2 \cdots h_n$, where $h_k \in H_k$ for all k.

The proof is omitted.

24.8 Definition. Whenever normal subgroups $H_1, H_2, \ldots, H_n$ of a group G can be found such that $G \cong H_1 \times H_2 \times \cdots \times H_n$ then we say that G is the *internal direct product* of the subgroups, and write $G = H_1 \times H_2 \times \cdots \times H_n$.

For this to hold, it is necessary that the subgroups are normal subgroups and that they satisfy the conditions of the theorem.

Since internal and external direct products are isomorphic, people usually do not bother to make a distinction between them.

For finite subgroups of a group, the following simple counting proposition can be useful.

24.9 Proposition. *Let H, K be finite subgroups of a group G such that $H \cap K = \{1\}$. Then $|HK| = |H| \cdot |K|$.*

Proof. By definition, $HK = \{xy \mid x \in H, y \in K\}$. The number of elements is $|H| \cdot |K|$ precisely when all the listed products are distinct, so that's what needs to be shown. So suppose that $x_1y_1 = x_2y_2$ where $x_1, x_2 \in H$, $y_1, y_2 \in K$. Left multiply the equation $x_1y_1 = x_2y_2$ by x_1^{-1} and right multiply by y_2^{-1} to get $y_1y_2^{-1} = x_1^{-1}x_2$. Since H, K are subgroups of G, the left hand side $y_1y_2^{-1} \in K$ and the right hand side $x_1^{-1}x_2 \in H$. Since they are equal, both elements are in $H \cap K$. Since $H \cap K = \{1\}$, it follows that $y_1y_2^{-1} = 1$ and $x_1^{-1}x_2 = 1$; i.e., $y_1 = y_2$ and $x_1 = x_2$. □

Exercises

24.1. Verify that if G, H are groups then $G \times H$ is a group.

24.2. If G, H are abelian groups, show that their direct product $G \times H$ is abelian.

24.3. Show that $G \times \{1\} \cong G$.

24.4. Is the dihedral group $\mathbb{D}_n$ ever isomorphic to $\mathbb{Z}_n \times \mathbb{Z}_2$? Prove your answer.

24.5. Show that if m, n are relatively prime then $C_{mn} \cong C_m \times C_n$. (Here, C_n means the cyclic group of order n.)

24.6. Show that if m, n are relatively prime then $\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$.

24.7. Prove that $G \times H \cong H \times G$.

24.8. Prove that if $G \times H \cong G \times K$ then $H \cong K$. (This can be called a *cancellation property* for direct products.)

24.9. The symmetric group $\mathbb{S}_3$ has composition factors isomorphic to the cyclic groups C_2, C_3 of order 2 and 3. Show that $\mathbb{S}_3 \ncong C_2 \times C_3$. (This shows that not all groups are isomorphic to the product of their composition factors.)

24.10. If $|G| = mn$ where m, n are relatively prime and G has normal subgroups H, K of order m, n respectively, then show that $G = H \times K$.

24.11. Show that if H, K are normal subgroups of a group G, $HK = G$, and every $g \in G$ is *uniquely* expressible in the form $g = hk$ for some $h \in H$, $k \in K$ then $H \cap K = \{1\}$ and hence $G = H \times K$.

24.12. If $H_1, H_2, \ldots, H_n$ are normal subgroups of a group G, $H_1H_2\cdots H_n = G$, and every $g \in G$ is *uniquely* expressible as a product of the form $g = h_1h_2\cdots h_n$, where $h_k \in H_k$ for all k, then show that

$$H_k \cap (H_1 \cdots H_{k-1}H_{k+1} \cdots H_n) = \{1\}, \text{ for all } k = 1, \ldots, n$$

and thus $G = H_1 \times H_2 \times \cdots \times H_n$.

24.13. Show that if $G = H_1 \times H_2 \times \cdots \times H_n$ is the internal direct product of normal subgroups $H_1, H_2, \ldots, H_n$ then for any $i \neq j$ we have:

(a) $H_i \cap H_j = \{1\}$.

(b) $ab = ba$ for all $a \in H_i$, $b \in H_j$. [Hint: Argue that $aba^{-1}b^{-1}$ is in both H_i and H_j.]

24.14. (a) Show that the map $x \mapsto (x, 1)$ is an injective homomorphism from G into $G \times H$.

(b) Show that the image of this homomorphism is a normal subgroup of $G \times H$ isomorphic to G.

24.15. Show that the rule $(x, y) \mapsto x$ defines a surjective homomorphism p_1 mapping $G \times H$ onto G. Similarly, the rule $(x, y) \mapsto y$ defines a surjective homomorphism p_2 mapping $G \times H$ onto H. These maps are called *projections*. Describe their kernels and images.

24.16. If $G = H \times K$ is the direct product of two normal subgroups H, K then prove that $G/H \cong K$ and $G/K \cong H$.

24.17. Let G be a group and consider the direct product $G \times G$. Show that the set $S = \{(x, x) \mid x \in G\}$ is a subgroup of $G \times G$. Then show it is isomorphic to G. (It is known as the *diagonal* subgroup of $G \times G$.)

Chapter 8

Group Actions

25 Group actions

Group actions are useful for a variety of purposes. Not only do they give us new information about groups themselves, but there are also important applications to physics, chemistry, geometry and combinatorics.

25.1 Definition. Let G be a group and X a set. We say that G *acts on* the set X (on the left) if there is a mapping $G \times X \to X$, written as $(g, x) \mapsto g \cdot x$, satisfying the properties

$$(gh) \cdot x = g \cdot (h \cdot x); \qquad 1 \cdot x = x$$

for all $g, h \in G$ and all $x \in X$. When G acts on X, one equivalently says that X is a G-set.

As you might guess, if G acts (on the left) on X then people often abbreviate $g \cdot x$ to gx. It is also useful to consider right actions. A *right action* of a group G on a set X is a mapping $X \times G \to X$, written as $(x, g) \mapsto x \cdot g$, satisfying the properties

$$x \cdot (gh) = (x \cdot g) \cdot h); \qquad x \cdot 1 = x$$

for all $g, h \in G$ and all $x \in X$. For the sake of definiteness, all the actions that we will consider below are left actions, but all the results proved for left actions have counterparts for right actions.

25.2 Examples. 1. The dihedral group $\mathbb{D}_n$ acts on the set of vertices of a regular n-gon. It also acts on the set of edges of the regular n-gon.

2. Any matrix group G of $n \times n$ matrices over a field F acts on the vector space $V = F^n$ by matrix multiplication; more precisely, if $A \in G$ is a matrix in the group G and x is a (column) vector in F^n then the matrix product $Ax \in F^n$ gives the action of A on x.

3. In particular, the orthogonal group $\mathrm{O}(2)$ of 2×2 orthogonal matrices acts on the vector space $V = \mathbb{R}^2$ of points in the Euclidean plane. Given a point $x \in \mathbb{R}^2$ and a matrix $A \in \mathrm{O}(2)$, the action of A on x is given by $Ax \in \mathbb{R}^2$.

Usually we think of a given G-set X as a set of "points" so that we can use geometric language. The action of G moves points in X to other points in X. Starting from a given point and observing where it moves (as G acts) determines the orbit of the point.

25.3 Definition (Orbit and stabilizer)**.** If G is a group acting on a set X, the *orbit* of a point $x \in X$ is the set $O_x = \{gx \mid g \in G\}$. Note that $O_x \subset X$. The *stabilizer* of the point x is the subgroup $G_x = \{g \in G \mid gx = x\}$.

25.4 Lemma. *For any G-set X, the stabilizer G_x of a point $x \in X$ is always a subgroup of G.*

The proof is an easy exercise for the reader. Note that we do *not* claim that G_x is a normal subgroup. In fact, it is not always normal.

25.5 Proposition. *Let G be a group acting on a set X. Consider the relation $\sim$ on X defined by $x \sim y$ if and only if there exists some $g \in G$ such that $y = gx$. Then $\sim$ is an equivalence relation on X. The equivalence classes for $\sim$ are precisely the orbits. Thus the action of G partitions X into a disjoint union of its distinct orbits.*

Again, the proof is an easy exercise.

The following result could be called the *fundamental theorem* of group actions.

25.6 Theorem (The orbit-stabilizer theorem)**.** *Let G be a group acting on a set X. Let $x \in X$. Then $|O_x| = [G : G_x]$. In words: the cardinality of the orbit of x equals the index of the stabilizer of x.*

Proof. We need to find a bijection between O_x and $G/G_x =$ the set of left cosets of G_x in G. Given a point $g \cdot x \in O_x$, where $g \in G$, send it to the left coset gG_x. This rule defines a map $f : O_x \to G/G_x$.

The map f is well-defined since if $g \cdot x = h \cdot x$ for $g, h \in G$ then by left multiplication by h^{-1} we get $(h^{-1}g) \cdot x = x$, so $h^{-1}g \in G_x$ and thus $gG_x = hG_x$. The map f is injective since if $gG_x = hG_x$ then $h^{-1}g \in G_x$ so $(h^{-1}g) \cdot x = x$, so $g \cdot x = h \cdot x$. The map f is surjective because any left coset gG_x is the image of the point $g \cdot x$ in the orbit of x.

Thus, f is a bijective function from O_x to the set G/G_x of left cosets of G_x. The index $[G : G_x]$ is by definition the number of left cosets, so the proof is complete. □

Although the theorem holds generally, for any group (finite or infinite) acting on any set (finite or infinite), it is most useful when G is finite, in which case $[G : G_x] = |G|/|G_x|$ by Lagrange's theorem. This gives a nice corollary of the orbit-stabilizer theorem.

25.7 Corollary. *Suppose that G is a finite group acting on a set X. Then for any point $x \in X$ we have $|G| = |G_x| \cdot |O_x|$.*

Proof. Lagrange's theorem says that $[G : G_x] = |G|/|G_x|$. The orbit-stabilizer theorem says that $[G : G_x] = |O_x|$, so $|O_x| = |G|/|G_x|$ and the result follows. □

In case the G-set X is finite we have the following result, which can be used to count the number of points in the set X.

25.8 Corollary. *Let X be a finite G-set. Let $O_{x_1}, \ldots, O_{x_k}$ be the distinct orbits. Then $|X| = \sum_{i=1}^{k} [G : G_{x_i}]$.*

Proof. Since the action of G partitions X into disjoint orbits, we can write X as the union of its distinct orbits:

$$X = \bigcup_{i=1}^{k} O_{x_i} = O_{x_1} \sqcup O_{x_2} \sqcup \cdots \sqcup O_{x_n}.$$

Since the orbits listed are non-overlapping (i.e., pairwise disjoint) we have

$$|X| = \sum_{i=1}^{k} |O_{x_i}| = \sum_{i=1}^{k} [G : G_{x_i}]$$

where the last equality on the line above is justified by Theorem 25.6. This completes the proof. □

25.9 Remark. The set $\{x_1, \ldots, x_k\}$ in the preceding result is called a *complete set of orbit representatives*. To obtain such a set, one chooses exactly one element from each distinct orbit.

Here is one more piece of terminology that is often used in the theory of group actions.

25.10 Definition. Let X be a G-set. We say that the action of G is *transitive*, or that G acts *transitively*, if there is only one orbit. This means that we can get from any point $x \in X$ to any other point $y \in X$ by acting by a suitable group element.

25.11 Example. The action of $\mathbb{D}_n$ on the set of vertices (or the set of edges) of a regular n-gon is transitive.

Exercises

25.1. Let X be a G-set. Prove Lemma 25.4, that the stabilizer G_x of any point $x \in X$ is a subgroup of G.

25.2. Prove Proposition 25.5.

25.3. Let $\mathbb{D}_4$ act on the set of vertices of a square by its natural action.
   (a) Is the action transitive?
   (b) Compute the stabilizer of a vertex.

25.4. Let $\mathbb{S}_4$ act on the set $\{1,2,3,4\}$ by $\alpha \cdot j = \alpha(j)$, for each $j \in \{1,2,3,4\}$.
   (a) Is the action transitive?
   (b) Compute the stabilizer of 4. What group is it isomorphic to?

25.5. Let $\mathbb{S}_n$ act on the set $\underline{\mathbf{n}} = \{1,\ldots,n\}$ by $\alpha \cdot j = \alpha(j)$, for each $j \in \underline{\mathbf{n}}$.
   (a) Is the action transitive?
   (b) Compute the stabilizer of n. What group is it isomorphic to?

25.6. Is the action of $\mathrm{GL}_2(\mathbb{R})$ on $\mathbb{R}^2$ given by $A \cdot X = AX$ (ordinary matrix multiplication) transitive? Justify your answer.

25.7. Let $X = \{1,2,3\}$. Consider the action of the alternating group $\mathbb{A}_3 = \{(1), (1,2,3), (3,2,1)\}$ on X defined by $\alpha \cdot j = \alpha(j)$.
   (a) How many orbits are there?
   (b) Describe the orbits completely.
   (c) Compute the stabilizer of 3.

25.8. Let X be the set of line segments connecting any two vertices of the square, i.e., the edges and the diagonals. The group $\mathbb{D}_4$ acts on the set X in a natural way: if L is a line segment connecting vertices i and j and $g \in \mathbb{D}_4$ then $g \cdot L$ is the line segment connecting vertices $g \cdot i$ and $g \cdot j$.
   (a) Describe the orbits completely. (You may wish to number the vertices of the square.) How many orbits are there?
   (b) Compute the stabilizer of one of the diagonals.
   (c) Compute the stabilizer of one of the edges.

26 Applications of group actions

We now consider a few examples of the theory of group actions. As we will see, group actions are useful to count the size of various sets; thus group actions are a useful tool in combinatorics.[1]

26.1 Example. The rotation group $\mathrm{SO}_2(\mathbb{R})$ acts on the set $\mathbb{R}^2$ by ordinary matrix multiplication: $A \cdot x = Ax$ for any $A \in \mathrm{SO}_2(\mathbb{R})$ and any $x \in \mathbb{R}^2$ (considered as a column vector). The orbit of any point $(a, b) \in \mathbb{R}^2$ is a circle of radius $\sqrt{a^2 + b^2}$.

26.2 Example. Consider the symmetry group $G = \mathbb{D}_n$ of a regular n-gon in the plane. Let v be a given vertex of the n-gon. Clearly $\mathbb{D}_n$ has n different rotations, and the rotation subgroup $\langle r \rangle$ generated by the basic rotation r acts transitively on the set X of all n vertices. Of all the rotations, only the identity fixes v. There is just one reflection that fixes v, namely the reflection across the (unique) line of symmetry of the figure passing through the vertex v. This proves that the stabilizer G_v of the point v has order 2, since G_v consists of just the identity and the indicated reflection. Hence, by Corollary 25.7 we conclude that $|\mathbb{D}_n| = |O_v| \cdot |G_v| = n \cdot 2 = 2n$. Group actions provide a tool to prove rigorously that $|\mathbb{D}_n| = 2n$, something we found difficult earlier because we lacked the appropriate theory and terminology.

26.3 Example. Let G be the group of proper symmetries of a cube. (Recall that proper symmetries are rotations.) We are going to use Corollary 25.7 to count the number of elements of G. Clearly G acts on the set of 8 vertices of the cube. The action is transitive since you can get from any chosen vertex to any other, by an appropriate sequence of rotations. Let us fix a chosen vertex, call it v. Then $|O_v| = 8$. Moreover, the only rotations fixing v are the three rotations whose axis lies along the diagonal line segment connecting v to its opposite vertex. (It helps to hold an actual cube in your hands to see this.) So $|G_v| = 3$. Thus by Corollary 25.7 we conclude that $|G| = 8 \cdot 3 = 24$. So the octahedral group has 24 elements. (Recall that the symmetry group of the cube and the octahedron are the same group, since the cube and octahedron are dual polyhedra.)

By similar methods you should be able to count the number of proper symmetries of a tetrahedron (12) and the number of proper symmetries of a dodecahedron or its dual, an icosahedron (60). Without the aid of the orbit =stabilizer theorem, counting the size of these groups would be rather daunting

[1] Combinatorics, roughly speaking, is the science of counting.

26.4 Example. Consider the group $\mathbb{S}_n$ acting on the set $\underline{\mathbf{n}} = \{1, \ldots, n\}$ in the natural way: $\alpha \cdot j = \alpha(j)$. This action is transitive because one can get from any i to any j by a suitable permutation. So if we fix a chosen number i then $O_i = \{1, \ldots, n\}$. The number of permutations that fix i is $|\mathbb{S}_{n-1}|$, since we are free to permute the other $n-1$ positions arbitrarily. Thus by Corollary 25.7 we conclude that $|\mathbb{S}_n| = n|\mathbb{S}_{n-1}|$. Since $|\mathbb{S}_1| = 1$, we conclude by a simple induction on n that $|\mathbb{S}_n| = n!$ for any positive integer n. This gives a new proof (via the orbit-stabilizer theorem) that $|\mathbb{S}_n| = n!$.

26.5 Definition. (Permutation Representation) Let X be a G-set. Given $g \in G$, consider the map $\varphi_g : X \to X$ given by the rule $x \mapsto g \cdot x$. In other words, $\varphi_g(x) = g \cdot x$. Then the map φ_g is a bijection (exercise), so φ_g is a permutation of the set X. In other words, $\varphi_g \in \mathbb{S}_X$, the group of permutations of X. The rule $\varphi(g) = \varphi_g$ for each $g \in G$ defines a function φ from G to $\mathbb{S}_X$. This map $\varphi\colon G \to \mathbb{S}_X$ is called the *permutation representation* of the given G-set X.

26.6 Lemma. *The permutation representation $\varphi\colon G \to \mathbb{S}_X$ is a group homomorphism.*

Proof. We need to check that $\varphi(gh) = \varphi(g)\varphi(h)$. In other words, we must check that $\varphi_{gh} = \varphi_g \circ \varphi_h$. We can verify this equality of functions by verifying that the functions on the two sides of the equality act the same on every possible input. So consider any $x \in X$. Then by definition, we have

$$\varphi_{gh}(x) = (gh) \cdot x, \qquad (\varphi_g \circ \varphi_h)(x) = \varphi_g(\varphi_h(x)) = \varphi_g(h \cdot x) = g \cdot (h \cdot x).$$

These are the same by the definition of group actions. □

The permutation representation provides a way to model abstract group elements $g \in G$ by permutations φ_g of X, in such a way that the group multiplication is reproduced in the permutations.

Every group action gives rise to a permutation representation in this way. We now apply this observation to prove the following important result.

26.7 Theorem (Cayley's theorem)**.** *Every group is isomorphic to a group of permutations on some set X. In particular, every finite group G is isomorphic to a subgroup of $\mathbb{S}_n$ where $n = |G|$.*

Proof. Consider the action of G on G itself by left multiplication: $g \cdot x = gx$ for any $g, x \in G$. Here we are taking the set X to be G itself. We use the permutation representation of Definition 26.5. By Lemma 26.6, we know

that $\varphi : G \to \mathbb{S}_G$ is a group homomorphism. The kernel of φ is trivial, since $\varphi(g) = id_G$ implies that $\varphi_g = id_G$, which in turn implies that $gx = x$ for all $x \in G$. This forces $g = 1$.

Thus we have an injective homomorphism $\varphi : G \to \mathbb{S}_G$. This homomorphism induces an isomorphism of G onto its image, which is a subgroup of $\mathbb{S}_G$. In other words, we have produced an isomorphism from G to a group of permutations, as was desired.

To get the last statement, in case G is finite, just number the elements of G from 1 to n. Then permutations of G can be regarded as permutations of the set $\mathbf{n} = \{1, \ldots, n\}$; i.e., $\mathbb{S}_G \cong \mathbb{S}_n$ where $n = |G|$. □

In the proof of Cayley's theorem above, we considered the action of G on itself by left multiplication. Another way that a group G can act on itself is by conjugation. Analysis of this action leads to important structural information about the group.

26.8 Definition. Given $g, x \in G$ define $g \cdot x = gxg^{-1}$. This action of G on itself is called *conjugation*. The element gxg^{-1} is called a *conjugate* of x.

The orbits for the conjugation action are called *conjugacy classes*, and G is the disjoint union is its distinct conjugacy classes. If two elements of G lie in the same conjugacy class, they are said to be *conjugate* in G. We write $C(x)$ for the conjugacy class of x; i.e., $C(x) = \{gxg^{-1} : g \in G\}$.

26.9 Definition. Given an element $x \in G$, its stabilizer is the subgroup G_x given by

$$\{g \in G \mid gxg^{-1} = x\} = \{g \in G \mid gx = xg\}.$$

This subgroup is known as the *centralizer* of x, denoted by $Z_G(x)$.

In words, the centralizer of x is the set of all $g \in G$ which commute with the element x. In this context, the orbit-stabilizer theorem (Theorem 25.6) says that

$$|C(x)| = [G : Z_G(x)]$$

for any $x \in G$. (Because, $C(x) = O_x$ and $Z_G(x) = G_x$ in the earlier notation.)

Notice that the center $Z(G)$ of the group G is contained in every centralizer: $Z(G) \subset Z_G(x)$, for any $x \in G$. In fact, $Z(G)$ is equal to the intersection of all the centralizers: $Z(G) = \bigcap_{x \in G} Z_G(x)$.

26.10 Definition. For any subset S of G, the centralizer of S by $Z_G(S) = \bigcap_{x\in S} Z_G(x)$. Then $Z_G(S) = \{g \in G \mid gx = xg, \forall\ x \in S\}$. In words, the centralizer of a subset S is the set of all elements of G commuting with all elements of S.

In this notation, the center $Z(G)$ is equal to the centralizer of G; i.e., $Z(G) = Z_G(G)$. This is clear from the definitions.

Let me point out that if z is an element of the center $Z(G)$ then the conjugacy class of z (the orbit) is just the singleton set $\{z\}$ (since z commutes with all elements, so $gzg^{-1} = z$ for any g) and $Z_G(z) = G$.

These notions lead us to the following important theorem, which gives new information about finite groups.

26.11 Theorem (The class equation)**.** *Let G be a finite group, $Z(G)$ the center of G. Let $x_1, \ldots, x_t$ be a complete set of representatives for the conjugacy classes that are disjoint from $Z(G)$. Then*

$$|G| = |Z(G)| + \sum_{j=1}^{t} [G : Z_G(x_j)].$$

Proof. This is a restatement of Corollary 25.8. Label the elements of $Z(G)$ by $z_1, z_2, \ldots, z_s$. Then $z_1, \ldots, z_s, x_1, \ldots, x_t$ is a complete set of representatives of the conjugacy classes, so by Corollary 25.8 we have

$$|G| = \sum_{i=1}^{s} [G : Z_G(z_i)] + \sum_{j=1}^{t} [G : Z_G(x_j)].$$

But each $[G : Z_G(z_i)]$ in the first sum is equal to 1 since $Z_G(z_i) = G$ by the remarks above, so the first sum is equal to s, the number of elements in $Z(G)$, and the result is proved. □

26.12 Definition. Let p be a prime. A *p-group* is a group in which every element is of prime power order p^r, for some positive integer r.

Note that any group of order p^r for a positive integer r must be a p-group, by Lagrange's theorem. The study of p-groups is quite important for understanding finite groups. It turns out that p-groups are the most difficult class of finite groups to understand, partly because there are so many of them. For instance, it is known that there are 267 different groups of order $64 = 2^6$, up to isomorphism. We will have more to say about p-groups later.

Here are two immediate applications of the class equation, each of which is a result about p-groups.

26.13 Corollary. *Let p be a prime number.*

(a) Any group of order p^r ($r \geq 1$) must have at least p elements in its center.

(b) All groups of order p^2 are abelian.

Proof. (a) By the class equation, $p^r = |Z(G)| + \sum[G : Z_G(x_j)]$. But each $Z_G(x_j)$ is a *proper* subgroup of G since otherwise x_j would be in the center. So, by Lagrange's Theorem, the order of $Z_G(x_j)$ must be a proper divisor of p^r. But that means the order is of the form p^m for some $m < r$ and so each $[G : Z_G(x_j)]$ is divisible by p. It follows that $|Z(G)|$ must be divisible by p. Part (a) is proved.

(b) If we can show that $Z(G) = G$ then we are done, because $Z(G)$ is obviously abelian. So suppose not. Then there exists some $x \in G$ with $z \notin Z(G)$. Now $Z_G(x)$ is a subgroup of G containing $Z(G)$. By part (a), $Z(G)$ must have at least p elements, so the same is true of $Z_G(x)$. By Lagrange's Theorem, it follows that $|Z_G(x)|$ is either p or p^2. But $|Z_G(x)|$ cannot be equal to p^2 or else x would belong to the center, contrary to our assumption on x. So $|Z_G(x)|$ must equal p. But this implies that $Z_G(x) = Z(G)$. But $x \in Z_G(x)$ (x commutes with itself) so $x \in Z(G)$. This is a contradiction. The contradiction forces $Z(G) = G$, so G is abelian. Part (b) is proved. □

Next we consider the conjugation action on subsets of G. Recall that in set theory the power set of a set S is the collection $\mathcal{P}(S)$ of all subsets of S. Any group G acts on its power set $X = \mathcal{P}(G)$ by conjugation, as follows.

26.14 Definition. Given any subset S of G, and any element $g \in G$, the conjugate $gSg^{-1} = \{gsg^{-1} \colon g \in G\}$ of S by g is another subset of G. Thus the rule $g \cdot S = gSg^{-1}$ for any $g \in G$, $S \in \mathcal{P}(G)$ defines an action of G on the set $\mathcal{P}(G)$.

The orbit of a given subset S for this action is the set of all conjugates gSg^{-1} of S; this set is sometimes denoted $C(S)$. The power set $\mathcal{P}(G)$ of G is the disjoint union of these conjugacy classes. The stabilizer of the subset S is

$$N_G(S) = \{g \in G \mid gSg^{-1} = S\} = \{g \in G \mid gS = Sg\},$$

which is called the *normalizer* of S in G. Note that the set equality $gS = Sg$ does *not* mean that $gs = sg$ for each $s \in S$; rather it means that for every

$s \in S$ there is some $s' \in S$ such that $gs = s'g$. In this context, the orbit-stabilizer theorem (Theorem 25.6) says that

$$|C(S)| = [G : N_G(S)]$$

for any subset S of G.

Note that if H is a given subgroup of G then $H \triangleleft G$ if and only if $N_G(H) = G$. This follows readily from the definitions.

Exercises

26.1. Compute the conjugacy classes of $\mathbb{S}_3$. How many elements are in the centralizer $Z_{\mathbb{S}_3}(\alpha)$ if α is a 3-cycle?

26.2. Compute the conjugacy classes of $\mathbb{S}_4$. How many elements are in the centralizer $Z_{\mathbb{S}_4}(\alpha)$ if α is a 4-cycle?

26.3. Let G be a group. Prove that:

(a) The center $Z(G)$ is a subgroup of any centralizer $Z_G(x)$, for any $x \in G$.

(b) $Z(G) = \bigcap_{x \in G} Z_G(x)$.

26.4. Prove that for any subset S of a group G, $Z_G(S)$ must be a subgroup of $N_G(S)$.

26.5. Prove that if $H < G$ then $H \triangleleft G$ if and only if $N_G(H) = G$.

26.6. (a) Prove that the map φ_g of 26.5 is a bijection.

(b) Verify that φ (see 26.5) is a homomorphism.

26.7. Use a permutation representation of the dihedral group $\mathbb{D}_3$ to find a permutation group isomorphic to $\mathbb{D}_3$. Write out a list of the elements of the permutation group, and explain how the isomorphism is defined.

26.8. Use a permutation representation of the cyclic group $\mathbb{Z}_n$ to find a permutation group isomorphic to $\mathbb{Z}_n$. Explain how the isomorphism is defined.

26.9. If G is a group of order p^r where r is a positive integer and p is a prime, then show that every subgroup of G has order p^k for some integer $k \leq r$.

26.10. If G is a group of order p^r where r is a positive integer and p is a prime, then show that $|Z(G)| = p^k$ where $k \geq 1$.

27 Burnside's Lemma

Another important application of group actions is the result commonly known as Burnside's lemma. This result was proved originally by Georg Frobenius in 1887, so it is not due to Burnside. Burnside included it in his popular 1897 book *On the Theory of Groups of Finite Order*. By an accident of history, the result has become known as Burnside's lemma.

27.1 Theorem (Burnside's lemma)**.** *Let G be a finite group acting on a finite set X. Then the number N of orbits is given by*

$$N = \frac{1}{|G|} \sum_{g \in G} |X^g|$$

where $X^g = \{x \in X \mid g \cdot x = x\}$ is the set of fixed points of g.

Proof. First we prove the result in case G acts transitively. Then $N = 1$ and we need to show that $|G| = \sum_{g \in G} |X^g|$. Let

$$T = \{(g, x) \in G \times X \mid g \cdot x = x\}.$$

Fix some $x \in X$. The pair $(g, x) \in T$ if and only if $g \in G_x$, so the number of such pairs is $|G_x|$. Since the action is transitive, G_x is conjugate to G_y for any $y \in X$, so $|G_y| = |G_x|$ for all $y \in X$. So if we sum over y we get

$$|T| = \sum_{y \in X} |G_y| = \sum_{y \in X} |G_x| = |X| \cdot |G_x| = |G|$$

where the final equality is by the orbit-stabilizer theorem. On the other hand, fix some $g \in G$ and count $|T|$ another way. The pair $(g, x) \in T$ for some $x \in X$ if and only if $g \cdot x = x$, so the set of $x \in X$ with this property is just X^g. Summing over all $g \in G$ we get

$$|T| = \sum_{g \in G} |X^g|.$$

By transitivity of equality, the right hand side of each of the last two displayed equations are equal, which proves the theorem in the transitive case.

Now we consider the general case. Observe that G acts transitively on each of its N orbits. Thus the formula we just proved applies to each orbit. Also the total number of fixed points of a group element is the sum of the number of fixed points in each orbit. Hence

$$N|G| = \sum_{g \in G} |X^g|.$$

This proves the result after dividing both sides by $|G|$. $\square$

This simple counting formula has many useful applications. We give just one example here to whet the reader's appetite.

27.2 Example. Suppose that we wish to count the number of ways to color the vertices of a regular pentagon by red and green. There are two ways to color each vertex, and five vertices, so the simplest answer to our question is that there are $32 = 2^5$ colorings.

But we are usually interested in more sophisticated counting questions. Suppose that we are making a necklace with five red and green beads. We do not wish to distinguish between patterns which are the same under a symmetry of the pentagon, because such patterns produce the same necklace.

To count these, let X be the set of all 32 patterns. The symmetry group $\mathbb{D}_5$ of the pentagon acts on X, and we want to know how many orbits there are. The Burnside lemma will tell us the answer, as soon as we have computed the number of fixed points of each symmetry.

Case 1. $X^1 = X$. Obviously each element of X is fixed by $1 \in \mathbb{D}_5$. So $|X^1| = 32$.

Case 2. $|X^r| = 2$. The only patterns fixed by a rotation r are the ones in which all colors are the same, either all red or all green. This holds true for all three non-trivial rotations.

Case 3. $|X^d| = 8$. Recall that a basic reflection d fixes one vertex and interchanges the other four vertices in opposite pairs. The only patterns fixed by the basic reflection d are therefore those that have the same color on opposite vertices. So there are two colorings for the fixed vertex, and two each for the opposite pairs, for a total of $2^3 = 8$ colorings left fixed by d. The same is true of each of the other four reflections.

Now we apply the Burnside formula. The identity element of $\mathbb{D}_5$ produces 32 fixed points, each of the four non-trivial rotations produces 2 fixed points, and finally each of the five reflections gives 8 fixed points, so the number of orbits is

$$N = \frac{1}{10}(32 + 4 \cdot 2 + 5 \cdot 8) = 8.$$

This solves our problem. There are exactly 8 distinct colorings by two colors of a necklace with five beads.

Chemists use Burnside's lemma to count chemical compounds. For example, a benzene molecule can be modeled by six carbon atoms in a regular

hexagon in a plane. One of three radicals can be attached to each carbon atom to form a benzene molecule. Counting the number of possible benzene molecules is thus an exercise in Burnside's lemma.

There is a generalization of Burnside's lemma known as the *Polya enumeration theorem*. Applications of group theory to counting problems abound, and we can only hint at the possibilities here.

Exercises

27.1. Determine the number of necklaces with 6 beads of two possible colors.

27.2. Determine the number of necklaces with 4 beads of three possible colors.

27.3. Determine the number of necklaces with 5 beads of three possible colors.

27.4. A benzene molecule can be modeled by six carbon atoms in a regular hexagon in a plane. One of three radicals can be attached to each carbon atom to form a benzene molecule. Count the number of possible benzene molecules using Burnside's lemma.

27.5. In how many ways can the faces of a cube be colored by three colors up to rotational symmetry? [Hint: The answer should be 57.]

Chapter 9

Further Topics

28 The Sylow Theorems

It is rather remarkable that three of the most important general theorems about finite groups were proved by a high school teacher. He was Ludwig Sylow, in Norway, and he proved his famous theorems in a ten page paper published in 1872. If p^r is the *largest* power of a prime p dividing the order of G, then Sylow showed:

(i) G has at least one subgroup of order p^r;
(ii) any two such subgroups are conjugate;
(iii) G has $1 + kp$ such subgroups, for some non-negative integer k.

All of these statements are proved using the theory of group actions. Let us look at more precise statements of these facts. First we need some additional terminology.

28.1 Definition. Let p be a prime divisor of the order of a finite group G. A *p-subgroup* of G is any subgroup whose order is a power of p. If p^r is the highest power of p dividing the order of G, then any subgroup of order p^r is called a *Sylow p-subgroup* of G.

Note that any Sylow p-subgroup is also a p-subgroup, but not vice versa.

28.2 Example. Suppose $|G| = 250 = 2 \cdot 5^3$. The prime divisors of G are just $p = 2$ and $p = 5$. Then the 2-subgroups of G are just the subgroups of order 2, and they are also Sylow 2-subgroups. The 5-subgroups of G are the subgroups of order 5, $25 = 5^2$, and $125 = 5^3$. The subgroups of order 125 are the Sylow 5-subgroups.

28.3 Theorem (First Sylow Theorem)**.** *Let G be a finite group and p a prime divisor of $|G|$. Let p^r be the highest power of p dividing $|G|$. Then for every integer s satisfying $0 \leq s \leq r$ there exists a subgroup of G of order p^s. In particular, a Sylow p-subgroup of G (of order p^r) must exist.*

Proof. Proceed by induction on the order of G. If G has order 1 there is nothing to prove. So assume $|G| > 1$ and assume inductively that the theorem has been proved for all groups of order less than $|G|$. By the inductive hypothesis: If G has a proper subgroup H such that p^s divides the order of H, then H has a subgroup of order p^s, and hence so does G.

Now we apply the class equation (Theorem 26.11) for G, which states that

$$|G| = |Z(G)| + \sum [G : Z_G(x_j)]$$

where the x_j range over a complete set of representatives of the conjugacy classes of G not contained in the center $Z(G)$. Each $Z_G(x_j)$ is a proper subgroup of G (or else x_j would be in the center) so by the inductive hypothesis, if p^s divides the order of any $Z_G(x_j)$ then G contains a subgroup of order p^s and we are done.

The remaining case is that p^s does not divide the order of any $Z_G(x_j)$. Then p divides each index $[G : Z_G(x_j)]$, so from the class equation we conclude that p must divide $|Z(G)|$. But $Z(G)$ is abelian, so this means (by a standard lemma, proved in Lemma 28.4 below) that the center $Z(G)$ has an element a of order p. The subgroup $P = \langle a \rangle$ generated by a has order p, so we are done if $s = 1$.

So assume that $s > 1$. Observe that P is a normal subgroup of G since P is contained in the center $Z(G)$. Hence the quotient group G/P is defined. The order of G/P is $(p^r m)/p = p^{r-1}m$, for some integer m, so by the inductive hypothesis G/P has a subgroup H of order p^{s-1}. By the correspondence theorem this subgroup H corresponds with a subgroup H' of G containing P, such that $H'/P \simeq H$. Thus $|H'| = p^s$ and we are done. □

In order to complete the proof of the first Sylow theorem, we need the following simple result, which amounts to Cauchy's theorem (see Theorem 28.7) in the abelian case.

28.4 Lemma. *Let G be a finite abelian group and let p be a prime divisor of $|G|$. Then G has an element of order p.*

Proof. By induction on the order of G. If $|G| = 1$ there is nothing to prove. Assume that $|G| > 1$ and that the theorem has been proved already for all

abelian groups of order less than $|G|$. If G has no proper subgroup (other than $\{1\}$) then G must be cyclic and the statement of the theorem is easy to see.

The remaining case is that G has a proper subgroup $H \neq |1|$. If p divides $|H|$ then we are done by the inductive hypothesis. So assume that p does not divide $|H|$. Since G is abelian, $H \triangleleft G$. So G/H is an abelian group of order $|G|/|H|$. Now p must divide $|G/H|$ since p divides $|G|$ but not $|H|$. Moreover, $|G/H| < |G|$. Thus, by the inductive hypothesis, G/H has an element aH of order p. (The elements of G/H are left cosets, by definition.) Let $b = a^k$ where $k = |H|$. Then one can check that the order of b is p, since $(aH)^p = H$ implies that $a^p \in H$, so by a corollary to Lagrange's theorem $(a^p)^k = 1$, and thus $b^p = a^{kp} = 1$. This completes the proof, as no smaller positive power of b can be 1. □

28.5 Theorem (Second Sylow Theorem)**.** *Let G be a finite group and p a prime divisor of $|G|$. If Q is a p-subgroup of G and P is any Sylow p-subgroup of G, then Q must be contained in some conjugate of P. In particular, any two Sylow p-subgroups of G must be conjugate.*

28.6 Theorem (Third Sylow Theorem)**.** *Let G be a finite group and p a prime divisor of $|G|$. Write $|G| = p^r m$ where m is not divisible by p. If n_p is the number of Sylow p-subgroups of G, then:*

(a) $n_p \equiv 1 \pmod{p}$*; and*
(b) n_p *divides* m*.*

Proofs of the second and third Sylow theorems can be found in virtually any text on abstract algebra, so they are not reproduced here.

Our first application of Sylow's theorems generalizes Lemma 28.4 to include the non-abelian case.

28.7 Theorem (Cauchy's Theorem)**.** *Let p be a prime divisor of the order of a finite group G. Then G must have an element of order p.*

Proof. By the first Sylow theorem, G has a subgroup of order p. Any element of that subgroup (except the identity) must have order p, by Lagrange's theorem. □

Notice how easy the proof is! This shows the power of the Sylow theorems. Note carefully, however, that we needed an independent proof of Lemma 28.4 in the abelian case, since it is used in the proof of the first Sylow theorem.

Recall that we have earlier defined a p-group (where p is a prime) to be any group in which the order of every element is a power of p. It follows (see Exercise 28.4) from Lagrange's theorem and Cauchy's theorem that the order of any finite p-group must be of the form p^r for some $r \geq 1$.

Here is another application of Sylow's theorems.

28.8 Proposition (the pq theorem)**.** *Let $|G| = pq$ where $p < q$ and p, q are primes. If p does not divide $q - 1$ then G is cyclic.*

Proof. Let H be a Sylow p-subgroup of G and let K be a Sylow q-subgroup of G. By the third Sylow theorem, the number of Sylow p-subgroups has the form $1 + kp$ and divides q. If this number is q then p divides $q - 1$, a contradiction. Thus the number of Sylow p-subgroups must be 1, so $H \triangleleft G$.

Similarly, one can show there is only one Sylow q-subgroup, so $K \triangleleft G$. Now $H \cap K = \{1\}$ by Lagrange's theorem, so

$$|HK| = |H| \cdot |K| / |H \cap K| = pq$$

which proves that $HK = G$. Hence $G = H \times K$ is the direct product of its two Sylow subgroups. Since direct products of abelian groups are abelian, this shows that G is abelian. In fact, if x is a generator of the cyclic group H and if y is any generator of the cyclic group K then x and y commute and the order of xy is the least common multiple of p and q, which is pq, so xy has order pq. Thus xy generates G, so G is cyclic. □

The last result shows that, up to isomorphism, there is just one group (which must be the cyclic group) of order pq, where $p < q$ are two distinct primes such that $q \not\equiv 1 \pmod{p}$. So up to isomorphism we have just one group of order 15, one of order 33, one of order 35, etc.

28.9 Example. We apply the Sylow theorems to classify groups of order 21. Assume that $|G| = 21 = 3 \cdot 7$. Note that the pq-theorem does not apply since 7 is congruent to 1 mod 3. Let n_7 be the number of Sylow 7-subgroups of G. By the third Sylow theorem,

$$n_7 \equiv 1 \pmod{7} \quad \text{and} \quad n_7 \mid 3.$$

The positive divisors of 3 are $1, 3$ but only 1 is congruent to 1 mod 3, so $n_7 = 1$. So there is just one subgroup H of order 7. By the second Sylow theorem, H is stable under conjugation and hence normal. We have proved that any group of order 21 must have a normal subgroup of order 7. Such

a group is not simple, so at this point we also know that there is no simple group of order 21.

We can be even more precise. Let n_3 be the number of Sylow 3-subgroups. Then the third Sylow theorem implies that $n_3 = 1$ or 7. Let $x \in G$ be a generator of H. Let $y \in G$ be an element of order 3; such an element must exist by Cauchy's theorem. Note that $H = \langle y \rangle$ is a Sylow 3-subgroup. We have $x^7 = 1$, $y^3 = 1$, and (since H is normal) $yxy^{-1} \in H$. Since $H = \langle x \rangle = \{1, x, x^2, \ldots, x^6\}$ is cyclic, it follows that $yxy^{-1} = x^k$ for some positive integer $k < 7$.

What are the possibilities for k? The fact that $y^3 = 1$ provides information on this question, as follows:

$$\begin{aligned} x &= y^3xy^{-3} = y^2(yxy^{-1})y^{-2} = y^2x^ky^{-2} = y(yx^ky^{-1})y^{-1} \\ &= y(yxy^{-1})^ky^{-1} = y(x^k)^ky^{-1} = yx^{k^2}y^{-1}. \end{aligned}$$

Thus $x = yx^{k^2}y^{-1} = (yxy^{-1})^{k^2} = (x^k)^{k^2} = x^{k^3}$. Since x has order 7, it follows that k^3 must be congruent to 1 mod 7, so $k = 1, 2$, or 4. We consider these cases below. Note that $G = \langle x, y \rangle$ is generated by x, y by Proposition 24.9.

Case 1. If $k = 1$ then $yxy^{-1} = x$; i.e., $xy = yx$. Since x, y generate G as noted above, this means that G must be abelian. Furthermore, this implies that K is normal in G, so in fact $G = H \times K$ is the direct product of H and K. This implies that $G \cong Z_7 \times \mathbb{Z}_3 \cong \mathbb{Z}_{21}$ is itself cyclic.

Case 2. If $k = 2$ then $yxy^{-1} = x^2$; i.e., $yx = x^2y$. The relations $x^7 = 1$, $y^3 = 1$, and $yx = x^2y$ actually determine G uniquely. The existence of this group can be checked with a little more work. It is clearly not abelian.

Case 3. If $k = 2$ then $yxy^{-1} = x^4$. In this case $y^2xy^{-2} = yx^4y^{-1} = x^{16} = x^2$. Thus if we replace y by y^2, which also generates K, then we are back in the previous case.

This analysis proves that, up to isomorphism, there are just two groups of order 21, just one of which is abelian (the cyclic group of order 21).

We finish this section by displaying a brief table of the number of groups, up to isomorphism, of order up to 24. This can be justified by the Sylow theorems in conjunction with other results we have proved, but the task is not easy. You can find much more extensive tables online.

n	number	n	number	n	number
1	1	9	2	17	1
2	1	10	2	18	5
3	1	11	1	19	1
4	2	12	5	20	5
5	1	13	1	21	2
6	2	14	2	22	2
7	1	15	1	23	1
8	5	16	14	24	15

Table 9.1: The number of groups of order 1–24, up to isomorphism

Exercises

28.1. Show that the element b in the proof of 28.4 has order p.

28.2. Let p be a given prime number. Prove that if G is a finite group with just one Sylow p-subgroup, then that subgroup must be normal.

28.3. Show that there is only one group of order 33, up to isomorphism. (In other words, every group of order 33 is isomorphic to $\mathbb{Z}_{33}$.)

28.4. Prove that if G is a finite group in which every element has order some power of p (where p is prime) then $|G| = p^r$ for some $r \geq 1$. (Such groups are called p-groups.)

28.5. Prove that no group of order pq, where p, q are primes, is simple. (Consider the possibility $p = q$ as well as $p \neq q$.)

28.6. Use the Sylow theorems to prove that there is no simple group of order 30.

28.7. Use the Sylow theorems to prove that there is no simple group of order 56.

28.8. Show that any quotient of a solvable group must be solvable.

28.9. Prove that if $G/Z(G)$ is cyclic then G must be abelian.

28.10. Show that if $|G| = 2p$ where p is an odd prime then G must be isomorphic to either $\mathbb{Z}_{2p}$ or $\mathbb{D}_p$.

28.11. Show that if $|G| = 12$ then either $G \cong \mathbb{A}_4$ or G has a normal subgroup of order 3. [Hint: If G does not have a normal subgroup of order 3, consider the action of G on the set G/H by left multiplication, where H is one of the Sylow 3-subgroups. This action induces a homomorphism from G into $\mathbb{S}_4$.]

29 Simplicity of $\mathbb{A}_n$

Our final application of the theory of group actions will be to prove the theorem (due to Galois) that the alternating groups $\mathbb{A}_n$ are all simple, except for $\mathbb{A}_4$. This is the key result that shows that the general polynomial of degree 5 or higher is not solvable in terms of radicals.

We begin with an analysis of the conjugacy classes of the symmetric group $\mathbb{S}_n$. We need to examine $\mathbb{S}_n$ because Galois proved that $\mathbb{S}_n$ is the symmetry group associated to a general polynomial of degree n.

29.1 Example. Let's start by looking at the example of $\mathbb{S}_5$. Here is a list of all 120 elements of $\mathbb{S}_5$, produced by the GAP[1] software package:

```
gap> G := SymmetricGroup(5);
Sym( [ 1 .. 5 ] )
gap> List(G);
[ (), (1,5), (1,2,5), (1,3,5), (1,4,5), (2,5), (1,5,2), (1,2), (1,3,5,2),
  (1,4,5,2), (2,3,5), (1,5,2,3), (1,2,3), (1,3)(2,5), (1,4,5,2,3), (2,4,5),
  (1,5,2,4), (1,2,4), (1,3,5,2,4), (1,4)(2,5), (3,5), (1,5,3), (1,2,5,3),
  (1,3), (1,4,5,3), (2,5,3), (1,5,3,2), (1,2)(3,5), (1,3,2), (1,4,5,3,2),
  (2,3), (1,5)(2,3), (1,2,3,5), (1,3,2,5), (1,4,5)(2,3), (2,4,5,3),
  (1,5,3,2,4), (1,2,4)(3,5), (1,3,2,4), (1,4)(2,5,3), (3,4,5), (1,5,3,4),
  (1,2,5,3,4), (1,3,4), (1,4)(3,5), (2,5,3,4), (1,5,3,4,2), (1,2)(3,4,5),
  (1,3,4,2), (1,4,2)(3,5), (2,3,4), (1,5)(2,3,4), (1,2,3,4,5), (1,3,4,2,5),
  (1,4,2,3,5), (2,4)(3,5), (1,5,3)(2,4), (1,2,4,5,3), (1,3)(2,4),
  (1,4,2,5,3), (4,5), (1,5,4), (1,2,5,4), (1,3,5,4), (1,4), (2,5,4),
  (1,5,4,2), (1,2)(4,5), (1,3,5,4,2), (1,4,2), (2,3,5,4), (1,5,4,2,3),
  (1,2,3)(4,5), (1,3)(2,5,4), (1,4,2,3), (2,4), (1,5)(2,4), (1,2,4,5),
  (1,3,5)(2,4), (1,4,2,5), (3,5,4), (1,5,4,3), (1,2,5,4,3), (1,3)(4,5),
  (1,4,3), (2,5,4,3), (1,5,4,3,2), (1,2)(3,5,4), (1,3,2)(4,5), (1,4,3,2),
  (2,3)(4,5), (1,5,4)(2,3), (1,2,3,5,4), (1,3,2,5,4), (1,4)(2,3), (2,4,3),
  (1,5)(2,4,3), (1,2,4,3,5), (1,3,2,4,5), (1,4,3,2,5), (3,4), (1,5)(3,4),
  (1,2,5)(3,4), (1,3,4,5), (1,4,3,5), (2,5)(3,4), (1,5,2)(3,4), (1,2)(3,4),
  (1,3,4,5,2), (1,4,3,5,2), (2,3,4,5), (1,5,2,3,4), (1,2,3,4), (1,3,4)(2,5),
  (1,4)(2,3,5), (2,4,3,5), (1,5,2,4,3), (1,2,4,3), (1,3)(2,4,5), (1,4,3)(2,5)
 ]
```

As you can see, the list is utter chaos. How can we impose any order on this chaos, and make sense of the list? Let's ask GAP to compute the elements of each conjugacy class, and see what happens:

```
gap> CC := ConjugacyClasses(G);
[ ()^G, (1,2)^G, (1,2)(3,4)^G, (1,2,3)^G, (1,2,3)(4,5)^G, (1,2,3,4)^G,
  (1,2,3,4,5)^G ]
gap> List( CC[1] );
[ () ]
gap> List( CC[2] );
[ (1,2), (1,3), (1,4), (1,5), (2,3), (2,4), (2,5), (3,4), (3,5), (4,5) ]
gap> List( CC[3] );
[ (1,2)(3,4), (1,3)(2,4), (1,4)(2,3), (2,5)(3,4), (2,4)(3,5), (2,3)(4,5),
  (1,5)(3,4), (1,4)(3,5), (1,3)(4,5), (1,4)(2,5), (1,5)(2,4), (1,2)(4,5),
  (1,3)(2,5), (1,2)(3,5), (1,5)(2,3) ]
gap> List( CC[4] );
[ (1,2,3), (1,2,4), (1,2,5), (1,3,2), (1,3,4), (1,3,5), (1,4,2), (1,4,3),
  (1,4,5), (1,5,2), (1,5,3), (1,5,4), (2,3,4), (2,3,5), (2,4,3), (2,4,5),
  (2,5,3), (2,5,4), (3,4,5), (3,5,4) ]
gap> List( CC[5] );
[ (1,2,3)(4,5), (1,2,4)(3,5), (1,2,5)(3,4), (1,3,2)(4,5), (1,3,4)(2,5),
  (1,3,5)(2,4), (1,4,2)(3,5), (1,4,3)(2,5), (1,4,5)(2,3), (1,5,2)(3,4),
  (1,5,3)(2,4), (1,5,4)(2,3), (1,5)(2,3,4), (1,4)(2,3,5), (1,5)(2,4,3),
  (1,3)(2,4,5), (1,4)(2,5,3), (1,3)(2,5,4), (1,2)(3,4,5), (1,2)(3,5,4) ]
```

[1] The GAP Group (`http://www.gap-system.org`).

```
gap> List( CC[6] );
[ (1,2,3,4), (1,2,4,3), (1,3,2,4), (1,3,4,2), (1,4,2,3), (1,4,3,2),
  (2,3,4,5), (2,4,3,5), (2,4,5,3), (2,5,3,4), (2,3,5,4), (2,5,4,3),
  (1,3,4,5), (1,4,3,5), (1,4,5,3), (1,5,3,4), (1,3,5,4), (1,5,4,3),
  (1,4,5,2), (1,5,2,4), (1,2,4,5), (1,4,2,5), (1,5,4,2), (1,2,5,4),
  (1,5,2,3), (1,3,5,2), (1,5,3,2), (1,2,5,3), (1,2,3,5), (1,3,2,5) ]
gap> List( CC[7] );
[ (1,2,3,4,5), (1,2,3,5,4), (1,2,4,3,5), (1,2,4,5,3), (1,2,5,3,4),
  (1,2,5,4,3), (1,3,2,4,5), (1,3,2,5,4), (1,3,4,2,5), (1,3,4,5,2),
  (1,3,5,2,4), (1,3,5,4,2), (1,4,2,3,5), (1,4,2,5,3), (1,4,3,2,5),
  (1,4,3,5,2), (1,4,5,2,3), (1,4,5,3,2), (1,5,2,3,4), (1,5,2,4,3),
  (1,5,3,2,4), (1,5,3,4,2), (1,5,4,2,3), (1,5,4,3,2) ]
```

Ah, that's much better. Some order appears in the chaos. It seems that conjugacy classes might be a nice way to organize the elements in a large group. Let's analyze the results of the above computer calculation. What do you notice about each conjugacy class? A quick look reveals that each class consists of all elements that have the same *cycle type*. Each element of the second class is a 2-cycle, each element of the third is a product of two 2-cycles, and each element of the sixth class is a 4-cycle, and so forth.

There is obviously a nice theorem here. Let's formulate and prove it. First, we need the notion of a partition of n, which is an important concept in combinatorics.

29.2 Definition. Let n be a given positive integer. A *partition* of n is by definition any set $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ of positive integers that add up to n.

By standard convention, we will write partitions as ordered k-tuples of the form $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ where the numbers are ordered from biggest to smallest. With this convention, the partitions of n for the first few values of n are displayed in the table below:

n	partitions of n
1	(1)
2	(2), (1^2)
3	(3), $(2,1)$, (1^3)
4	(4), $(3,1)$, (2^2), $(2,1^2)$, (1^4)
5	(5), $(4,1)$, $(3,2)$, $(3,1^2)$, $(2^2,1)$, $(2,1^3)$, (1^5)
6	(6), $(5,1)$, $(4,2)$, $(4,1^2)$, (3^2), $(3,2,1)$, $(3,1^3)$, (2^3), $(2^2,1^2)$, $(2,1^4)$, (1^6)

In the table, we used the notational trick of writing a^k in place of k repeated values of a. Thus, for instance, $(1^5) = (1,1,1,1,1)$. This useful shorthand is commonly used by people dealing with partitions.

Let us display another table, indicating not the partitions themselves but just counting their number. In fact, the number of partitions of a given n is usually denoted as $\mathfrak{p}(n)$, and the function $\mathfrak{p}$ is called the *partition function*.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\mathfrak{p}()n)$	1	2	3	5	7	11	15	22	30	42	56	77	101	135	176

An interesting (and infamous) open problem in mathematics is to find a finite closed formula for the number $\mathfrak{p}()n)$ of partitions of n. Nobody knows how to do this, and it would be astonishing if someone solved it at this point.

Now that we know about partitions, we can return to the problem of describing the conjugacy classes in the symmetric group $\mathbb{S}_n$.

Let α be a permutation in $\mathbb{S}_n$. We know how to write α as a product of disjoint cycles. For convenience, let's do this in such a way that every number from 1 to n actually appears, by inserting 1-cycles for all the fixed points. (Recall that fixed points are usually omitted in the notation.) For example, if $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 7 & 2 & 6 & 5 & 4 & 3 \end{smallmatrix}\right)$ then we have $\alpha = (1)(2,7,3)(4,6)(5) = (2,7,3)(4,6)(1)(5)$. Recall that disjoint cycles commute, so we can reorder the factors any way we like; we have here chosen to write the product of cycles in decreasing order by cycle length. Let us adopt this ordering convention: in this section we will always write the product of cycles in order of decreasing cycle length, from longest to shortest. With this convention, the cycle type of a permutation $\alpha \in \mathbb{S}_n$ is a partition of n.

29.3 Definition. The *cycle type* of a permutation α is the partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ of lengths of each cycle in the product.

For example, if $\alpha = (2,7,3)(4,6)(1)(5)$ as above then the cycle type of α is the partition $(3,2,1,1) = (3,2,1^2)$. For another example, if $\alpha = (7,3,5)(2,1,4)(8,6,10)(9) \in \mathbb{S}_{10}$ then the cycle type of α is the partition $(3,3,3,1) = (3^3,1)$.

What is the cycle type of the identity permutation in $\mathbb{S}_n$? It has exactly n fixed points, so in our convention it must be written as a product of n 1-cycles. Thus the cycle type of the identity permutation in $\mathbb{S}_n$ is the partition $(1^n) = (1,1,\ldots,1)$.

The cycle type of any permutation in $\mathbb{S}_n$ is a partition of n. For every partition of n, there exists a permutation $\alpha \in \mathbb{S}_n$ with that cycle type.

29.4 Example. The various cycle types for permutations in $\mathbb{S}_5$ are just the partitions of 5: (5), $(4,1)$, $(3,2)$, $(3,1^2)$, $(2^2,1)$, $(2,1^3)$, and (1^5). Notice that there are seven partitions of 5, and in Example 29.1 we found exactly seven conjugacy classes in $\mathbb{S}_5$. This observation suggests the following general result.

29.5 Theorem. *Two permutations in $\mathbb{S}_n$ are conjugate if and only if they have the same cycle type. Hence, the partitions of n label the conjugacy*

classes of $\mathbb{S}_n$. The number of distinct conjugacy classes is the same as the number $\mathfrak{p}()n)$ of partitions of n.

Proof. We only need to check the first claim, since all the other claims follow from it immediately.

($\Rightarrow$): Assume that $\alpha, \beta \in \mathbb{S}_n$ are conjugate permutations. This means that there exists some permutation $\gamma \in \mathbb{S}_n$ such that $\beta = \gamma\alpha\gamma^{-1}$. Now I claim that if

$$\alpha = (i_1, \ldots, i_r)(j_1, \ldots, j_s)(\ldots)$$

is the disjoint cycle decomposition of α then the disjoint cycle decomposition of $\beta = \gamma\alpha\gamma^{-1}$ is

$$\beta = (i_1\gamma, \ldots, i_r\gamma)(j_1\gamma, \ldots, j_s\gamma)(\ldots)$$

and this proves that α and β have the same cycle type, as desired.

($\Leftarrow$): Conversely, if α, β have the same cycle type, say

$$\alpha = (i_1, \ldots, i_r)(j_1, \ldots, j_s)(\ldots)$$
$$\beta = (k_1, \ldots, k_r)(m_1, \ldots, m_s)(\ldots)$$

then define γ to be the permutation sending $i_1 \to k_1$, $\ldots$, $i_r \to k_r$, $j_1 \to m_1$, $\ldots$, $j_s \to m_s$, and so on. A calculation shows that $\beta = \gamma\alpha\gamma^{-1}$, so α, β are conjugate. This completes the proof. □

29.6 Example. We already calculated the conjugacy classes of $\mathbb{S}_5$ in Example 29.1. Counting the number of elements in each class gives us the following tabulation for the class equation of $\mathbb{S}_5$:

$$120 = 1 + 10 + 15 + 20 + 20 + 30 + 24.$$

(This is just the fact that G is the union of its conjugacy classes; see 25.8.) Note that the numbers on the right of the class equation all divide $|G|$. This is no accident, because of the fundamental orbit-stabilizer theorem (see 25.6 and 26.8), which says that the size of an orbit $C(x)$ is the same as the index $[G : Z_G(x)]$, which divides the group order $|G|$ by Lagrange's theorem.

29.7 Example. Now let's fire up GAP to compute the conjugacy class orders in the alternating group $\mathbb{A}_5$:

```
gap> G := AlternatingGroup(5);
Alt( [ 1 .. 5 ] )
gap> CC := ConjugacyClasses(G);
[ ()^G, (1,2)(3,4)^G, (1,2,3)^G, (1,2,3,4,5)^G, (1,2,3,5,4)^G ]
gap> Size( CC[1] ); Size( CC[2] ); Size( CC[3] ); Size( CC[4] ); Size( CC[5] );
1
15
20
12
12
```

As we see, there are just five classes in $\mathbb{A}_5$, and their cycle types are given by the partitions (1^5), $(2^2, 1)$, $(3, 1^2)$, (5), and (5). This is an interesting observation, because it shows that there are two classes in $\mathbb{A}_5$ of the same cycle type. (Note that Theorem 29.5, which is a theorem about symmetric groups, fails for alternating groups.) The GAP calculation above gives us the following for the class equation of $\mathbb{A}_5$:

$$60 = 1 + 15 + 20 + 12 + 12.$$

This calculation is going to be very useful in proving that $\mathbb{A}_5$ is a simple group, the main goal of this section.

Before leaving this example, notice that there is only one 1 in each of the class equations considered here. This proves that the center of $\mathbb{S}_5$ and the center of $\mathbb{A}_5$ are trivial, since $x \in Z(G) \Leftrightarrow Z_G(x) = G \Leftrightarrow |C(x)| = 1$.

We have the following general observation.

29.8 Lemma. *Let N be a normal subgroup of a group G. If $x \in N$ then $C(x) \subset N$. Thus N is a union of certain conjugacy classes of G.*

Proof. This follows immediately from the definitions. If N is normal then N is closed under conjugation, so of course N contains all conjugates of any of its elements. □

We are finally ready to prove the following result, originally proved by Galois. This is the culmination of all the information developed in this section.

29.9 Theorem. *[Galois] The alternating group $\mathbb{A}_5$ is a simple group.*

Proof. Suppose that $\mathbb{A}_5$ has a proper normal subgroup N. So the order $|N|$ satisfies $1 < |N| < 60$. Also, $|N|$ must be a divisor of 60, by Lagrange's theorem, so $|N| = 2$, 3, 4, 5, 6, 10, 12, 15, 20, or 30. By the preceding lemma, N is a union of conjugacy classes, so $|N|$ is a sum of the numbers 1, 15, 20, 12, 12 because of the class equation for $\mathbb{A}_5$ that we computed in Example 29.7. But of course we *must* include the number 1 in the sum, because H must contain the identity element and its conjugacy class is a singleton. This is a contradiction: no such sum equals any divisor of 60. This contradiction proves that $\mathbb{A}_5$ has no proper normal subgroup, hence is simple. □

It is easy to see that $\mathbb{A}_2$ and $\mathbb{A}_3$ are simple, since $\mathbb{A}_2$ is trivial and $\mathbb{A}_3$ is of order 3, thus cyclic of prime order. (All cyclic groups of prime order are simple.) See Exercise 29.2 for an outline of a proof that $\mathbb{A}_n$ is simple for all $n \geq 6$. Thus all the alternating groups are simple, except for $\mathbb{A}_4$.

Exercises

29.1. Find a proper normal subgroup of $\mathbb{A}_4$, thus proving that $\mathbb{A}_4$ is not a simple group.

29.2. Let N be any proper normal subgroup of $\mathbb{A}_n$ ($n \geq 6$).

(a) Show that N contains an element $\alpha \neq 1$ which has a fixed point $i \in \{1, \ldots, n\}$.

(b) Show that N contains the subset G_i of *all* permutations in $\mathbb{A}_n$ which fix i.

(c) Show that G_i is isomorphic with $\mathbb{A}_{n-1}$. (So N contains an isomorphic copy of $\mathbb{A}_{n-1}$.)

(d) Show that N contains every G_i, for each $1 \leq i \leq n$.

(e) Show that $\mathbb{A}_n$ must be simple, by showing that no N, with these properties, exists.

30 Classification of finite abelian groups

We have now developed enough general theory to classify all the finite abelian groups, up to isomorphism. This means that we can list all possibilities for a given order n, and prove that any abelian group of order n must be isomorphic to one of the groups on the list.

Let p be a prime. By definition, a *p-group* is a group in which every element has order a power of p. Recall that Cauchy's theorem says that a finite group G has an element of order p, for every prime p dividing the order of G. It follows immediately from Cauchy's theorem and Lagrange's theorem that any finite p-group must be of order p^k for some $k \geq 1$.

30.1 Lemma. *Any finite abelian group is isomorphic to the direct product of its Sylow subgroups.*

Proof. Let G be an abelian group of order n. Since any subgroup of an abelian group is normal, all the Sylow subgroups of G are normal. By the fundamental theorem of arithmetic, n can be factored in the form $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ where $p_1, p_2, \ldots, p_r$ are the distinct prime factors of n. Let G_j be the Sylow p_j-subgroup of G of order $p_j^{k_j}$. Each Sylow subgroup is normal since every subgroup of an abelian group is normal. By Lagrange's theorem, the order of any element of G_j is a power of p_j and the order of any element of the product of the other Sylow subgroups is coprime to p_j. Thus the intersection of any G_j with the product of the other Sylow subgroups is trivial, and it follows that $G = G_1 \times G_2 \times \cdots \times G_r$. □

30.2 Remark. It is not hard to show that the subgroup G_j appearing in the above proof may also be described as $G_j = \{x \in G \mid x^{q_j} = 1\}$ where $q_j = p_j^{k_j}$. This satisfying concrete description is often useful for computations.

The lemma reduces our task (understanding all finite abelian groups) to the task of understanding finite abelian p-groups. For this it will be useful to use the language of partitions.

30.3 Definition. A *partition* of a positive integer k is a sequence $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ of positive integers such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ and $\lambda_1 + \cdots + \lambda_m = k$. The number m is called the *length* of the partition λ.

There is just one partition of 1, namely (1). The partitions of 2 are (2) and $1, 1$). The partitions of 3 are (3), $(2, 1)$, and $(1, 1, 1)$. The partitions of 4 are (4), $(3, 1)$, $(2, 2)$, $(2, 1, 1)$, and $(1, 1, 1, 1)$. When writing partitions,

people often use an *exponential shorthand notation* for repeated entries, in which λ_j^t stands for λ_j repeated t times. For example, in the shorthand notation, $(4, 2^2, 1^3) = (4, 2, 2, 1, 1, 1)$.

It is easy to construct abelian groups of order p^k, for a given fixed prime p. Pick any partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ of the exponent k, and define

$$G(p, \lambda) = \mathbb{Z}_{p^{\lambda_1}} \times \mathbb{Z}_{p^{\lambda_2}} \times \cdots \times \mathbb{Z}_{p^{\lambda_m}}.$$

Since $G(p, \lambda)$ is a direct product of cyclic groups, it is a direct product of abelian groups, and hence is abelian. Since $\lambda_1 + \cdots + \lambda_m = k$, the order of $G(p, \lambda)$ is p^k, so $G(p, \lambda)$ is an abelian group of order p^k.

30.4 Example. The partitions of 5 are (5), $(4, 1)$, $(3, 2)$, $(3, 1^2)$, $(2^2, 1)$, $(2, 1^3)$, and (1^5). For any prime p, we have the following abelian groups of order p^5.

λ	$G(p, \lambda)$
(5)	$\mathbb{Z}_{p^5}$
$(4, 1)$	$\mathbb{Z}_{p^4} \times \mathbb{Z}_p$
$(3, 2)$	$\mathbb{Z}_{p^3} \times \mathbb{Z}_{p^2}$
$(3, 1^2)$	$\mathbb{Z}_{p^3} \times \mathbb{Z}_p \times \mathbb{Z}_p$
$(2^2, 1)$	$\mathbb{Z}_{p^2} \times \mathbb{Z}_{p^2} \times \mathbb{Z}_p$
$(2, 1^3)$	$\mathbb{Z}_{p^2} \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p$
(1^5)	$\mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p$.

In this example, it is not difficult to verify that all the groups in the list are pairwise non-isomorphic, because you can always find an element in one product group of a different order than all elements of the other. (For example, $\mathbb{Z}_{p^5}$ has an element of order p^5, but none of the others do, so $\mathbb{Z}_{p^5}$ is not isomorphic to any of the others.)

Based on the example, one naturally expects that if $\lambda \neq \mu$ are distinct partitions of k then $G(p, \lambda) \ncong G(p, \mu)$. This is indeed true, but somewhat awkward to prove at the moment, so we defer the proof until later in the analysis.

It would be nice if the abelian p-groups we just constructed, namely the ones of the form $G(p, \lambda)$, where λ is a partition, give *all* the finite abelian p-groups up to isomorphism. We will prove that this is indeed the case, following a note by G. Navarro[2] published in 2003.

The key fact we need to prove is that any abelian p-group is isomorphic to a product of cyclic groups. The proof rests on the following pair of lemmas.

[2]Navarro, Gabriel: *On the fundamental theorem of finite abelian groups.* Amer. Math. Monthly 110 (2003), no. 2, 153–154.

30.5 Lemma. *Suppose that G is a finite abelian p-group, where p is a prime. If G has a unique subgroup of order p then G is cyclic.*

Proof. By induction on $|G|$. Consider the homomorphism $f : G \to G$ given by $f(x) = x^p$. This is a homomorphism because G is abelian. Let K be the kernel of f. By hypothesis, K is the only subgroup of G of order p. Then $G/K \cong f(G)$. If $K = G$ then G is cyclic and we are done. Otherwise, K is a proper subgroup of G and hence $f(G)$ is not the trivial subgroup. Every subgroup of $f(G)$ is a subgroup of G, so $f(G)$ has a unique subgroup of order p, and thus is cyclic by the inductive hypothesis. So G/K is cyclic. Then there is some $y \in G$ such that yK generates G/K. Clearly $y \neq 1$ as yK has order $|G|/p$ in G/K. Let $H = \langle y \rangle$ be the cyclic subgroup of G generated by y. Then $HK = G$. (Otherwise the order of HK/K would be strictly less than $|Gp$, in violation of the fact that yK has order $|G|/p$.) Now by Cauchy's theorem H has a subgroup of order p, which must be K. Hence $K \subset H$ and $G = HK = H = \langle y \rangle$, so G is cyclic. □

30.6 Lemma. *If G is a finite abelian p-group, let C be a cyclic subgroup of maximal order. Then $G = C \times B$ for some subgroup B.*

Proof. Again we use induction on $|G|$. If G is cyclic then $G = C \times \{1\}$ and we are done. Otherwise, by the previous lemma G has at least two subgroups of order p, but C has only one. Let K be a subgroup of order p which is not contained in C. Then $C \cap K = \{1\}$. No homomorphic image of G has a cyclic subgroup of order larger than $|C|$, so $CK/K \cong C$ is cyclic of maximal order in G/K. By the inductive hypothesis, $G/K = CK/K \times B/K$ for some subgroup B of G. Since $K \subset B$, it follows that $G = (CK)B = CB$. Furthermore, $C \subset CK$ and $B \cap CK = K$, so $C \cap B = C \cap B \cap CK = C \cap K = \{1\}$. Hence $G = C \times B$. □

30.7 Theorem (classification of abelian p-groups)**.** *Suppose that G is a finite abelian p-group, of order p^k for some $k \geq 1$. Then there is some partition λ of k such that $G \cong G(p, \lambda)$. Furthermore, $G(p, \lambda) \ncong G(p, \mu)$ for $\lambda \neq \mu$.*

Proof. The proof is by induction on $|G|$. If G is cyclic, then $G \cong \mathbb{Z}_{p^k}$ and we are finished. Otherwise, let C be a cyclic subgroup of G of maximum possible order. Then $|C| = p^{\lambda_1}$ where $\lambda_1 < k$. By Lemma 30.6, $G = C \times B$ with $|B| < |G|$. Note that B is also a p-group; in fact $|B| = p^{k-\lambda_1}$. By the inductive hypothesis, B is isomorphic to a product of cyclic groups of the form $G(p, (\lambda_2, \ldots, \lambda_m))$, where $(\lambda_2, \ldots, \lambda_m)$ is a partition of $k - \lambda_1$.

Thus $G = C \times B \cong G(p, \lambda)$, where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$. This proves the first claim. It also proves the second claim, since if $C \times B = C \times B'$ then $B \cong B'$. □

As an easy corollary, we obtain the desired classification theorem for finite abelian groups.

30.8 Theorem (classification of finite abelian groups)**.** *Any finite abelian group is isomorphic to a direct product of cyclic groups. If $n = p_1^{k_1} \cdots p_t^{k_r}$ is the prime power factorization of n then the number of isomorphism classes of finite abelian groups of order n is $\mathfrak{p}(k_1) \cdots \mathfrak{p}(k_r)$, where $\mathfrak{p}(k)$ is the number of partitions of k.*

Proof. Combine Lemma 30.1 and Theorem 30.7. □

The function $\mathfrak{p}(k)$ appearing in the last theorem is called the *partition function*. There is no known explicit formula for $\mathfrak{p}(k)$. We are now finished with the classification of all finite abelian groups.

30.9 Example. Let $n = 1200 = 2^4 \cdot 3 \cdot 5^2$. Since $\mathfrak{p}(4) = 5$, $\mathfrak{p}(1) = 1$, and $\mathfrak{p}(2) = 2$, it follows that there are exactly $10 = 5 \cdot 1 \cdot 2$ different abelian groups of order 1200, up to isomorphism. We can easily list all ten isomorphism types of the abelian groups of order 1200. They are indexed by ordered pairs (λ, μ) of partitions, such that λ is a partition of 4 and μ is a partition of 2 (we can omit the unique partition of 1 from our indexing), as follows:

(λ, μ)	$G(2,\lambda) \times G(3,(1)) \times G(5,\mu)$
$((4),(2))$	$\mathbb{Z}_{2^4} \times \mathbb{Z}_3 \times \mathbb{Z}_{5^2}$
$((4),(1^2))$	$\mathbb{Z}_{2^4} \times \mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_5$
$((3,1),(2))$	$\mathbb{Z}_{2^3} \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_{5^2}$
$((3,1),(1^2))$	$\mathbb{Z}_{2^3} \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_5$
$((2^2),(2))$	$\mathbb{Z}_{2^2} \times \mathbb{Z}_{2^2} \times \mathbb{Z}_3 \times \mathbb{Z}_{5^2}$
$((2^2),(1^2))$	$\mathbb{Z}_{2^2} \times \mathbb{Z}_{2^2} \times \mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_5$
$((2,1^2),(2))$	$\mathbb{Z}_{2^2} \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_{5^2}$
$((2,1^2),(1^2))$	$\mathbb{Z}_{2^2} \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_5$
$((1^4),(2))$	$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_{5^2}$
$((1^4),(1^2))$	$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_5$.

The orders of the individual factors in each product above are called the *elementary divisors* of the group. Each isomorphism type is uniquely determined by its set of elementary divisors.

The acute reader might be alarmed to notice that the cyclic group of order 1200 appears to be missing from the above list. However, it is isomorphic to the first group on the list, so it is not actually missing.

Because of the fact that $\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$ whenever m, n are relatively prime, there are many other ways to express each of the above isomorphism types as products of cyclic groups. With this in mind, we rewrite the above table in a different looking form (but still isomorphic correspondingly):

(1200)	$\mathbb{Z}_{1200}$
$(5 \mid 240)$	$\mathbb{Z}_5 \times \mathbb{Z}_{240}$
$(2 \mid 600)$	$\mathbb{Z}_2 \times \mathbb{Z}_{600}$
$(10 \mid 120)$	$\mathbb{Z}_{10} \times \mathbb{Z}_{120}$
$(4 \mid 300)$	$\mathbb{Z}_4 \times \mathbb{Z}_{300}$
$(20 \mid 60)$	$\mathbb{Z}_{20} \times \mathbb{Z}_{60}$
$(2 \mid 2 \mid 300)$	$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{300}$
$(2 \mid 10 \mid 60)$	$\mathbb{Z}_2 \times \mathbb{Z}_{10} \times \mathbb{Z}_{60}$
$(2 \mid 2 \mid 2 \mid 150)$	$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{150}$
$(2 \mid 2 \mid 10 \mid 30)$	$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{10} \times \mathbb{Z}_{30}$.

Although the second table looks very different from the first, it is merely another way of describing the isomorphism types as products of cyclic groups. In this description, the products are indexed by sequences $(d_1 \mid d_2 \mid \cdots)$ of divisors of the group order such that each d_i divides the next d_{i+1} and such that the product of all the d_i is equal to the group order. Such sequences are called *invariant factors* of the group. There is an algorithm for going going back and forth between elementary divisors and invariant factors, which we leave to the reader.

It would be nice to go on to solve the problem of classifying non-abelian finite groups up to isomorphism. Alas, it is unknown how to do this in general. This remains an unsolved problem in group theory.

Exercises

30.1. Prove that $\mathbb{Z}_m \times \mathbb{Z}_n \cong \mathbb{Z}_{mn}$ if m, n are relatively prime.

30.2. Use the result of the previous exercise to express $\mathbb{Z}_{330}$ as an isomorphic product of simple groups. What are the composition factors of $\mathbb{Z}_{330}$?

30.3. Show that the composition factors of any abelian p-group are all isomorphic to $\mathbb{Z}_p$.

30.4. List all the partitions of 6 and 7. What is $\mathfrak{p}(6)$ and $\mathfrak{p}(7)$?

30.5. List all abelian groups of order 30, up to isomorphism, giving both the elementary divisors and invariant factors for each type.

30.6. List all abelian groups of order 48, up to isomorphism, giving both the elementary divisors and invariant factors for each type.

30.7. List all abelian groups of order 64, up to isomorphism, giving both the elementary divisors and invariant factors for each type.

30.8. List all abelian groups of order 100, up to isomorphism, giving both the elementary divisors and invariant factors for each type.

30.9. Prove that the number of isomorphism types of abelian groups of order 128 is 15.

30.10. Prove that there are 35 different abelian groups of order $2592 = 2^5 \cdot 3^4$ up to isomorphism.

30.11. Show that the generating function for the number $\mathfrak{p}(n)$ of partitions of n is

$$\sum_{n=0}^{\infty} \mathfrak{p}(n) x^n = \prod_{k=1}^{\infty} \left(\frac{1}{1 - x^k} \right).$$

This means that you can compute $\mathfrak{p}(n)$ by taking the coefficient of x^n in the product of the various geometric series expansions on the right hand side. This formula is due to Euler.

31 Presentations of groups

The idea of presenting a group by generators and relations is known as *combinatorial group theory*. The idea originated in a paper by Walther von Dyck in 1882; it is useful in algebraic topology and geometry. We give a brief crash course in the main ideas.

31.1 Definition. The *free group* on generators $a, b, c, \ldots$ is the group consisting of all possible strings (*e.g.* $aaba^{-1}bbbac^{-1}$) of the generators and their inverses, taken in any order, with the understanding that whenever a symbol x and its inverse x^{-1} appear next to one another, we are allowed to reduce the string by omitting the pair (*i.e.* we cancel xx^{-1} and $x^{-1}x$ whenever they appear). A string is *reduced* if no such cancellations are possible, and the elements of the free group are precisely the reduced strings of any length. Obviously, any free group is infinite, since there is no limit to the length of the strings.

By definition, there are no other relations between the generators in a free group, besides the relation which says we can cancel an element multiplied by its inverse. In a free group, as in any group, we allow ourselves to write a^n as a shorthand for the string $aaa\cdots a$ (n times repeated) and similarly a^{-n} as a shorthand for $a^{-1}a^{-1}\cdots a^{-1}$ (repeated n times).

An important element of the free group is the *empty string* ϵ, which is the only string of length zero. Given two strings s_1 and s_2 in a free group, we multiply then by juxtaposition; that is, $s_1 s_2$ is the string obtained by joining the symbols of s_1 with the symbols of s_2 to make a new string. For instance, if $s_1 = a^3b^2a^{-2}bab$ and $s_2 = b^{-1}a^{-2}b^5cb$ then

$$\begin{aligned} s_1 s_2 &= a^3b^2a^{-2}bab\,b^{-1}a^{-2}b^5cb \\ &= a^3b^2a^{-2}ba\,a^{-2}b^5cb \\ &= a^3b^2a^{-2}ba^{-1}b^5cb \end{aligned}$$

remembering our rules for reducing adjacent pairs of symbols and their inverse. It is easy to see that the free group is a group, with ϵ serving as its identity element. What is the inverse of a given string?

The free group on one generator a is obviously abelian, since the only strings we can form using one symbol a are powers of a, and powers of a surely commute with one another. It is easy to see that the free group G on a single generator a is isomorphic with the infinite (additive) cyclic group $\mathbb{Z}$; the isomorphism is given by the map $G \to \mathbb{Z}$ defined by the rule $a^m \to m$.

A free group with more than one generator is never abelian, since the equality $ab = ba$ for two generators a, b would be a non-trivial relation (and thus the group would no longer be "free").

31.2 Theorem. *Any group is isomorphic to a quotient of some free group.*

Proof. (Sketch) Here's a sketch of the procedure to prove this fact, and it gives a method of constructing the group as a quotient of a free group. Given a group G, let $S = \{a, b, c, \dots\}$ be a set of elements of G which generates G (so that $G = \langle S \rangle$). At worst, we could take $S = G$ but usually we can pick a much smaller generating set, as we have seen in numerous examples. Let F be the free group on the generators in the set S, and in the free group S let N be the subgroup consisting of all strings in the generators which evaluate to the identity in the given group G. Then one can prove that N is a normal subgroup of F, and moreover that $F/N \simeq G$. This is done by the first isomorphism theorem, using a natural homomorphism from F onto G. □

31.3 Remark. In practice, one would like to choose the generating set S to be as small as possible. Also, it is customary to specify the normal subgroup N by finding a (smallest possible, or at least nice in some way) set of generators for it. These generators of N are known as the *defining relations* of G.

31.4 Example. An example of a group given by generators and relations is the dihedral group $\mathbb{D}_n$, which is given by two generators r, d subject to the defining relations

$$r^n = 1; \quad d^2 = 1; \quad rdr = d.$$

It is easy to see that these relations hold in $\mathbb{D}_n$, but rather harder to show that any other relations between these generators will be consequences of these relations. To do this, you need to prove that the corresponding elements

$$r^n; \quad d^2; \quad rdrd$$

generate a normal subgroup N of the free group F on the generators r, h such that $F/N \simeq \mathbb{D}_n$. This can be shown with a bit of work.

31.5 Example. The symmetric group $\mathbb{S}_n$ is generated by transpositions, we have proved a long time ago. In fact, it is generated by the the *adjacent interchanges*; these are the special transpositions

$$t_1 = (1, 2), t_2 = (2, 3), \dots, t_{n-1} = (n-1, n).$$

It is not hard to verify that these $n-1$ generators satisfy the relations

$$t_i^2 = 1; \quad t_i t_j = t_j t_i \text{ (if } |i-j| > 1);$$
$$t_i t_{i+1} t_i = t_{i+1} t_i t_{i+1}$$

for all values of the indices i, j for which the equations make sense in $\mathbb{S}_n$. The last two relations are known as the *braid relations*.

It is harder to verify, yet true, that these relations actually *define* $\mathbb{S}_n$, in the sense that every other relation between generators is a consequence of these. In other words, the corresponding elements

$$t_i^2; \quad t_i t_j t_i t_j \text{ (if } |i-j| > 1); \quad t_i t_{i+1} t_i t_{i+1} t_i t_{i+1}$$

obtained from the relations (by writing each relation in the form $R = 1$) generates a normal subgroup N such that $\mathbb{S}_n \simeq F/N$ where F is the free group on the symbols $t_1, \ldots, t_{n-1}$.

31.6 Definition. In general, if G is a group given by generators $g_1, g_2, \ldots$ with defining relations $R_1, R_2, \ldots$ then we will write

$$G \simeq \langle g_1, g_2, \cdots \mid R_1, R_2, \ldots \rangle$$

to indicate this. Such a description of G (by generators and relations) is called a *presentation* of G.

A presentation $G = \langle g_1, g_2, \cdots \mid R_1, R_2, \ldots \rangle$ describes G as F/N where F is the free group generated by $g_1, g_2, \ldots$, and N is the normal subgroup of G generated by $R_1, R_2, \ldots$. Note that if one is given a generator R for N then R is a string in the symbols $g_1, g_2, \ldots$ (and their inverses) which is identity in G; sometimes the relation R is expressed as an equation $R = 1$, or any equation equivalent to it. For instance, one of the generators of the set of relations in $\mathbb{D}_n$ is $rdrd$, and it is customary to write this as the relation $rdrd = 1$ in $\mathbb{D}_n$, which is equivalent to the equation $rdr = d$ (since $d^2 = 1$).

31.7 Examples. In the notation just introduced, we write out presentations for each of the groups C_n = cyclic group of order n, $\mathbb{D}_n$ = the dihedral group of symmetries of a regular n-gon, and $\mathbb{S}_n$ = symmetric group on n letters:

$$C_n = \langle x \mid x^n = 1 \rangle$$
$$\mathbb{D}_n = \langle r, d \mid r^n = 1,\ d^2 = 1,\ rdr = d \rangle$$
$$\mathbb{S}_n = \langle t_1, \ldots, t_{n-1} \mid t_i^2 = 1,\ t_i t_j = t_j t_i \text{ (if } |i-j| > 1),\ t_i t_{i+1} t_i = t_{i+1} t_i t_{i+1} \rangle.$$

The presentation of $\mathbb{S}_n$ is called the *Coxeter presentation*. There is an elaborate theory of *Coxeter groups*, which are groups defined by generators and relations subject to certain conditions. Coxeter groups have applications to Lie groups.

Although it is satisfying to describe a group by means of generators and relations, there are some issues with this approach. Here are three fundamental problems formulated by Max Dehn in 1911. These problems are important for presentation theory as well as applications. Let G be a group defined by means of a given presentation. Dehn's problems are:

I. (Word problem) For an arbitrary word w in the generators, decide in a finite number of steps whether or not $w = 1$ in G.

II. (Conjugacy problem) For two arbitrary words w_1, w_2 in the generators, decide in a finite number of steps whether or not w_1 and w_2 are conjugate in G.

III. (Isomorphism problem) For an arbitrary group H defined by means of another presentation, decide in a finite number of steps whether or not $G \cong H$.

Unfortunately, it has been proven that all three problems are in general *undecidable* in the sense of mathematical logic. Roughly speaking, a problem is undecidable if it is not possible to design a Turing machine to solve it. It is generally understood that algorithms are equivalent to Turing machines, so the undecidability of these problems effectively means that there is no general algorithm to solve them.

That may sound rather discouraging, but in fact there are many classes of presentations for which all three problems have been solved, so the situation is not so dire as it may seem. Much more is known about combinatorial group theory; the book *Combinatorial Group Theory* by Magnus, Karass, and Solitar is a classic reference.

Exercises

31.1. Find a presentation by generators and relations of the Klein four group K_4, using two generators.

31.2. Show that the group given by the presentation $\langle a, b \mid a^5 = b^2 = 1, ba = a^2 b \rangle$ is isomorphic to $\mathbb{Z}_2$.

31.3. Show that the group $G = \langle x, y \mid x^2 = y^n = 1, xyx = y^{-1} \rangle$ is isomorphic to $\mathbb{D}_n$.

31.4. Find a minimal presentation of the group $\mathbb{Z}_2 \times \mathbb{Z}_3$.

31.5. What is the minimum number of generators needed to generate the group $G = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$? Find a presentation of this group.

31.6. Show that the quaternion group $Q = \{\pm 1, \pm i, \pm j, \pm k\}$, in which $i^2 = j^2 = k^2 = -1$, $(-1)^2 = 1$, and the symbols i, j, k multiply like standard unit vectors according to usual cross-product rules in $\mathbb{R}^3$, is presented by $\langle a, b \mid a^2 = b^2 = (ab)^2 \rangle$.

31.7. Artin's *braid group* $\mathbb{B}_n$ can be defined by the presentation

$$\langle t_1, \ldots, t_{n-1} \mid t_i t_j = t_j t_i \text{ (if } |i-j| > 1) \; ; \quad t_i t_{i+1} t_i = t_{i+1} t_i t_{i+1} \rangle.$$

Show that $\mathbb{S}_n$ is a homomorphic image of $\mathbb{B}_n$ and compute the kernel.

31.8. (a) Show that the symmetric group $\mathbb{S}_n$ is generated by the n-cycle $c = (1, 2, 3, 4, \ldots, n)$ and the transposition $t = (1, 2)$.

(b) (May be difficult) Find a set of defining relations on these generators that gives a presentation of $\mathbb{S}_n$ by these two generators and the relations you found.

31.9. (May be difficult) Find a presentation by generators and relations for the alternating group $\mathbb{A}_n$. [Hint: You may wish to start with the fact that $\mathbb{A}_n$ is generated by 3-cycles.]

Index