1

GROUPS and GEOMETRY:

A bridge to the Math major

Marshall M. Cohen **Kevin M. Pilgrim**

Marshall M. Cohen
Department of Mathematics
Malott Hall
Cornell University
Ithaca, NY 14853-4201
marshall@math.cornell.edu

Kevin M. Pilgrim
Department of Mathematics
Rawles Hall
Indiana University
Bloomington, IN 47405
pilgrim@indiana.edu

Contents

Preface

The beauty and unity of higher mathematics is nowhere better illustrated than in the interwoven subjects of group theory and geometry. The purpose of this book is to introduce the student to higher mathematics via a study of of the isometry group of the Euclidean plane. (Informally, an *isometry* is a *symmetry* or a *congruence in the sense of Euclid*).

Thus, **mathematically**, this book provides an introduction to group theory — usually thought of as a topic in "modern algebra" — and a study of the isometries of the Euclidean plane — a topic in "geometry". The combination of these topics culminates in a classification of the *wallpaper groups*. Wallpaper groups are precisely defined in Chapter V, but may be thought of for now as a means of classifying "wallpaper patterns" by certain common features.

Pedagogically. the purpose of this book is to present this rich and beautiful mathematical subject so as to help the student grow into modern mathematics. Thus we not only give careful rigorous proofs, but we try at each point to elucidate the mathematical culture — to bring the relatively inexperienced student into the world of modern mathematics. This is as important for those who will be users of mathematics (indeed, for any educated person in the 21^{st} century) as for those who hope to become mathematicians. In line with this pedagogical goal there are comments in the text, in footnotes and in Appendices, concerning the logic of proof, the nature of mathematical definition, the question of what it means for a function to be "well defined", etc.

To reinforce both our mathematical and pedagogical goals, the tool with which we study the isometries of the Euclidean plane is the geometry of complex numbers. Although we assume some familiarity with 2×2 matrices and sophomore linear al-

gebra, the heart of our discussion of isometries is in the language of functions of the complex plane to itself, which is introduced at the very outset. This enriches the mathematical discussion and further prepares the student for advanced work in mathematics, for example the study of complex analysis or hyperbolic geomety.

Our hope is that, at the end of a one-semester course, this book will have helped the reader to learn some geometry, some group theory and a bit about complex numbers. We hope that she or he will have experienced some of the richness and unity of mathematics, and will have become comfortable with the culture of mathematics and with the nature of mathematical definition (wherein we model our intuitions) and of mathematical proof. But, though this book can serve as a guide, these hopes can be achieved, of course, only if the reader works very hard, asks questions without embarrassment, discusses these topics with teacher and students, and enters into this world with abandon.

MMC, KMP

Chapter I

Symmetry in the plane

I.1 Introduction

In what ways is each of the plane figures drawn in Figures 1-6 symmetric? How do we recognize "**symmetry**"? (The figures are drawn on page 2.)

We respond in terms of the *motions of each figure onto itself*. We give some sample motions of each figure and we ask the reader (by adding the unfinished sentence "Further motions are ...") to join in and identify some motions for Figures 3 – 6 which we've omitted.

1. The triangle in Figure 1 may be rotated about its center by 1/3 of a complete turn either clockwise or counterclockwise (to bring the triangle back to itself). It may also be "flipped", or *reflected*, in the line joining a vertex to the midpoint of the opposite side.

2. The object in Figure 2 may be rotated by $1/6, 2/6, 3/6, 4/6$, or $5/6$ of a complete turn clockwise or counterclockwise.

3. The circle in Figure 3 can be rotated about its center by any angle. Further motions are ...

4. The graph in Figure 4 can be shifted, or *translated*, to the right or to the left. It can also be reflected through the vertical lines through the peaks and troughs. Further motions are ...

5. The pattern of circles in Figure 5 may be rotated by any angle about the common center. It may also be *contracted* or *expanded* by a factor of 2; i.e. if we shrink or magnify the picture by a factor of 2, the pattern is brought back to itself (though an individual circle will go to another circle). Further motions are ..

6. The pattern in Figure 6 may be translated horizontally or vertically. It may also be rotated by π radians about the centers of the rectangles and reflected in vertical lines passing through the sides of the rectangles. Further motions are ...

In each of the above cases, something which we do to the figure to bring it to itself, like rotation by a 1/3 turn counterclockwise in Figure 1.1, is called a *symmetry* of the figure.

Symmetries as functions

In our attempt to *describe* what the symmetries of the above figures are, we used phrases like "may be rotated to bring it back to itself". That is, we are implicitly thinking of a symmetry as a transformation, or *function* which carries the figure to itself.

If we wish to model a symmetry of a plane figure using mathematics, we are then led to using the language of functions. At this point, some care is in order, since a function consists of three ingredients:

a *domain*, a *range*, a *rule*.

The symmetry of the triangle in Figure 1 given by rotating it by 1/3 turn counterclockise can be realized by rotating the entire plane by 1/3 turn about the center of the triangle. That is, this symmetry of the triangle can be thought of as a function from the plane to itself: the domain and range are both the whole plane, and the rule is "rotate the plane counterclockwise by 1/3 turn about the center of the triangle". Similarly, the symmetries in the other examples can be thought of as functions from the plane to itself.

There are some important consequences of this point of view.

First, notice that if, for example, we rotate the triangle in Figure 1 by one full turn (equivalently, by 360 degrees or by 2π radians), then every point is sent to itself; i.e. no point is moved at all. Thus, as a *function*, rotation by one full turn is the *identity function* on the plane. Similarly, the counterclockwise rotation by 1/3 turn ($= 2\pi/3$ radians) mentioned above, gives the same function from the plane to itself as does clockwise rotation by a 2/3 turn ($4\pi/3$ radians).

Note, in particular, that we will not pay attention to the process in time by which the final result is achieved — only to the resulting function. A rich theory would occur if we did take the time process into account, but we won't take that road.

Second, since we view a symmetry as a function from the plane to itself, we may *compose* these symmetries. So for example if we first reflect the triangle in a line, and then apply the same reflection again, then, as a function, the resulting composition is a function which does not move any point, i.e. composing a reflection with itself gives the identity function on the plane.

Special properties of these functions

Finally, the particular functions we have identified as symmetries of the figures above are rather special.

To explain this last point, consider the following kinds of functions f which arise as symmetries in the above examples: *rotation* about a given point; *translation* in a given direction by a given magnitude; *reflection* in a given line; *dilating* or *expanding* by a given factor. Although we have not yet precisely defined these functions, we can still see that they have certain special properties.

Each of these functions is *invertible*. Loosely, this means that each of these functions can be "undone". For example, if we rotate by 1/3 turn counterclockwise, and then rotate by 1/3 turn clockwise, we get the identity function. Similarly, if we reverse the order of composition, we again get the identity function. More formally, there exists another function, which we call f^{-1}, from the plane to itself such that the two compositions $f \circ f^{-1}$ and $f^{-1} \circ f$ are *both* the identity function on the plane. This is equivalent to saying that f, as a function from the plane to itself, is a *bijection*; see Appendix D for the definition of bijection.

Further, rotations, translations, and reflections preserve the shape and size of figures. In particular, they *preserve the distance between points.* That is, if we fix some unit of measurement of distance on the plane, then the distance between two points P and Q is the same as the distance between the points $f(P)$ and $f(Q)$, where f is a rotation, translation, or reflection.

In particular, suppose Δ is a triangle with vertices A, B, C. Then $f(\Delta)$ is a triangle with vertices $A' = f(A), B' = f(B), C' = f(C)$. Moreover, the fact that f preserves distances implies that $\overline{AB} = \overline{A'B'}$, $\overline{BC} = \overline{B'C'}$, and $\overline{CA} = \overline{C'A'}$, i.e. that lengths of corresponding sides are the same.

Although contractions and expansions, do not preserve distances, they do preserve *ratios* of distances. So, if f is a contraction or expansion, then using the notation above, we have $\overline{AB}/\overline{A'B'} = \overline{BC}/\overline{B'C'} = \overline{CA}/\overline{C'A'}$, i.e. that the ratios of corresponding sides are the same.

Often the triangles Δ, Δ' in the first case are said to be *congruent*, and those in the

second case are said to be *similar*.

Choosing a definition of symmetry

How, then, should we precisely *define* a symmetry of a plane figure? Should we include contractions and expansions, or only distance-preserving functions? We need to choose how special we want our functions arising as symmetries to be, i.e. we need to choose a particular collection of functions with which to work. There are many possible choices. The point is that, once a choice has been made, we will get what one might call the corresponding *theory*, or collection of theorems. For example, if we choose our symmetries to include dilations and expansions, then we can prove that any symmetry of a figure will take an angle onto an equal angle, but will not necessarily take a curve (like one of the circles above) to a curve of equal size. If we made our notion of symmetry very restrictive by requiring that a symmetry of a plane figure be a translation which carries the figure to itself, then the triangle in Figure 1 would have no symmetries at all, other than the identity function!

In this book, we choose to define a **symmetry** of a plane figure P to be

a distance-preserving function from the plane
to itself which takes the figure P onto P.

A distance-preserving function from the plane to the plane is often called a *Euclidean isometry* and we will refer to it simply as an *isometriy*. See I.7.1 and I.4.1 for the formal definitions.

A brief outline of this book

To keep the big picture in mind, the reader is encouraged to re-read this outline as the course progresses.

In this chapter, we give (§§I.2, I.3) several ways of thinking about the Euclidean plane and functions from the plane to itself. In §I.4 we give examples and general

properties of symmetries. In §I.5 and I.6 we classify the symmetries – say exactly which functions they are and what are the formulas for them. In §I.7, we abstract four crucial properties enjoyed by the set of symmetries of a plane figure. These four properties, when abstracted, become the axioms for a mathematical object known as a *group*. In Chapter II, we introduce the abstract notion of group.

In Chapter II, we begin (§II.1 and II.2) by listing axioms, basic properties, and many numerical examples. The groups which arise as symmetry groups of plane figures, however, are special–they are examples of *transformation groups*, which we study by themselves in §II.4, II.5, and II.8. In §II.6 and II.7 we return to the study of abstract groups by looking *inside* one group to find other groups, called *subgroups*.

In Chapter III, we return to the study of isometries, having now developed tools from the theory of groups. In §III.1 these tools are used to shed light on the earlier Classification Theorem. In §III.2, *congruence* of plane figures is closely related to *conjugacy* of isometries of the plane. Here, the interplay between algebra and geometry is profound, and the results help us understand the structure of the group of isometries of the plane. In §III.3 we discuss the relationship between the group of all isometries of the plane and the subgroup of isometries which send the origin to itself. We conclude in §III.4 with a classification of the finite subgroups of the group of plane isometries.

In chapter IV, we follow a fundamental mathematical practice by looking not just at objects with some structure (in our case, groups) but also at functions between them which preserve this structure. In the setting of groups, these functions are called *homomorphisms*. The definition and examples are given in §IV.1. The study of homomorphisms is then shown in §IV.2-3 to be intimately connected with the study of subgroups via the *kernel* of a homomorphism, which is a special kind of subgroup – a *normal subggroup*. In §IV.4 the construction of the *quotient group* of a group modulo a normal subgroup is introduced. These concepts are united in the First Isomorphism Theorem – one of the high points of this book. This theorem leads us to the *group extension problem*, which examines how a group can be built from two other groups (the kernel and image of a homomorphism) in a way which will be particularly cogent for our analysis of the isometry groups — wallpaper groups — which are studied in the final chapter. We lay the ground for this analysis in §IV.6 by introducing a certain homomorphism which sends each isometry to an isometry which fixes the origin.

In Chapter V, all of the preceding algebra and geometry coalesce to give a

classification of the type of symmetry groups seen in classical two-dimensional artistic patterns — the *wallpaper groups*. We give (1) a precise criterion for calling two wallpaper groups "the same"; (2) a list of groups of the type being discussed [there turns out to be 17 wallpaper groups], and (3) a proof showing that *any* wallpaper group is "the same as" one of the groups on the list. Wallpaper groups are quite special in the sense that they are *discrete* and allow translations in two independent directions. Some new ideas particular to the study of discrete groups enter here as well. In particular, the classification of wallpaper groups depends on the *chrystallographic restriction* which says that only very special angles can appear in such infinitely repeating discrete patterns.

In several places, certain aspects of mathematical culture may appear with which the reader is unfamiliar: the language of implication (if–then, if and only if, suffices, etc.); the language of functions (one-to-one, onto, bijection, well-defined); the critical reading of proofs; the construction of examples and counterexamples; the complex exponential. As an aid, several Appendices on these and other topics are included and referenced in what follows.

Exercises I.1

1. Figures 3-6 admit symmetries other than the ones described. For each figure, find as many symmetries as you can.

2. The triangle in Figure 1 and the object in Figure 2 both have six symmetries, counting the "trivial" symmetry which is just the identity function on the plane. Are there nonetheless differences between these two sets of symmetries?

3. Figures 3-6 all admit infinitely many symmetries. Are there nonetheless differences between these two sets of symmetries?

4. Find two symmetries f and g of the real line (i.e., the x-axis) which take each real number x to $-x$, but such that f and g are not the same functions of the plane to the plane. *Note: part of this exercise is to precisely formulate this question.*

5. Find all the symmetries you can of the wallpaper pattern shown in Figure 7. Compare and contrast this collection of symmetries with that in Figure 6.

Fig. 7. A "wallpaper pattern"

I.2 The Euclidean plane and complex numbers

In this section, we present several different ways of thinking about the plane. Throughout this book we will pass back and forth between these various points of view as it suits our purposes. We discuss below how we may think of the plane

1. as the usual Cartesian "x-y" plane;
2. as the plane of *complex numbers* $z = x + iy$;
3. as a *real vector space*.

Notation. Throughout this book, we will denote by

- $\mathbb{Z}$, the set of all integers $\{... - 2, -1, 0, 1, 2, ...\}$;
- $\mathbb{R}$, the set of all real numbers;
- $\mathbb{R}^2$, the set of all *ordered pairs* of real numbers (x, y);
- $\mathbb{C}$, the set of all *complex numbers* $z = x + iy$.

By *ordered pair* we mean, for example, that the pair $(1, 2)$ is different from the pair $(2, 1)$. We think of $\mathbb{R}$ as a line and of $\mathbb{R}^2$ as a plane.

The fundamental identification

As sets, we identify the Cartesian plane $\mathbb{R}^2$ with $\mathbb{C}$ by

$$(x, y) \longleftrightarrow z = x + yi.$$

We will take this point of view throughout most of this book. To make the connection explicit, we introduce some terminology:

Definition I.2.1 (Real and imaginary parts) *Let $z = x+yi \in \mathbb{C}$. The real number x is called the* real part *of z, denoted* $\mathrm{Re}(z)$*, and the real number y is called the* imaginary part *of z, denoted* $\mathrm{Im}(z)$.

So the complex number $1 + 2i$ can be thought of as the point $(1, 2)$ in the plane $\mathbb{R}^2$. Since $\mathbb{R}^2$ is also a vector space, we can think of $1+2i$ as the vector with component 1 in the x-direction and 2 in the y-direction. We think of the real numbers as a *subset* of the complex numbers via $x \in \mathbb{R}$ corresponding to $x + 0i \in \mathbb{C}$.

Conventions: We usually write, for example, 7 instead of $7+0i$. Similarly, we write $4i$ instead of $0 + 4i$, and 0 instead of $0 + 0i$.

Complex addition and multiplication

Definition I.2.2 (Complex arithmetic) *Let $z = a + bi$ and $w = c + di$. Complex addition and multiplication are defined by*

$$\begin{aligned} z + w &= (a + c) + (b + d)i \\ zw &= (ac - bd) + (ad + bc)i \end{aligned}$$

and for a nonzero complex number $z = x + yi$, the reciprocal of z is given by

$$\frac{1}{z} = \left(\frac{x}{x^2 + y^2}\right) + \left(\frac{-y}{x^2 + y^2}\right) i.$$

Some basic facts about complex arithmetic are:

- **Familiar properties hold.** E.g. for all complex numbers z, w, u we have

$$\begin{aligned} z + w &= w + z, \qquad zw = wz, \qquad z(w + u) = zw + zu, \\ 0 + z &= z, \qquad 1z = z, \\ z\tfrac{1}{z} &= 1 \quad (z \neq 0), \quad \text{which justifies the name “reciprocal”.} \end{aligned}$$

  See Exercise 1.

- **Extension of ordinary arithmetic.** Since $\mathbb{R}$ can be thought of as a subset of $\mathbb{C}$, it is reasonable to pause and ask if the complex arithmetic operations, when applied to real numbers, yield the same answer as their ordinary, real counterparts. Indeed, they do. [1]

- **Algebraic completeness.** The *Fundamental Theorem of Algebra* says that every polynomial of degree $d > 0$, $p(z) = a_d z^d + ... + a_1 z + a_0$ $(a_0, \ldots a_d \in \mathbb{C})$, has a root in the complex plane, i.e. there is a complex number r such that $p(r) = 0$. For example, the polynomial $z^2 + 1$ has as one root i, since $i^2 = -1$, and $z^2 + i$ has as one of its roots the complex number $-(\sqrt{2}/2) - (\sqrt{2}/2)\, i$. Though important, we will not need this theorem.

Because of these facts, we can be more casual in writing a complex number – e.g., $5 + (-6)i = 5 - 6i$ and $\frac{3}{5} + \frac{4}{5}\, i = \frac{3+4i}{5}$.

Complex arithmetic has a fundamental symmetry:

Definition I.2.3 (Complex conjugation) *Let $z = x + yi$. The* complex conjugate *of z, denoted $\overline{z}$, is defined by $\overline{z} = x - yi$.*

By a symmetry of the arithmetic, we mean the following:

Proposition I.2.4 (Properties of conjugation) *Complex conjugation has the following properties: for all $z, w \in \mathbb{C}$,*

1. $\overline{\overline{z}} = z$
2. $\overline{z + w} = \overline{z} + \overline{w}$
3. $\overline{zw} = \overline{z}\ \overline{w}$
4. $\overline{\left(\frac{1}{z}\right)} = (1/\overline{z}) \qquad (z \neq 0)$
5. $\overline{z} = z$ *if and only if z is a real number.*

The proof is Exercise 2.

[1] See Appendix D for more on the idea of "extending" functions.

Arithmetic and the fundamental identification

Definition I.2.5 (Vector arithmetic) *Let* $\vec{v} = (a, b), \quad \vec{w} = (c, d) \quad \in \mathbb{R}^2$ *be vectors and* $\lambda \in \mathbb{R}$ *a scalar. Vector addition and scalar multiplication are defined by*

$$\begin{aligned} \vec{v} + \vec{w} &= (a + c \quad , \quad b + d) \\ \lambda\vec{v} &= (\ \lambda a \quad , \quad \lambda b\ ). \end{aligned}$$

Since we've been viewing the same plane in two different ways, we should look for relationships between the complex and vector operations. Under the identification of the complex plane with the vector space $\mathbb{R}^2$ we have

- **complex addition corresponds to vector addition:**
  If $z = a + bi = (a, b) = \vec{z}, \quad w = c + di = (c, d) = \vec{w}$, then $z + w = \vec{z} + \vec{w}$,
- **multiplication of a complex number by a real number corresponds to scalar multiplication:** In the notation above,
  if λ is a real number, then the complex number $\lambda z = \lambda\vec{z}$.

The geometry of complex arithmetic

Geometrically, addition of complex numbers (equivalently, vectors in the plane) is done via the so-called *Parallelogram Law*; see Figure 1. Note also that multiplication of a complex number by a positive real scalar (equivalently, multiplication of a vector by a positive scalar) preserves direction but alters "size"; see Figure 2.1. More precisely:

Definition I.2.6 (Modulus) *The* modulus*, or* absolute value*, of a complex number* $z = x + yi$*, denoted* $|z|$*, is the nonnegative real number* $|z| = \sqrt{x^2 + y^2}$.

Geometrically, the modulus is the distance from the origin 0 to the point z. Some properties of the modulus are given below:

Fig. 1. The parallelogram law for vector addition.

Proposition I.2.7 (Properties of the modulus) *The function*[2] $z \mapsto |z|$ *has the following properties: for all* $z, w \in \mathbb{C}$,

1. $|z| = 0$ *if and only if* $z = 0$.
2. $z\overline{z} = |z|^2$
3. $|\overline{z}| = |z|$
4. $\left|\frac{1}{z}\right| = \frac{1}{|z|}$
5. $|zw| = |z|\ |w|$
6. $|z + w| \leq |z| + |w|$;
   $|z + w| = |z| + |w|$ *with* $w \neq 0 \iff z = \lambda w$ *for some* $\lambda \geq 0$.

The proof is Exercise 3. ∎

[2]See Appendix D for a discussion of the symbol $\mapsto$.

To formulate a vector analog of the modulus of a complex number, we need a way of measuring lengths of vectors. This is usually done by means of an *inner*, or *dot* product; see Appendix F.

Definition I.2.8 (Inner product on $\mathbb{R}^2$) *If $\vec{u} = (a, b))$ and $\vec{v} = (c, d)$, the* inner product *of $\vec{u}$ and $\vec{v}$, denoted by $\langle \vec{u}, \vec{v} \rangle$, is the real number given by*

$$\langle \vec{u}, \vec{v} \rangle = ac + bd.$$

In particular, if $\vec{u} = \vec{v} = (x, y)$, we see that

$$\langle \vec{v}, \vec{v} \rangle = x^2 + y^2$$

which is the square of the distance from the origin $(0, 0)$ to (x, y).

Definition I.2.9 (Norm of a vector) *The* norm *of a vector $\vec{v} = (x, y)$, denoted by $||\vec{v}||$, is the nonnegative real number given by*

$$||\vec{v}|| = \sqrt{\langle \vec{v}, \vec{v} \rangle}.$$

Thus, the norm of a vector $\vec{v} = (x, y)$ is, equivalently:

- the length of $\vec{v}$;
- the distance from the origin to (x, y);
- the modulus of the complex number $z = x + yi$.

Using the modulus of a complex number, or equivalently, using the norm of a vector, we can define distances:

Definition I.2.10 (Distance in the plane) *If $z, w \in \mathbb{C}$ the* distance between z and w *is the nonnegative real number $|z - w|$.*

This is the usual distance function in the plane, and it may be expressed in coordinate form or vector form, as well as in the complex form:
if $z = a + bi = (a, b) = \vec{z}$ and $w = c + di = (c, d) = \vec{w}$, then

$$|z - w| \;=\; \sqrt{(a-c)^2 + (b-d)^2} \;=\; ||\vec{z} - \vec{w}||$$

These three ways of defining distance are therefore all the same. The complex notation is rather compact, so it is the one which we adopt:

Proposition I.2.11 (Properties of distance) *The distance function* $(z, w) \;\mapsto\; |z - w|$ *satisfies the following properties:*

1. **Nondegeneracy.** *For all* $z, w \in \mathbb{C}$, $|z-w| \geq 0$. *Equality holds* $\iff z = w$.
2. **Symmetry.** *For all* $z, w \in \mathbb{C}$, $|z - w| = |w - z|$.
3. **Triangle Inequality.** *For all* $z, w, u \in \mathbb{C}$,

$$\begin{aligned} |z - w| &\leq |z - u| + |u - w|. \\ |z - w| &= |z - u| + |u - w| \iff u = (1-t)z + tw \;\; \textit{for some } t \textit{ with } 0 \leq t \leq 1. \end{aligned}$$

The proof, using properties of the modulus (Proposition I.2.7), is left to the reader as Exercise 4. ∎

Polar coordinates

Any point $(x, y) \in \mathbb{R}^2$ can be written in *polar coordinates*

$$(r, \theta), \quad r \geq 0, \; \theta \in \mathbb{R}$$

where $r = \sqrt{x^2 + y^2}$ and where, if $(x, y) \neq (0, 0)$, θ is the angle, measured in radians and proceeding counterclockwise, from the positive $x-$axis to the ray joining $(0, 0)$ to (x, y); see Figure 2. Note that if $z = x + yi = (x, y)$, then $r = \sqrt{x^2 + y^2}$ is the

Fig. 2. Polar coordinates.

modulus $|z|$ of the complex number z. The point $(0,0) \in \mathbb{R}^2$ has non-unique polar coordinates $(0,\theta)$, θ any real number.

We emphasize here that we require $r \geq 0$. [3]

A subtlety is that *there is more than one possible choice of polar coordinate for a given point* (x,y). In fancier language, *polar coordinates, unlike Cartesian coordinates, are not unique.* For example, the point $(1,1)$ in Cartesian coordinates can be specified in polar coordinates by $(\sqrt{2}, \pi/4)$ or by $(\sqrt{2}, -7\pi/4)$. In general, the polar coordinates

$$(r_1, \theta_1) \text{ and } (r_2, \theta_2)$$

represent the same point of the plane if and only if either

- $r_1 = r_2 = 0$ (in which case both points are the origin), or
- $r_1 = r_2$ and $\theta_2 - \theta_1 = 2n\pi$ for some $n \in \mathbb{Z}$.

[3] In calculus, you may have looked at parametric curves in polar coordinates where the radial coordinate r was allowed to be negative. In contrast, we adopt the convention that r is always non-negative.

The Cartesian coordinates (x, y) of a point can be recovered from the polar coordinates (r, θ) by

$$x = r\cos\theta,\ y = r\sin\theta.$$

One way around this non-uniqueness is to require that the set of possible angles θ lie in some restricted set, e.g. $\theta \in (-\pi, \pi]$. With this requirement, the polar coordinates of any point other than the origin are indeed uniquely defined.

Definition I.2.12 (Arg and arg) *Let $z = x + yi = (x, y)$ be a nonzero complex number. The* argument *of z, denoted* $\arg(z)$*, is the* set *of all possible angular coordinates for the point. The* standard argument *of z, denoted* $\mathrm{Arg}(z)$*, is the* particular *angular coordinate whose values lie in $[0, 2\pi)$. Thus*

$$\arg(z) = \{\theta \in \mathbb{R} \mid \theta = \mathrm{Arg}(z) + 2n\pi, \quad n \in \mathbb{Z}\}.$$

Note that, while $\mathrm{Arg} : \mathbb{C} - \{0\} \to \mathbb{R}$ is a (non-continuous) function, arg is *not* a function from $\mathbb{C} - \{0\}$ to the real numbers.

Polar coordinates and the complex exponential

Using polar coordinates, we can give a geometric interpretation of complex multiplication.

Definition I.2.13 (Complex exponential) *For $\theta \in \mathbb{R}$, the* complex exponential $e^{i\theta}$ *is defined by*

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

Note: This notation for $\cos\theta + i\sin\theta$ can be motivated by considering the power series for each of the functions involved (Exercise 5).
Geometrically this function wraps the line around the unit circle.

Proposition I.2.14 (Properties of the complex exponential) .
The complex exponential function taking $\mathbb{R}$ to $\mathbb{C}$ by the rule $\theta \mapsto e^{i\theta}$ satisfies:

1. $e^{i\,0} = 1$.
2. *If* $z \in \mathbb{C}$ *then* $|z| = 1 \iff$ *there exists* $\theta \in \mathbb{R}$ *such that* $z = e^{i\theta}$.
3. *for all* $\theta_1, \theta_2 \in \mathbb{R}$, $e^{i(\theta_1+\theta_2)} = e^{i\theta_1}e^{i\theta_2}$.
4. *For all* $\theta \in \mathbb{R}$, $e^{i(-\theta)} = \overline{e^{i\theta}} = \frac{1}{e^{i\theta}}$.

The proof is Exercise 6. ∎

Let $z = x + yi$ have polar coordinates r, θ. Since $x = r\cos\theta$, and $y = r\sin\theta$ we have

$$\begin{aligned} z &= r\cos\theta + ir\sin\theta \\ &= r(\cos\theta + i\sin\theta) \\ &= re^{i\theta}, \end{aligned}$$

where we have used the definition of the complex exponential. Combining this with our observations about polar coordinates in the plane, we see that any nonzero complex number z may be written uniquely in the form

$$z = re^{i\theta}, \quad r > 0,\ \theta \in [0, 2\pi)$$

where (r, θ) are the polar coordinates of z, i.e. $r = |z|$ and $\theta = \mathrm{Arg}(z)$. Since $e^{i\theta_1} = e^{i\theta_2}$ if and only if $\theta_1 - \theta_2 = 2n\pi$ for some $n \in \mathbb{Z}$, we see that if $z_1 = r_1e^{i\theta_1}$ and $z_2 = r_2e^{i\theta_2}$ then $z_1 = z_2$ if and only if either $r_1 = r_2 = 0$ (in which case $z_1 = z_2 = 0$), or $r_1 = r_2 > 0$ and $\theta_1 - \theta_2 = 2n\pi$ for some $n \in \mathbb{Z}$.

We're now ready to give the geometric version of complex multiplication.

Proposition I.2.15 (Geometry of complex multiplication) *Let* $z, w \in \mathbb{C}$. *If, in polar coordinates,*

$$z = (r_1, \theta_1) \quad \text{and} \quad w = (r_2, \theta_2)$$

then, in polar coordinates,

$$zw = (r_1r_2,\ \theta_1 + \theta_2).$$

In other words, *to multiply two complex numbers, multiply their moduli, and add their arguments.* See Figure 3.

Fig. 3. To multiply two complex numbers, multiply their moduli and add their arguments.

Proof: We have $z = r_1 e^{i\theta_1}$ and $w = r_2 e^{i\theta_2}$, so

$$zw = r_1 e^{i\theta_1} r_2 e^{i\theta_2} = r_1 r_2 e^{i\theta_1} e^{i\theta_2} = r_1 r_2 e^{i(\theta_1+\theta_2)}$$

where we have used property 3. of Proposition I.2.14. ∎

Orthogonality

Orthogonality is nicely expressed in terms of complex arithmetic. The reader is probably acquainted with the fact (Exercise 7) that $\langle \vec{v}, \vec{w} \rangle = ||\vec{v}||\ ||\vec{w}||\ \cos\theta$ where θ is the angle (measured either clockwise or counterclockwise) between $\vec{v}$ and $\vec{w}$.

Two nonzero vectors $\vec{v}, \vec{w}$ are called *orthogonal* if the angle between them is a right angle. This holds $\iff \cos\theta = 0 \iff \langle \vec{v}, \vec{w} \rangle = 0$.

The geometric interpretation of complex multiplication given above — add the arguments and multiply the moduli — leads us to an expression of orthogonality in terms of complex numbers. View the complex number z as a vector from the origin. Then every vector orthogonal to z is gotten by multiplying z by $re^{i\pi/2}$ or by $re^{-i\pi/2}$. But $e^{\pm i\pi/2} = \pm i$. Hence, we have the proposition:

Proposition I.2.16 *If $z, w \in \mathbb{C}$ are thought of as vectors from the origin, with $w \neq 0$, then*

$$\begin{aligned} w \text{ is orthogonal to } z \quad &\Longleftrightarrow \quad \frac{z}{w} = \pm ri = \text{a pure imaginary number} \\ &\Longleftrightarrow \quad Re\left(\frac{z}{w}\right) = 0. \end{aligned}$$

■

Geometric objects

We consider **circles** and **lines**.

A circle with center $c \in \mathbb{C}$ and radius $r > 0$ is the set of all points whose distance from c is equal to r. Hence this circle can be written as

$$\{z \mid \quad |z - c| = r\}$$

and, when written in this form, the numbers c, r are unique.

Lines, however, are a bit more complicated to describe. For example, in $\mathbb{R}^2$, the lines defined by the equations

$$y = 2x + 1 \quad \text{and} \quad 20x - 10y = -10$$

are the same.

Definition I.2.17 (Lines and rays) *A* line *is a subset L of $\mathbb{C}$ of the form*

$$L = \{z \mid z = z_0 + tz_1,\ t \in \mathbb{R}\},$$

where $z_0, z_1 \in \mathbb{C}$ and $z_1 \neq 0$. If we require $t \geq 0$ the resulting set is called a ray.

Notice that, by definition, lines, rays and segments are *subsets* of $\mathbb{C}$. For the line L given in the definition, the vector described by the complex number z_1 points in one direction and the vector $-z_1$ points in the other direction. If we wish to think of L

as having a direction, then, since the direction being pointed out does not depend on the length of z_1, we define the *direction* of L to be the unit vector

$$\text{direction of } L \;=\; \frac{z_1}{|z_1|}$$

and we say that "the direction of L is determined by z_1". If we do not wish to think of the line as having a preferred direction, then we say that "the direction of L is determined up to sign by z_1" or "is determined by $\pm z_1$" and

$$\text{the direction of } L \text{ up to sign} = \pm\frac{z_1}{|z_1|}.$$

However, our *description* of a line given in the definition, unlike that of a circle, is not unique. So, we need to know when two different *descriptions* yield the same line.

Proposition I.2.18 (When lines coincide) *Let $z_0, w_0, z_1, w_1 \in \mathbb{C}$ with $z_1 \neq 0$, $w_1 \neq 0$, and let lines L, L' be defined by*

$$\begin{aligned} L &= \{z \mid z = z_0 + tz_1,\ t \in \mathbb{R}\} \qquad \textit{and} \\ L' &= \{z \mid z = w_0 + sw_1,\ s \in \mathbb{R}\}. \end{aligned}$$

Then $L = L'$ if and only if the following two conditions are satisfied:

1. *L and L' have the same direction up to sign, i.e.*

$$w_1 = \lambda z_1 \quad \textit{for some } \lambda \in \mathbb{R},\ \lambda \neq 0,$$

2. *L and L' have a point in common, i.e.*

$$z_0 + sz_1 = w_0 + tw_1 \quad \textit{for some } s, t \in \mathbb{R}.$$

Proof: We first show that $L = L'$ implies 1. and 2.
Clearly the lines have a point in common if $L = L'$, so 2. is satisfied. To see that 1. is satisfied, note that if $L = L'$ then z_0 is on L' and w_0 is on L. Therefore

$$z_0 = w_0 + sw_1 \qquad \text{and} \qquad w_o = z_0 + tz_1, \quad \text{for some } s, t \in \mathbb{R}$$

Therefore $z_0 - w_0 = sw_1 = -tz_1$.

If $z_0 \neq w_0$ then $s \neq 0$ and $t \neq 0$ so we may conclude that

$$w_1 = \lambda z_1 \quad \text{where} \quad \lambda = -\frac{t}{s},$$

and 1. is proved in this case.

Otherwise, if $z_0 = w_0$, then $z_0 + z_1 \in L = L'$ and $z_0 + z_1 \neq w_0$. Therefore

$$z_0 + z_1 = w_0 + sw_1 = z_0 + sw_1 \quad \text{for some} \quad s \neq 0.$$

Thus, in this case also, $w_1 = \lambda z_1$, where $\lambda = s^{-1} \neq 0$, and 1. is proved in this case. The two cases cover all possibilities.

Next we show that 1. and 2. imply that $L = L'$.
Let a, b be numbers given by 2. such that $z_0 + az_1 = w_0 + bw_1$. Hence

$$\begin{aligned} L &= \{z \mid z = z_0 + tz_1,\ t \in \mathbb{R}\} \\ &= \{z \mid z = (w_0 + bw_1 - az_1) + tz_1,\ t \in \mathbb{R}\} \\ &= \{z \mid z = w_0 + (b - a\eta + t\eta)w_1,\ t \in \mathbb{R}\} \quad \text{(using 1. to write } z_1 = \eta w_1,\ \eta \neq 0) \\ &= \{z \mid z = w_0 + tw_1,\ t \in \mathbb{R}\} \\ &= L' \end{aligned}$$

The second-to-last equality follows from the fact that as t ranges over all of $\mathbb{R}$, so does $(b - a\eta + t\eta)$, since a, b and η are constants. ∎

The parametric definition of a straight line given above is not as natural as the notion which we learn early in our education that "the shortest distance between two points is a straight line". The latter is a characterization of straight lines in terms of distance, which we express in the following Proposition.[4]

Proposition I.2.19 *If $z, w \in L \subset \mathbb{C}$, then L is a straight line if and only if L consists precisely of all the points $u \in \mathbb{C}$ such that*

$$\begin{aligned} & |u - z| + |z - w| &= |u - w| \\ OR \qquad & |z - u| + |u - w| &= |z - w| \\ OR \qquad & |z - w| + |w - u| &= |z - u|. \end{aligned}$$

[4]Euclid's Definition 4 expressed this by saying "*A straight line is a line which lies evenly with the points on itself.*" [T. Heath, Euclid, The Elements, Vol. 1.]

The **proof** of this proposition uses the triangle inequality (Proposition I.2.11) and is left to the reader as Exercise 8. ∎

Some familiar geometric properties

We say that lines L and L' are *parallel* if they have the same direction up to sign. We say that points $z_0, z_1, \ldots, z_n$ are *collinear* if they all lie on a common line L.

To gain some fluency working with complex descriptions of lines, you will be asked to prove

- (Exercise 2.8) Given any pair of distinct points z_0 and $w_0 \in \mathbb{C}$, there is exactly one line containing both z_0 and w_0.
- (Exercise 2.9) If lines L and L' are parallel, then they are equal or disjoint.

Exercises I.2

1. Prove from the definitions that the familiar axioms for addition and multiplication listed below the definitions hold.
2. Prove Proposition I.2.4. *Note:* in part 5., you must prove two things:
   - *if* $\overline{z} = z$, *then* z is real, and
   - *if* z is real, *then* $\overline{z} = z$.

   See Appendix A for more discussion of implications.
3. Prove Proposition I.2.7. *Note:* Again, when a claim of if-and-only-if is made, there are *two* implications to prove. The only tricky part, really, is 6. To show this, start with $|z + w|^2 = (z + w)(\overline{z + w})$, expand and use properties of conjugation.
4. Prove Proposition I.2.11. *Hint:* appeal to Proposition I.2.7.

5. Motivate the definition of $e^{i\theta}$ by writing out the power series for $\cos\theta$, $\sin\theta$ and (assuming that complex power series work the same as real power series) for $e^{i\theta}$.

6. Prove Proposition I.2.14. *Hint:* for 3., use addition formulas for sin and cos.

7. (a) Prove that
$$\langle \vec{v}, \vec{w} \rangle = ||\vec{v}||\ ||\vec{w}||\ \cos\theta$$
where θ is the angle (measured either clockwise or counterclockwise) between $\vec{v}$ and $\vec{w}$. *Hint:* use properties of inner products to reduce to the case when $\vec{v}$ and $\vec{w}$ are unit vectors. Then draw a picture and use trig identities.

   (b) If $w \neq 0$ express $\mathrm{Re}(v/w)$ in terms of $\langle\, v,\, w \,\rangle$ and $|w|$. Compare with Proposition I.2.16.

8. Prove Proposition I.2.19, which characterizes straight lines in terms of distance.

9. Prove from the definitions that given any pair of distinct points z_0 and $w_0 \in \mathbb{C}$, there is exactly one line containing both z_0 and w_0.

10. Prove from the definitions that if lines L and L' are parallel, then they are equal or disjoint.

11. Let $z_0 = a + ib$ and $w_0 = c + id$ be two distinct complex numbers and let L be the line through w_1 and w_2.

   (a) Write an equation in the two real variables x and y whose solution set is the line L.

   (b) Write an equation in the two complex variables $z = x + yi$ and $\overline{z} = x - iy$ whose solution set is the line L. Hint: write x and y in terms of z and $\overline{z}$.

I.3 Functions from the plane to the plane

Having different descriptions of the plane is useful since functions can then be defined in different ways, according to our convenience. For example, consider the following rules:

1. $z \mapsto (-1)z$
2. $(x, y) \mapsto (-x, -y)$
3. $(r, \theta) \mapsto (r, \theta + \pi)$
4. $re^{i\theta} \mapsto re^{i(\theta+\pi)}$
5. $$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$
6. Rotate the plane by π radians counterclockwise.
7. Reflect each point through the origin to the opposite point.[1]

These rules all define the same function from the plane to itself. What is different about them is our *description* of the rule in each case. In the first case, we view the plane as $\mathbb{C}$ and express our function using complex multiplication: the point $z \in \mathbb{C}$ is sent to the point $(-1)z$. In the second, we view it as $\mathbb{R}^2$ equipped with rectangular coordinates and express our function explicitly in these coordinates, i.e. we write $(x, y) \mapsto (g(x, y), h(x, y))$ with $g(x, y) = -x$ and $h(x, y) = -y$. In the third, we view it as $\mathbb{R}^2$ equipped with polar coordinates, and express our function explicitly in these coordinates. In the fourth we view the plane as $\mathbb{C}$ again and we write $z = re^{i\theta}$. (Note that $e^{i\pi} = -1$.) In the fifth, we view the plane as a vector space and express our function (which happens to be a linear transformation from $\mathbb{R}^2$ to itself) by giving a matrix for it. In the last two , we give, in words, geometric descriptions of our function.

[1] "Reflect point z through point a" means "map z to that point on the straight line L_{az} which is on the other side of a from z and equally far away. In other words, $z \mapsto a - (z - a) = 2a - z$.

Being able to pass back and forth between these different points of view means that we will be able to adopt whichever one is most convenient. The following Proposition is quite useful in this regard.

Proposition I.3.1 *Let $\alpha = a + ib = se^{i\eta} \in \mathbb{C}$ and $z = x + iy = re^{i\theta} \in \mathbb{C}$ (where $r, s, \theta, \eta \in \mathbb{R}$). Then the following descriptions define the same function from the plane to itself, and this function is a real linear transformation:*

1. $z \mapsto \alpha z$
2. $(x, y) \mapsto (ax - by,\ bx + ay)$
3. *writing vectors as column vectors,*
$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$
4. $re^{i\theta} \mapsto sre^{i(\theta+\eta)}$

Remark: Geometrically this function $z \mapsto se^{i\eta}z$ rotates each point z by the angle $\eta = \arg\ \alpha$ around the origin and multiplies its amplitude $|z|$ by $R = |\alpha|$. This is explained in Proposition I.4.10

Proof: To see that 2. gives the same function as 1., note that

$$\alpha z = (a + ib)(x + iy) = (ax - by) + i(bx + ay)$$

by the definition of complex multiplication. The complex number $(ax-by)+i(bx+ay)$ is identified with the point $(ax - by, bx + ay) \in \mathbb{R}^2$, so that the functions in 1. and 2. are the same.

Computing the matrix product on the right-hand side of 3. above we get

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix}.$$

Thus, under our identification of the plane with the vector space $\mathbb{R}^2$, the functions in 2. and 3. are the same.

Finally these four functions are all the same since 1. gives the same function as 4. if one merely substitutes the exponential form of a and z into 1.

If we let $T(z) = \alpha z$ then T is a linear transformation of $\mathbb{C}$ when $\mathbb{C}$ is viewed as a real vector space. This is because, for all $z_1, z_2 \in \mathbb{C}$ and $\lambda \in \mathbb{R}$,

$$T(z_1 + z_2) = \alpha(z_1 + z_2) = \alpha z_1 + \alpha z_2 = T(z_1) + T(z_2)$$

and

$$T(\lambda z) = \alpha(\lambda z) = \lambda \alpha z = \lambda T(z).$$

(Alternatively, as a function from $\mathbb{R}^2$ to $\mathbb{R}^2$, this is a linear transformation since it is given by matrix multiplication as in 3.) ■

Exercises I.3

1. Let $\alpha = e^{2\pi i/3}$. Find real numbers a, b such that the functions

   $$z \mapsto \alpha z$$

   and

   $$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

   are the same. Give a geometric interpretation of this function (no proof required).

2. Find a complex number α in the form $\alpha = re^{i\eta}$ such that the functions

   $$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

   and

   $$z \mapsto \alpha z$$

   are the same. Give a geometric interpretation of this function (no proof required).

3. Give a formula for $R(z)$ if $R : \mathbb{C} \longrightarrow \mathbb{C}$ is the rotation about the origin which takes $4 + 7i$ to $\sqrt{\frac{13}{2}}\,(-1 + 3i)$.

I.4 Isometries: definition and examples

Having in hand a precise notion of the plane, the distance between points, geometric quantities and objects such as lines and angles, and various ways of defining functions on the plane, we now define and investigate the class of functions that we will use to model symmetries.

Definition I.4.1 (Isometry)
An isometry $f : \mathbb{C} \to \mathbb{C}$ *is a distance-preserving function from the plane to itself. This means that*

$$|f(z) - f(w)| = |z - w| \qquad \textit{for all } z, w \in \mathbb{C}.$$

The set of all isometries is denoted by Isom $\mathbb{C}$.

The first and simplest example of an isometry is the identity map:

Definition I.4.2 (Identity map) *The function* $id_{\mathbb{C}} : \mathbb{C} \to \mathbb{C}$ *given by*

$$id_{\mathbb{C}}(z) = z \qquad \textit{for all } z \in \mathbb{C}$$

is called the identity map *of* $\mathbb{C}$.

If the domain and range are understood, we will often drop the subscript $\mathbb{C}$ and write id instead of $\mathrm{id}_{\mathbb{C}}$.

Lemma I.4.3 *The identity map is an isometry of* $\mathbb{C}$.

Proof: Let $z, w \in \mathbb{C}$ be any pair of points. Then

$$|\mathrm{id}(z) - \mathrm{id}(w)| = |z - w| \quad \text{(by the definition of id.)}$$

Therefore id is an isometry. ∎

If $f : \mathbb{C} \to \mathbb{C}$ is an isometry then we shall prove (I.5.3) that f is a bijection. We prove immediately the first half of this — namely, that f is one-one. (To review the definition of these terms see Definition D.1).

Lemma I.4.4 *If $f : \mathbb{C} \to \mathbb{C}$ is an isometry then f is a one-to-one function.*

Proof: $x_1 \neq x_2 \implies |x_1 - x_2| \neq 0 \implies |f(x_1) - f(x_2)| \neq 0 \implies f(x_1) \neq f(x_2)$
The second implication is due to the fact that f preserves distances. ■

New isometries from old

Before giving the basic examples of isometries, we provide some tools for producing examples from known examples. We prove that the composition of two isometries is an isometry and that the inverse of an isometry is an isometry.

Proposition I.4.5 (Composition) .
The composition of two isometries is an isometry.

Proof: Let $f, g : \mathbb{C} \to \mathbb{C}$ be two isometries. . Let $z, w \in \mathbb{C}$. Then

$$\begin{aligned} |f(g(z)) - f(g(w))| &= |g(z) - g(w)| \quad \text{(since } f \text{ is distance-preserving)} \\ &= |z - w| \quad \text{(since } g \text{ is distance-preserving)} \end{aligned}$$

Therefore $f \circ g$ is an isometry. ■

In the exercises, you will be asked to extend this proposition by using mathematical induction to prove that an arbitrary finite composition of isometries, $f_1 \circ f_2 \circ \cdots \circ f_n$, is an isometry.

Since every isometry $f : \mathbb{C} \to \mathbb{C}$ is a bijection, it has an inverse $f^{-1} : \mathbb{C} \to \mathbb{C}$. (See Appendix D and Proposition D.2.). Is the function f^{-1} also an isometry?

Proposition I.4.6 (Inverses) *The inverse of an isometry is again an isometry.*

Proof: Let z, w be elements of the range $\mathbb{C}$ of f, which is the domain of f^{-1}. Then

$$|f^{-1}(z) - f^{-1}(w)| = |f(f^{-1}(z)) - f(f^{-1}(w))| = |z - w|,$$

where the first equality follows from the fact that f is distance-preserving and the second from the fact that $f \circ f^{-1}(u) = u$ for all $u \in \mathbb{C}$. Hence f^{-1} is an isometry. ■

Examples

We shall introduce, in turn, the following types of isometries.

translations, rotations, reflections, glide reflections.

Translations

Definition I.4.7 *Let $b \in \mathbb{C}$. The function $T_b : \mathbb{C} \to \mathbb{C}$ given by*

$$T_b(z) = z + b$$

is called a translation by b. *The real number $|b|$ is called the* magnitude *of the translation T_b. If $b \neq 0$ the number $b/|b|$ is called the* direction *of the translation T_b.*

When $b = 0$ we get the identity map, viewed as a translation ("the trivial translation") by the vector 0.

Lemma I.4.8 *Translations are isometries.*

Proof: Let z, w be any pair of points in $\mathbb{C}$. Then

$$\begin{aligned} |T_b(z) - T_b(w)| &= |(z+b) - (w+b)| \quad \text{(by the definition of } T_b) \\ &= |z - w|. \end{aligned}$$

Therefore T_b is distance-preserving. ∎

Remark:

- The inverse of the translation T_b is T_{-b} since

$$\begin{aligned} T_{-b} \circ T_b(z) &= T_{-b}(z+b) = (z+b) - b = z \qquad \text{and} \\ T_b \circ T_{-b}(z) &= T_b(z-b) = (z-b) + b = z \qquad \text{for all } z \in Z. \end{aligned}$$

- Translations preserve directions of all lines in the plane. (Exercise 5a.)

Rotations

Definition I.4.9 (Rotation about the origin) *Let $\eta \in \mathbb{R}$. The function $R_\eta : \mathbb{C} \to \mathbb{C}$ given in polar coordinates by*

$$R_\eta(r, \theta) = (r,\ \theta + \eta)$$

is called a rotation about the origin by angle η.

When $\eta = 0$ we get the identity map, viewed as a rotation ("the trivial rotation") by 0 radians.

Are rotations well-defined?

Before giving formulas for rotations about points other than the origin and checking whether rotations are isometries, we must check that R_η is a *well-defined* function.

The point is that each point in the plane has many different labels in terms of polar coordinates, and our *formula* for rotation is given *in terms of these labels.* If R_η is to be a function, it must give only one result for each point. Thus we must check that our formula for R_η gives the same answer when two different polar coordinates (r_1, θ_1) and (r_2, θ_2) represent the same point of $\mathbb{C}$. See Appendix E.

To verify that rotations about the origin are well-defined, suppose first that $r = 0$. Then for *any* $\theta \in \mathbb{R}$, $(r,\ \theta) = (0,\ \theta))$ represents the origin and then

$$R_\eta(0,\ \theta) = (0,\ \theta + \eta) = \text{the origin.}$$

Therefore the origin goes to the origin, no matter which polar coordinates are used for the origin. So suppose $r \neq 0$. Then the polar coordinates $(r,\ \theta + 2n\pi),\ n \in \mathbb{Z}$ all describe the same point in the plane, and no other polar coordinates describe that point. But for any $n \in \mathbb{Z}$,

$$R_\eta(r,\ \theta + 2n\pi) = (r,\ \theta + \eta + 2n\pi).$$

However for all $n \in \mathbb{Z}$, the points in the plane with polar coordinates $(r,\ \theta + \eta + 2n\pi)$ are the same, so our function R_η is indeed well-defined.

We can give two other descriptions of R_η—one in terms of complex multiplication, the other using matrices.

Proposition I.4.10 *Let $a = e^{i\eta} = \cos\eta + i\,\sin\eta$.*
Then the following are all descriptions of the same function $R_\eta : \mathbb{C} \longrightarrow \mathbb{C}$:

1. *rotation of the plane by angle η about the origin,*
2. *$(r,\ \theta) \mapsto (r,\ \theta + \eta)$ (in polar coordinates),*
3. *$z \mapsto az = e^{i\eta}z$, for all $z \in \mathbb{C}$, and*
4. *writing vectors as column vectors,*

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \cos\eta & -\sin\eta \\ \sin\eta & \cos\eta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Proof: : 1. and 2. represent the same function by definition. To see that 3. agrees with these, write z as $z = re^{i\theta}$. Then 3. becomes

$$\begin{aligned} re^{i\theta} \mapsto e^{i\eta}re^{i\theta} &= r(\cos\eta + i\sin\eta)(cos\theta + i\sin\theta)) \\ &= r\Big((\cos\eta\,\cos\theta - \sin\eta\,\sin\theta) + i(\cos\eta\,\sin\,\theta + \sin\eta\,\cos\theta)\Big) \\ &= r\Big(\cos(\eta+\theta) + i\sin(\eta+\theta)\Big) \\ &= re^{i(\theta+\eta)} \end{aligned}$$

Thus az has polar coordinates $(r,\ \theta + \eta)$ and we see that 3. agrees with 1. and 2.

Finally 4. agrees with the previous three formulas for R_η by Proposition I.3.1. ∎

Corollary I.4.11

1. *The rotation R_η is a linear transformation of $\mathbb{R}^2$.*
2. *R_η changes the direction of every line in the plane by η radians.*

Proof: 1. follows from the fact that R_η has a matrix representation. (See Appendix F.) 2. is left as Exercise 5b. ∎

Now that we have the convenient description of rotation via multiplication by a complex number of absolute value one, we can prove

Lemma I.4.12 *Rotations about the origin are isometries.*

Proof: Let the rotation $R_\eta : \mathbb{C} \to \mathbb{C}$ be given by $R_\eta(z) = az,\ a = e^{i\eta}$.
Then for any pair $z, w \in \mathbb{C}$,

$$\begin{aligned} |R_\eta(z) - R_\eta(w)| &= |az - aw| \\ &= |a(z-w)| \\ &= |a| \cdot |z-w| \\ &= 1 \cdot |z-w| \\ &= |z-w|. \end{aligned}$$

∎

How should we define a rotation about a point which is not the origin?

On the one hand, without a choice of coordinates, it seems that rotation about one point should not be essentially different from rotation by the same angle about some other point, since without coordinates no point is distinguished. On the other hand, the imposition of a coordinate system means that one point, namely the origin, is distinguished. To treat another point z_0 as the origin, we write each point z as $z = z_0 + (z - z_0)$ and rotate the difference vector $z - z_0$ about the point z_0. We accomplish this as follows.

Definition I.4.13 (Rotation about arbitrary point) *Let $\eta \in \mathbb{R}$ and $z_0 \in \mathbb{C}$. The function $R_{z_0,\eta} : \mathbb{C} \to \mathbb{C}$ given by*

$$\begin{aligned} R_{z_0,\eta}(z) &= T_{z_0} \circ R_\eta \circ T_{-z_0}(z) \\ &= z_0 + e^{i\eta}(z - z_0) \end{aligned}$$

is called the rotation about the point z_0 by angle η.

Fig. 4. Rotation about an arbitrary point z_0. Watch $z - z_0$.

We will sometimes refer to $R_{z_0,\eta}$ simply as a *rotation*.

In words, this definition says the following. First translate the point z_0 to the origin by T_{-z_0}, then rotate about the origin by angle η, and finally translate the origin back to z_0 by T_{z_0}. See Figure 4.

Since T_{z_0}, R_η, and T_{-z_0} are well-defined functions, so is their composition. We wish to show that $R_{z_0,\eta}$ is an isometry. We could do this "by hand", but it was for this purpose that we gave the Propositions above explaining how new isometries arise from old. At this point, translations and rotations about the origin are known ("old") isometries. From Propositions I.4.5 and I.4.6 we conclude

Proposition I.4.14 *Rotations are isometries.* ∎

Example I.4.15 (Rotating the line $y = mx + b$ to be horizontal)

Suppose that a line is given in the common form $y = mx + b, \quad m, b \in \mathbb{R}$. (Given two points on a non-vertical line, or a point and

a direction vector, this form is easily achieved.) Let us find the formula $f(z) = az$ for a rotation about the origin which will rotate this line to a horizontal line.

Since isometries take parallel lines to parallel lines (Exercise 4), it suffices to give a rotation which rotates the parallel line L given by $y = mx$ onto the x-axis. Suppose that this non-vertical line L makes an angle of η radians with the positive x-axis, where $-\pi/2 < \eta < \pi/2$. We give the formula for a rotation by $-\eta$ radians. (A rotation by $-\eta + \pi$ radians would also suffice.)

Since L has slope m, we see that $\tan \eta = m$. Therefore

$$\begin{aligned} \cos\eta &= \frac{1}{\sqrt{m^2+1}}, & \sin\eta &= \frac{m}{\sqrt{m^2+1}}, \\ e^{i\eta} &= \frac{1+im}{\sqrt{m^2+1}}, & e^{-i\eta} &= \frac{1-im}{\sqrt{m^2+1}}. \end{aligned}$$

Therefore the rotation we seek is given by

$$f(z) = e^{-i\eta}z = \frac{1-im}{\sqrt{m^2+1}}\, z.$$

Since $e^{i\pi} = -1$, the other rotation, by $-\eta + \pi$ radians, is given by

$$g(z) = \frac{-1+im}{\sqrt{m^2+1}}\, z.$$

Reflections

We may describe a reflection geometrically as follows. (See Figure 5.) Let L be a line in the plane and let P be a point in the plane. To find the image $M_L(P)$ of P under reflection in L, construct the line $L^\perp$ through P perpendicular to L. It intersects L in exactly one point, say C. Then $M_L(P)$ is defined to be the point Q on the line $L^\perp$ such that C is the midpoint of $\overline{PQ}$.

The simplest example of a reflection is reflection across the x-axis, which we will call M_0 in this section. This clearly takes the point $(x,\ y)$ to $(x,\ -y)$. In complex notation M_0 has the formula

$$M_0(z) = \overline{z}.$$

Fig. 5. Geometric description of a reflection.

Fig. 6. The choices of angle between a line and the positive x-axis.

The function M_0 is an isometry (Exercise 1).

In general, we wish to give our formal definition of reflection in the line L by a *formula* for M_L. Such a formula must use crucial data about the line, and the data we use will consist of a point z_0 on the line and the angle η which the line makes with the positive x-axis.

Of course there are many choices of $z_0 \in L$. Also, the angle η that L makes with the positive real axis is not unique: any two measurements differ by a multiple of π (see Figure 6.) We will have to check that our formula for M_L does not give a different function if we make different choices of z_0 or η. That is, we will have to show that our formula is well-defined.

Fig. 7. Reflection in a line through the origin.

We begin by defining reflection in a line L passing through the origin, $z_0 = 0$ (just as our discussion of rotations began with rotations about the origin).

Definition I.4.16 (Reflection in line through origin) *Let L be a line through the origin. Suppose that L makes an angle η with the positive real axis. The function $M_L : \mathbb{C} \to \mathbb{C}$ given in polar coordinates by*

$$M_L(r, \theta) = (r, 2\eta - \theta)$$

is called reflection in the line L.

See Figure 7 to see why the formula looks the way it does.

We now show that M_L is a well-defined function. If we chose to measure the angle between L and the x-axis by a different angle η', and if different polar coordinates (r, θ') were chosen to name the same point of the domain, then

$$(r,\ 2\eta' - \theta') = \Big(r,\ 2(\eta + k\pi) - (\theta + 2n\pi)\Big) = (r,\ 2\eta - \theta + 2\pi(k + n)).$$

Therefore $(r,\ 2\eta - \theta)$ and $(r,\ 2\eta' - \theta')$ represent the same point of the plane.

As we did with rotations about the origin (Proposition I.4.10), we now give complex and linear algebra formulas for reflection across a line through the origin. Also we show in 5. below that M_L is "the same as M_0 if we turn our head by an angle of η radians": it can be obtained by rotating L onto the x-axis, applying M_0, and then rotating back.

Proposition I.4.17 (Descriptions of reflections in lines through origin) *Suppose that the line L through the origin makes an angle η with the x-axis. Then the following are all descriptions of the same function $M_L : \mathbb{C} \longrightarrow \mathbb{C}$:*

1. *reflection across the line L*
2. *$(r, \theta) \mapsto (r, 2\eta - \theta)$ (in polar coordinates)*
3. *$z \mapsto e^{i2\eta}\,\overline{z} = R_{2\eta}(\overline{z}))$*
4. *writing vectors as column vectors,*

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \cos 2\eta & \sin 2\eta \\ \sin 2\eta & -\cos 2\eta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

5. *the composition $R_\eta \circ M_0 \circ R_{-\eta}$.*

Proof: 1. and 2. represent the same function by definition. To see that 3. agrees with these in complex coordinates, set $z = re^{i\theta}$. We have

$$\begin{aligned} M_L(z) &= re^{i(2\eta-\theta)} \qquad [\text{complex form of } (r,\ 2\eta - \theta)] \\ &= e^{i2\eta} re^{-i\theta} \\ &= e^{i2\eta}\overline{z} \\ &= R_{2\eta}(\overline{z}). \end{aligned}$$

We get the formula 4. by noting that $z = x + iy$, and $\overline{z} = x - iy \in \mathbb{C}$ correspond in $\mathbb{R}^2$ to $\begin{pmatrix} x \\ y \end{pmatrix}$ and to

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

Thus, using the matrix formula for a rotation about the origin (Proposition I.4.10) and the formula $M_L(z) = R_{2\eta}(\overline{z})$ achieved in 3., the function $M_L : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ becomes

$$\begin{aligned} M_L\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \cos 2\eta & -\sin 2\eta \\ \sin 2\eta & \cos 2\eta \end{pmatrix}\begin{pmatrix} x \\ -y \end{pmatrix} \\ &= \begin{pmatrix} \cos 2\eta & -\sin 2\eta \\ \sin 2\eta & \cos 2\eta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} \cos 2\eta & \sin 2\eta \\ \sin 2\eta & -\cos 2\eta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \end{aligned}$$

Finally, we see from 3. that $M_L = R_\eta \circ M_0 \circ R_{-\eta}$ because

$$R_\eta\, M_0\, R_{-\eta}(z) = R_\eta\Big(\overline{e^{-i\eta}z}\Big) = e^{i\eta}\Big(e^{i\eta}\overline{z}\Big) = e^{i2\eta}\overline{z} = M_L(z). \qquad \blacksquare$$

Example I.4.18 (Reflection in the line $y = mx$) *If a line L through the origin is given by the equation $y = mx$ ($m \in \mathbb{R}$), then the complex formula for reflection in this line is given by*

$$M_L(z) = \frac{1 - m^2 + 2im}{m^2 + 1}\ \overline{z}.$$

Proof: This follows by the same reasoning as in Example I.4.15. Let η be the angle this line makes with the x-axis, with $-\pi/2 < \eta < \pi/2$. Then

$$e^{i\eta} = \frac{1 + im}{\sqrt{m^2 + 1}} \qquad \text{and} \qquad e^{i2\eta} = \Big(e^{i\eta}\Big)^2 = \frac{1 - m^2 + 2im}{m^2 + 1}.$$

Thus the formula claimed is just a way of rewriting the fact that $M_L(z) = e^{i2\eta}\overline{z}$.

$\blacksquare$

How should we define reflection in an arbitrary line?

Again we wish to give an algebraic definition which captures the geometric description of a reflection given in Figure 5. We will give a formula using our parametric

description of a line, and then show that the formula is independent of our choice of description, i.e. that our formula is well-defined. For the moment, we will assume this.

Let $L = \{z \mid z = z_0 + tz_1, t \in \mathbb{R}\}$ where $z_1 \neq 0$. Suppose $z_1 = re^{i\eta}$. Thus $\eta = \arg(z_1)$ is an angle which L makes with the positive x-axis. Let L_η be the (parallel) line through the origin making the same angle with the x-axis, and let M_η be reflection in the line L_η. Let T_{-z_0} and T_{z_0} denote translation by $-z_0$ and z_0 respecively.

Definition I.4.19 (Reflection in arbitrary line) *The function* $M_L : \mathbb{C} \to \mathbb{C}$ *given by*

$$M_L(z) = T_{z_0} \circ M_\eta \circ T_{-z_0}(z)$$

is called reflection *in* L. *The line* L *is called the* axis *or* mirror[1] *of the reflection.*

Direct calculation gives

$$M_L(z) = e^{i2\eta}\overline{z} + (z_0 - e^{i2\eta}\overline{z_0}).$$

To see that this definition matches what we intuitively want, notice that T_{-z_0} carries L to L_η and T_{z_0} carries L_η to L. So what's going on is, first we move L to L_η by a *translation*, which does not change the direction of lines. (In particular, lines perpendicular to L go onto lines perpendicular to L_η.) Then we do reflection in L_η, and then we move L_η back to L by undoing this translation. We actually did something similar in our definition of a rotation about an arbitrary point and again in 5. of Proposition **??**. This process is called *conjugation*. We shall have more to say about it in Section III.4.

Proposition I.4.20 *Reflections are isometries.*

Proof: We combine the definition of M_L with the expression for M_η which comes from 5. in Proposition **??** — where $M_\eta = M_{L_\eta}$ here plays the role of "M_L" in that Proposition. This gives

$$M_L = T_{z_0}\, M_\eta\, T_{-z_0} = T_{z_0}\, R_\eta\, M_0\, R_{-\eta}\, T_{-z_0}.$$

[1] Reflections will usually be denoted by "M" because of the mirror involved.

Thus M_L is a composition of five isometries. Hence it is an isometry. ∎

We now verify that our formula for reflection gives a well-defined function, i.e. that the function which we call M_L depends only on the line L and not on the data z_0, z_1 which we used to describe L. If the same line is described using a different point w_0 and a different vector w_1, we wish to know that the formula given defines the same function M_L.

We got the angle $\eta = \arg(z_1)$ from the direction vector z_1; the angle $\eta_w = \arg(w_1)$ differs from this by a multiple of π, since they're both parallel to the line L. So the functions M_η and M_{η_w} which reflect across the line through the origin parallel to L are the same. (This was pointed out following Definition I.4.16.) Thus it suffices to prove that $T_{z_0} M_\eta T_{-z_0} = T_{w_0} M_\eta T_{-w_0}$.

Notice that, since both z_0 and w_0 are points on L, $w_0 - z_0$ is a vector in the direction of L. So there is a real number t for which $w_0 - z_0 = te^{i\eta}$. Writing the formula for M_L using w_0 gives

$$\begin{aligned} T_{w_0} \circ M_\eta \circ T_{-w_0}(z) &= e^{i2\eta}\overline{z} + (w_0 - e^{i2\eta}\overline{w_0}) \\ &= e^{i2\eta}\overline{z} + (z_0 + te^{i\eta}) - e^{i2\eta}(\overline{z_0 + te^{i\eta}}) \\ &= [e^{i2\eta}\overline{z} + z_0 - e^{i2\eta}\overline{z_0}] + \underbrace{[(te^{i\eta} - e^{i2\eta}te^{-i\eta}]}_{0} \\ &= T_{z_0} M_\eta T_{-z_0}. \end{aligned}$$

Therefore our definition is indeed independent of the choice of point z_0 on the line.

Glide reflections

Definition I.4.21 (Glide-reflections) *A* glide reflection *g is an isometry which can be written in the form*

$$g = T_c \circ M_L$$

where T_c is translation by a non-zero vector c which is parallel to the line L.

Note: If $z_0 \in L$ and L makes an angle η with the x-axis, then $c = te^{i\eta}$ for some $t \neq 0$. Therefore

$$g(z) = e^{i2\eta}\overline{z} + (z_0 - e^{i2\eta}\overline{z_0}) + te^{i\eta}.$$

Fig. 8. A glide-reflection g and its second iterate $g \circ g$

Proposition I.4.22 (Properties of glide-reflections) *If $g = T_c M_L$ is a glide reflection, as in the definition, then*

1. $g = T_c M_L = M_L T_c$.
2. $g \circ g = T_{2c}$.

Proof: The proof is most clearly given by Figure 8. It can also be proved by direct computation using the formula for $g(z)$ above. We leave this computation to the reader (Exercise 12). ∎

Exercises I.4

1. Show that the function $z \mapsto \overline{z}$ is an isometry of $\mathbb{C}$.
   Note: We used this fact in our proof that all reflections are isometries!
2. (a) Suppose that $f(z) = az + b$ and $g(z) = a\overline{z} + b$ for all $z \in \mathbb{C}$, where $a, b \in \mathbb{C}$ and $|a| = 1$. Prove that f and g are isometries.
   (b) Suppose that $f(z) = az + b$ or $f(z) = a\overline{z} + b$.
   Prove that: if f is an isometry then $|a| = 1$.
3. Let $f : X \to Y$ and $g : Y \to Z$ be functions. Prove:
   (a) If f and g are one-to-one, then so is $g \circ f$.
   (b) If f and g are onto, then so is $g \circ f$.
   (c) If f and g are bijections, then so is $g \circ f$.

(d) If f is a bijection, then so is f^{-1}.

4. Prove the following statements:

   (a) Isometries take lines onto lines.
   Hint: Isometries preserve distances. Use Proposition I.2.19.

   (b) Isometries take disjoint sets to disjoint sets.

   (c) Isometries take parallel lines to parallel lines.

5. (a) If T_b is translation by b and L is any line in the plane, then the direction of L equals the direction of $T_b(L)$.

   (b) If R_η is rotation about the origin by η radians, then every line in the plane changes its direction by η radians.

   (c) If $R_{z_0,\eta}$ is rotation about the point z_0 by η radians, then every line in the plane changes its direction by η radians.

6. Prove that the composition of two arbitrary translations is a translation.

7. Prove that the composition of two arbitrary rotations (perhaps about different centers) is either a rotation or a translation.
   Exactly when will $R_{z_0,\eta} \circ R_{z_1,\theta}$ be a translation?

8. Use the formula $M_L(z) = e^{i2\eta}(\overline{z - z_0}) + z_0$ to prove that, for any reflection in a line L,
$$M_L \circ M_L = id.$$
   Explain why this is obvious from the original synthetic description given of a reflection.

9. (a) Suppose that a is the perpendicular projection of the point $z \in \mathbb{C}$ onto the line L. Show that $M_L(z) = 2a - z$.
   Hint: Let z and a be the complex numbers labelling p and C in Figure 5. Draw vectors in this picture.

   (b) Find a formula for a.

10. Give formulas in the form $z \mapsto az + b$ or $z \mapsto a\overline{z} + b$ $(a, b \in \mathbb{C})$ for the following isometries of $\mathbb{C}$:

    (a) Reflection in the line L which is the solution set of the equation $y - x - 1 = 0$.

(b) Rotation counterclockwise by $\pi/4$ radians about the point $1+i$.

11. Extend Example I.4.18 to give a complex formula $f(z) = \cdots$ for reflection across the line $y = mx + b$ $(m, b \in \mathbb{R})$.

12. (a) Prove that glide-reflections are isometries.

    (b) Show that $z \mapsto \overline{z} + 1$ and $z \mapsto -\overline{z} + i$ are glide-reflections, and identify T_c and M_L in each case.

    (c) Prove that if $g = T_c \circ M_L$ is a glide-reflection, then $g = M_L \circ T_c$ by using the complex formula for $g(z)$.

    (d) Use (c) and the fact that $M_L \circ M_L = id$ to prove that $g \circ g = T_{2c}$

13. (a) Suppose that g and h are isometries of $\mathbb{C}$ and that $f = hgh^{-1}$. Prove that f is an isometry of C.

    We say that f and g are *conjugate*[2] isometries and that *f is gotten from g by conjugation.*

    (b) List all of the isometries in this section which have been defined by conjugation of other isometries.

    Note that conjugate isometries are very similar in nature. We return to this notion geometrically in Section III.2 and algebraically in Section IV.3.

14. **Induction.** Suppose we want to prove that a certain sequence of mathematical statements P_n, $n = 1, 2, 3, ...$ is true for all n. A *proof by induction on n* consists of establishing the following steps: 1. Prove, by hand, that P_1 is true. 2. For $n > 1$, *assume* that P_k is true for all $k = 1, 2, ...n-1$. Using this assumption, prove that P_n is true.

    (a) Show by induction that for every integer $n \geq 1$,

$$\sum_{k=1}^{k=n} k^2 = n(n+1)(2n+1)/6 \ .$$

    (b) Prove that if $f_1, ..., f_n$ are isometries, then so is $f_1 \circ f_2 \circ ... \circ f_n$.

[2]This use of the word is unrelated to conjugate complex numbers, $a + ib$ and $a - ib$.

I.5 The basic factorization of an isometry

We have given translations, rotations, reflections and glide reflections as basic examples of isometries. In this section we first prove that an isometry is determined by its values on any three non-collinear points. This then allows us to prove that every isometry factors as a product of some of our basic examples. Indeed (I.5.2) *every isometry is the composition of a translation (possibly the identity) and/or a rotation about the origin and/or a reflection across the x-axis.* In the next section we will go further and prove that every isometry is actually equal to one of our basic examples — translation, rotation, reflection or glide reflection.

The main ingredient in our analysis of an isometry is the following geometric proposition.

Proposition I.5.1 *Let $f : \mathbb{C} \to \mathbb{C}$ be an isometry which fixes three non-collinear points c_1, c_2, c_3. Then $f = id$.*

Proof: Let $z \in \mathbb{C} - \{c_1, c_2, c_3\}$, and set $r_i = |z - c_i|, \quad i = 1, 2, 3$. Let C_i denote the circle with center c_i and radius r_i. Then $z \in C_1 \cap C_2 \cap C_3$. Since $f : \mathbb{C} \to \mathbb{C}$ is distance-preserving, $f(C_i) = C_i$, and $f(z) \in C_i$ for each i. Hence we also have $f(z) \in C_1 \cap C_2 \cap C_3$. Thus the proof is complete once we verify that $C_1 \cap C_2 \cap C_3$ contains exactly one point.

Assume, on the contrary, that it contains two points, say z and another point w distinct from z. Then $\{z, w\}$ is contained in each of the intersections

$$C_1 \cap C_2, \quad C_2 \cap C_3, \quad C_1 \cap C_3.$$

By Exercise 1, each of these sets — being the intersection of two distinct circles — can contain at most two points. Hence our assumption implies that

$$\{z, w\} = C_1 \cap C_2 \cap C_3 = C_1 \cap C_2 = C_2 \cap C_3 = C_1 \cap C_3.$$

Let L be the line through z and w. Now let L_{12}, L_{23}, L_{13} be the lines through c_1 and c_2, through c_2 and c_3, and through c_1 and c_3, respectively. By Exercise 1, L is perpendicular to each of the lines L_{12}, L_{23}, L_{13}. In particular, the lines L_{12} and L_{23} are both perpendicular to L. Since they both pass through the same point, c_2, as well, they must coincide. So c_1, c_2, c_3 are collinear, contradicting our hypothesis. ∎

Theorem I.5.2 *If* $f : \mathbb{C} \longrightarrow \mathbb{C}$ *is an isometry then there exists* $b \in \mathbb{C}$, $\theta \in \mathbb{R}$, *and* $\epsilon \in \{0, 1\}$ *such that*

$$f = T_b \, R_\theta \, M^\epsilon.$$

where T_b *is translation by* b, R_θ *is rotation by* θ *radians about the origin, and* M *is reflection across the* x*-axis.*

Proof: The plan is to compose f with known isometries which have the net effect of taking $f(0)$, $f(1)$, and $f(i)$ back to 0, 1, and i.

If $b = f(0)$ the translation T_{-b} satisfies $T_{-b}f(0) = f(0) - f(0) = 0$.

Since $T_{-b}f$ is distance-preserving, it takes the circle with center 0 and radius 1 into itself. Therefore it takes the number 1 to $T_{-b}f(1) = e^{i\theta}$ for some $\theta \in \mathbb{R}$. The rotation $R_{-\theta}$ about the origin by angle $-\theta$ then satisfies

$$\begin{aligned} R_{-\theta}T_{-b}f(0) &= 0. \\ R_{-\theta}T_{-b}f(1) &= 1. \end{aligned}$$

Since $R_{-\theta}T_{-b}f$ is distance preserving, it takes the circle with center 0 and radius 1 to itself. It takes the circle of with center 1 and radius $\sqrt{2}$ to itself too. *Draw a picture!* Thus, it takes the intersection of these two circles, the set $\{i, -i\}$, to itself. In particular, it takes i to either itself, or to $-i$.

Let us temporarily denote $g = R_{-\theta}T_{-b}$. As a composition of isometries, g is an isometry, and we have just seen that $gf(0) = 0$, $gf(1) = 1$, and either $gf(i) = i$ or $gf(i) = -i$.

If $gf(i) = i$, then gf fixes the three non-collinear points 0, 1, i. Thus $gf = \mathrm{id}$ by I.5.1. If $gf(i) = -i$, then, compose with the reflection $M = M_0$ across the x-axis. We see that Mgf fixes the three non-collinear points 0, 1, i. Thus $Mgf = \mathrm{id}$. In either case we may write $M^\epsilon gf = \mathrm{id}$, $\epsilon \in \{0, 1\}$.

Since translations, rotations and reflections are known to be bijections, $M^\epsilon g$ is a bijection and we may use its inverse, along with the fact (D.4) that associativity holds for the composition of any three functions:

$$M^\epsilon g f = \mathrm{id} \implies (M^\epsilon g)^{-1}(M^\epsilon g f) = (M^\epsilon g)^{-1} \circ \mathrm{id} \implies (M^\epsilon g)^{-1}(M^\epsilon g) f = (M^\epsilon g)^{-1}$$
$$\implies f = (M^\epsilon g)^{-1} = g^{-1} M^{-\epsilon} = g^{-1} M^\epsilon.$$

Hence $f = T_b \circ R_\theta \circ M^\epsilon$ is a composition of isometries, and is therefore an isometry. ∎

We now prove the basic fact stated earlier:

Corollary I.5.3 *Suppose that $f : \mathbb{C} \longrightarrow \mathbb{C}$ is an isometry. Then f is a bijection of the plane onto itself.*

Proof: *We may write f as a composition $f = T_b R_\theta M^\epsilon$. We know that translations, rotations and reflections are bijections, as is the identity map. (Note that any of these maps might be the identity map in case $b = 0, \theta = 0$ or $\epsilon = 0$.) As a composition of bijections, f is a bijection.* ∎

The preceding Corollary tells us that each isometry has an inverse and we know (I.4.5, I.4.6) (i) the composition of isometries is an isometry and (ii) the inverse of an isometry is an isometry. This allows us to prove the following result:

Corollary I.5.4 *Suppose that f and g are isometries of $\mathbb{C}$ and there exist three non-collinear points a, b, c, such that $f(a) = g(a)$, $f(b) = g(b)$ and $f(c) = g(c)$. Then $f = g$.*

Another way to say this is that *"three non-collinear points determine an isometry."* This sort of expression is very common in mathematical writing.

Proof: The isometry $f^{-1}g$ fixes the the points a, b, c. Therefore

$$\begin{aligned} f^{-1} \circ g &= \mathrm{id} \\ f \circ (f^{-1} \circ g) &= f \circ \mathrm{id} \\ (f \circ f^{-1}) \circ g &= f \\ \mathrm{id} \circ g &= f \\ g &= f \end{aligned}$$

■

Exercises I.5

1. Show algebraically that two distinct circles C_1, C_2 can intersect in at most two points, and if they intersect in exactly two points z, w, then z and w lie on the line perpendicular to the line through the centers of C_1 and C_2.

   Hint: Do this first when C_1 is centered at the origin $(0,0)$ and C_2 is centered at $(a,0)$ on the x-axis. Thus these have equations $x^2 + y^2 = r_1^2$ and $(x-a)^2 + y^2 = r_2^2$. Solve for the points (x,y) which satisfy both equations.

2. (a) Suppose that P is a subset of the plane which contains three non-collinear points. If $f : P \longrightarrow P$ is a distance-preserving function of P to itself, prove that there is an isometry of the plane, $F : \mathbb{C} \longrightarrow \mathbb{C}$, such that $F \,|\, P = f$.
   Hint: Apply a variation of the proof of Theorem I.5.2.

   (b) Prove that the isometry F, constructed in the preceding part, is unique.

   (c) Suppose that P has three non-collinear points. Prove that

   $$\{f : P \longrightarrow P \mid f \text{ is distance} - \text{preserving}\} = \{f : \mathbb{C} \longrightarrow \mathbb{C} \mid f \text{ is an isometry}\}$$

   (This set will be defined (I.7.1 as Sym P.)

   (d) Are these two sets of functions necessarily equal if P does not contain three non-collinear points?

   What if P does not contain two non-collinear points?

I.6 Classification of isometries

We have given examples of four types of isometries:

translations, rotations, reflections, glide reflections.

In this section, we use I.5.2 to show first that, apart from the identity, there are no other isometries (Theorem I.6.1). In each of these four cases we show that an isometry f is given by a specific kind of formula of the form $f(z) = az + b$ or $f(z) = a\overline{z} + b$.

At this point, there is some logical housecleaning to do. Unlikely as it sounds, it is nonetheless conceivable that an isometry f might be expressible using more than one formula of the form $az + b$ or $a\overline{z} + b$. We rule this out in Theorem I.6.2.

Theorems I.6.1 and I.6.2 together will be referred to as *the classification of isometries* or *the Classification Theorem.*

Theorem I.6.1 *Let $f : \mathbb{C} \to \mathbb{C}$ be an isometry other than the identity. Then one of the following possibilities holds:*

i) *f is a* **translation** *given by*

$$\boxed{f(z) = z + b} \qquad \text{for some } b \neq 0 \in \mathbb{C}.$$

ii) *f is a* **rotation** *given by*

$$\boxed{f(z) = az + b} \qquad \text{for some } a,\ b \in \mathbb{C} \text{ with } |a| = 1,\ \ a = e^{i\theta} \neq 1.$$

It is a rotation by angle θ about the point $z_0 = b/(1 - a)$.

iii) *f is a* **reflection** *given by*

$$\boxed{f(z) = a\overline{z} + b} \qquad \text{for some } a,\ b \in \mathbb{C} \text{ with } |a| = 1,$$

where $(f \circ f)(0) = a\bar{b} + b = 0$. *The axis of reflection is the line* L *through* $b/2$ *which makes angle* $\theta/2$ *with the* x*-axis, where* $a = e^{i\theta}$. *If* $b \neq 0$, *this line has direction determined by* $\pm ib$.

iv) f *is a* **glide reflection** *given by*

$$\boxed{f(z) = a\overline{z} + b} \qquad \text{for some } a,\ b \in \mathbb{C} \text{ with } |a| = 1,\ \ a = e^{i\theta} \neq 1$$

where $(f \circ f)(0) = a\bar{b} + b \neq 0$.

f *is the composition* $f = T_c \circ M_L$ *of translation by* $c = (f \circ f(0))/2 = (a\bar{b}+b)/2$ *and the reflection* $M_L(z) = a\overline{z} + (b - a\bar{b})/2$. *The line* L *is the line through* $(b - a\bar{b})/4$ *and makes angle* $\theta/2$ *with the* x*-axis.* c *is parallel to the line* L.

Proof: We know from Theorem I.5.2 that any isometry f can be written as

$$f = T_b R_\theta M^\epsilon, \qquad b \in \mathbb{C}, \quad \theta \in \mathbb{R}, \quad \epsilon = 0 \text{ or } 1.$$

We shall denote $a = e^{i\theta} = R_\theta(1)$.

Then all isometries will fall into one or more of the following classes (the "more than one" possibility will be ruled out by the next theorem):

i) $\epsilon = 0, \quad a = 1,$ **ii)** $\epsilon = 0, \quad a \neq 1,$

iii) $\epsilon \neq 0, \quad (f \circ f)(0) = 0,$ **iv)** $\epsilon \neq 0, \quad (f \circ f)(0) \neq 0.$

We consider these cases in order.

i) We have $f = T_b\, R_0\, M^0 = T_b$. Thus f is translation by b and has formula

$$f(z) = z + b.$$

ii) Since $\epsilon = 0$ and $a = e^{i\theta}$, we have $f = T_b\, R_\theta$ so that

$$f(z) = e^{i\theta} z + b = az + b \qquad \text{for all } z \in \mathbb{C}.$$

[To see now that f is a rotation, we reason backwards: If f were a rotation, its center z_0 would be the unique point with $f(z_0) = z_0$. Solving the equation $z = az + b$, we

see that z_0 would equal $b/(1-a)$. Motivated by the formula for $R_{z_0,\theta}$ in Section I.4, we conjecture that the angle of rotation should be θ. So we make the following claim.][3]

Claim. *Let* $z_0 = b/(1-a)$, *where* $a = e^{i\theta}$. *Then*

$$T_{z_0} R_\theta T_{-z_0}(z) = az + b = f(z) \quad \textit{for all } z \in \mathbb{C}.$$

The claim is proved by direct computation (which is left to the reader). From the definition of rotation (I.4.10), we see that $f = R_{z_0,\theta}$ is indeed a rotation about the point z_0 by angle θ.

Note for (iii) and (iv): Since we are given that $\epsilon = 1$ we have

$$f(z) = T_b\, R_\theta\, M(z) = a\bar{z} + b.$$

Therefore: $\qquad f(0) = b, \qquad f(f(0)) = a\bar{b} + b.$

(iii): The hypothesis of (iii) gives us that $e^{i\theta}\bar{b} + b = a\bar{b} + b = 0$. So if $b \neq 0$, we have

$$\left(e^{i\theta/2}\right)^2 = e^{i\theta} = -b/\bar{b} = -b^2/b\bar{b} = -(b/|b|)^2 = \left(ib/|b|\right)^2.$$

Thus $e^{i\theta/2} = \pm ib/|b|$ when $b \neq 0$, and any line with direction $\pm ib$ makes angle $\theta/2$ with the x-axis.

[Some motivation: By hypothesis, $(f \circ f)(0) = 0$. If f is a reflection, it satisfies $f \circ f = \mathrm{id}$, so that it is conceivable that f is a reflection . If f is a reflection and $b \neq 0$, then the midpoint of the line segment from 0 to $f(0)$ (the point $b/2$) would be a fixed point and the axis of reflection would be perpendicular to this line segment. It would have direction $\pm ib$ and, by the note above, make an angle $\theta/2$ with the x-axis. Whether or not $b = 0$, we are now emboldened to make the following claim.]

Claim: *Let* $z_0 = b/2$. *Let* $\eta = \theta/2$. *Then*

$$T_{z_0} R_\eta\, M\, R_{-\eta} T_{-z_0}(z) = a\overline{z} + b = f(z) \quad \textit{for all } z \in \mathbb{C}.$$

[3]Usually this motivational paragraph would be kept in an author's head, and the reader would be given a clear, elegant, unmotivated claim. When confronted with such, it's the reader's job to figure out where it's coming from — not to just verify the steps the author gives.

This claim is true because

$$\begin{aligned} T_{z_0} R_\eta \, M \, R_{-\eta} T_{-z_0}(z) &= e^{i\eta}\left(\overline{e^{-i\eta}(z - z_0)}\right) + z_0 \\ &= e^{i\eta}\left(e^{i\eta}(\overline{z} - \overline{z_0})\right) + z_0 \\ &= e^{i2\eta}\overline{z} - e^{i2\eta}\bar{b}/2 + b/2 \\ &= a\overline{z} - (a\bar{b} + b)/2 + b \\ &= a\overline{z} + b \\ &= f(z). \end{aligned}$$

By Proposition I.4.17 and the definition (I.4.19) of reflection, we see that f is reflection in the line L which contains the point $b/2$ and makes the angle $\theta/2$ with the x-axis, completing the proof of (iii).

(iv) Again we have $f(z) = a\overline{z} + b$, and now we hypothesize that $f(f(0)) \neq 0$. Clearly f cannot be a reflection, since $f \circ f \neq \mathrm{id}$.

[Motivation: If f is a glide reflection, then we can write $f = T_c \circ M_L$ where c is parallel to the line L. Then $f(f(z)) = z + 2c$ for all $z \in \mathbb{C}$. In particular, $0 + 2c = f(f(0)) = a\bar{b} + b$. This tells us what c must be, and hence what T_c must be. Then M_L would have to equal $T_c^{-1} \circ f$ and we can compute a formula for M_L. Having done all this secretly on scratch paper, we make the following claim.]

Claim: *Let* $c = (a\bar{b} + b)/2$, $g(z) = a\overline{z} + (b - a\bar{b})/2$.

Then a) $g(g(0)) = 0$. *(So by (iii) g is a reflection in a line L. Denote $g = M_L$.)*

b) $T_c \circ M_L = f$

c) *c is parallel to the line L.*

This claim immediately implies (iv).

To prove the claim, a) and b) are straightforward calculations. To see c), note that $b \neq 0$ since $a\bar{b} + b \neq 0$. Also, since $M_L(z)$ is a reflection of the form $a\overline{z} + \text{constant}$, the axis of reflection L makes an angle of $\theta/2$ with the x-axis, by (iii). To see that c makes the same angle with the x-axis, and is thus parallel to L, notice that:

$$c = (a\bar{b}+b)/2 = (e^{i\theta}\bar{b}+b)/2 = e^{i\theta/2}\big(e^{i\theta/2}\bar{b}+e^{-i\theta/2}b\big)/2 = e^{i\theta/2}\cdot(\text{ non-zero real number}).$$

The non-zero real number comes about because $c \neq 0$ and $e^{i\theta/2}\bar{b} + e^{-i\theta/2}b$ is of the form $\bar{w} + w = (x + iy) + (x - iy)$, for $w = e^{i\theta/2}\bar{b}$.

Since c is a non-zero real multiple of $e^{i\theta/2}$, it makes the same angle, $\theta/2$, with the x-axis. Hence c is parallel to L as claimed. Part (iii) of this theorem implies that L passes through $(b - a\bar{b})/4$ and meets the x-axis in angle $\theta/2$. ∎

To make the classification of isometries of Theorem I.6.1 definitive[4], we must guarantee that an isometry f cannot be given by two different formulas for $f(z)$.

Theorem I.6.2 *If f is an isometry other than the identity, then f belongs to exactly one of the four classes in Theorem I.6.1. Furthermore, there is only one formula for f of the form $f(z) = az + b$ or $f(z) = a\overline{z} + b$.*

Proof: From I.5.2, $f = T_b\, R_\theta\, M^\epsilon$. There is no choice for what b is since

$$f(0) = T_bR_\theta M^\epsilon(0) = T_b(0) = b.$$

There is also no choice for what a is since

$$f(1) = T_bR_\theta M^\epsilon(1) = T_bR_\theta(1) = T_b(a) = a + b,$$

so that
$$f(1) - f(0) = a.$$

Finally, there is no choice for what ϵ is, for if

$$f = T_bR_\theta \circ \mathrm{id} = T_{b_1}R_{\theta_1}M,$$

then by the discussion above, $b_1 = b$ and $\theta_1 = \theta$. Therefore, for all $z \in \mathbb{C}$, we would have the following formulas:

$$f(z) = az + b = a\overline{z} + b.$$

[4]A *classification* of objects names or lists the objects so that there is a one-one correspondence between the set of objects and the set of things (names) in the list.

Applying this to $z = i$, we would get

$$f(i) = ai + b = a(-i) + b \quad \text{where} \quad a \neq 0 \quad \text{since} \quad |a| = 1.$$

But then it would follow that $i = -i$, which is surely a contradiction. Thus, given f, there is no choice as to the value of ϵ in writing $f = T_b R_\theta M^\epsilon$.

We conclude that f can only satisfy one of the possible formulas given by I.5.2 and then, from the proof of I.6.1 belongs to only one of the four classes. ∎

Examples illustrating the Classification Theorem.

- Let $f(z) = iz + (2 + i)$. In the Classification Theorem we have $a = i$, $\theta = \pi/2$, and $b = (2 + i)$. We conclude f is a rotation by $\pi/2$ radians counterclockwise about the center $c = b/(1 - a) = (2 + i)/(1 - i) = \frac{1}{2} + \frac{3}{2}i$.
- Let $f(z) = \overline{z} + i$. We have $a = 1$, $\theta = 0$, and $b = i$. Notice that $a\bar{b} + b = 0$. We conclude that f is a reflection in the line L through $i/2$ making angle zero with the x-axis, i.e. in the horizontal line through $i/2$.
- Let $f(z) = i\overline{z} + i$. We have $a = i$, $\theta = \pi/2$, and $b = i$. Notice that $a\bar{b} + b = 1 + i \neq 0$. We conclude that f is a glide-reflection $T \circ M_L$ with translation $T(z) = z + (1 + i)/2$ and reflection $M_L(z)$ in the line L making angle $\pi/4$ with the x-axis and passing through the point $(b - a\bar{b})/4 = (-1 + i)/4$. Notice that the line L is parallel to $(1 + i)/2$.

Formulas for isometries in linear algebra notation

To set up the linear algebra version of the formulas in Theorem I.6.1 we again use I.4.10 and I.4.17. With respect to the basis $(1, 0), (0, 1)$, the(function R_θ, (= rotation by angle θ about the origin) is a linear transformation given by the matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

and the reflection $M_{\theta/2}$ is given by the matrix

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

Definition I.6.3 *The* **orthogonal group** O_2 *is the set of all 2-by-2 matrices of the form*

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}, \quad (\theta \in \mathbb{R}),$$

equipped with the operation of matrix multiplication.

The elements $A \in \mathrm{O}_2$ are exactly those 2×2 matrices A such that $AA^T = I$. (See Exercise 3). Here A^T is the transpose of A, and I is the identity matrix.) Under matrix multiplication, O_2 does indeed form a group (Exercise **??**). For a complex number $b = b_1 + ib_2$, let $\vec{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ be the corresponding column vector in $\mathbb{R}^2$, and for $z = x + iy$, let $\vec{z} = \begin{pmatrix} x \\ y \end{pmatrix}$. The next theorem is an immediate consequence of Theorem I.6.1, together with the observations above.

Theorem I.6.4 *Every isometry* $F : \mathbb{C} \longrightarrow \mathbb{C}$ *may be written as a function* $F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ *in one and only way in the following form:*

$$F(\vec{z}) = A\vec{z} + \vec{b} \quad \textit{where} \quad A \in O_2,\ \vec{b} \in \mathbb{R}^2.$$

(The notation $A\vec{z}$ *denotes multiplication of the matrix* A *by the vector* $\vec{z}$*).*

■

Exercises I.6

1. Classify each of the following functions as a translation, rotation, reflection or glide reflection. In each case give the crucial data (e.g, translation vector, center and angle of rotation, axis of reflection, or the decomposition of glide reflection as reflection composed with translation by a vector parallel to the the axis of reflection).

   (a) $z \mapsto iz + i$

   (b) $f(z) = \Big(\dfrac{-1 - \sqrt{3}i)}{2}\Big)\overline{z} + (\sqrt{3} + i)$

   (c) $f(z) = \Big(\dfrac{7+i}{\sqrt{50}}\Big)\overline{z} + (3 - i)$

   (d) $f(z) = \Big(\dfrac{15 + 8i}{17}\Big)z + 6$

2. Let Δ be the equilateral triangle with vertices at 1, $e^{2\pi i/3}$, $e^{-2\pi i/3}$. Find an isometry which sends Δ to a triangle which is centered at $1+i$ with one vertex at $1 + 2i$. Find *all* isometries which meet the stated requirement. How are they related?

3. Characterization of the elements of O_2:

   (a) If A is an $n \times n$ matrix with $AA^T = I$ show that each column (viewed as a vector) has length 1 and that any two column vectors are orthogonal.

   (b) Show that the 2×2 matrices $A \in O_2$ are precisely those 2×2 matrices A satisfying $AA^T = I$.

4. Show that if $A \in O_2$ has determinant one, then the linear transformation $\vec{z} \mapsto A\vec{z}$ is a rotation about the origin, and find the angle in terms of the entries of A. Show that if A has determinant -1, then A is a reflection in a line through the origin, and find the angle with the x-axis in terms of the entries of A.

I.7 The structure of the set of symmetries of a plane figure

Let's now return to our original problem of modelling symmetries. We have chosen to describe them by using functions, and we restricted our functions to the class of isometries of $\mathbb{C}$. In the previous section, we gave several examples of types of isometries which arise as symmetries of the six figures in Section I.1, and pointed out some general features of isometries. We now formalize our notion of symmetries as follows.

Definition I.7.1 *[Symmetry of a plane figure] Let P be a subset of the plane. A* symmetry *of P is an isometry $f : \mathbb{C} \to \mathbb{C}$ such that $f(P) = P$. We denote by* $\mathrm{Sym}(P)$ *the set of all symmetries of P.*

Note that $\mathrm{Sym}(\mathbb{C}) = \mathrm{Isom}(\mathbb{C})$, the set of all isometries of $\mathbb{C}$.

This definition means that the isometry f takes the set P onto itself, though the individual points of P may move around on P. For example, if P is a square centered at the origin, with one side parallel to the x-axis, then the rotation $R_{3\pi/2}$ and the reflection $M_{3\pi/4}$ are symmetries of P.

Remark: A more natural definition of a symmetry of P would be that it is a distance-preserving bijection $f : P \longrightarrow P$, Why do we require also that f preserve distances between all points in $\mathbb{C}$ — not just those in P? There are two reasons:

- The definition we use leads to a unified theory, allowing us to reason about all subsets $P \subset \mathbb{C}$ with a single viewpoint.

- If P contains three non-collinear points then every such function $f : P \longrightarrow P$ extends to one and only one isometry $F : \mathbb{C} \longrightarrow \mathbb{C}$. (Exercise 2, Section I.5.) So for these sets P, which are the vast majority of those we are interested in, the theory is unchanged by our seemingly more restricted definition.

The algebraic structure of Sym(P)

1. If f, $g \in \mathrm{Sym}(P)$, then so is $f \circ g$.

   This holds, since (i) by Proposition I.4.5, the composition $f \circ g$ is an isometry; and (ii) since f and g send P onto itself, the composition $f \circ g$ does too.

2. If $f, g, h \in \mathrm{Sym}(P)$, then $f \circ (g \circ h) = (f \circ g) \circ h$.

   This follows from more general facts about composition of functions; see Appendix D.

3. There is a function, namely $\mathrm{id}_{\mathbb{C}}$, which is in $\mathrm{Sym}(P)$ and for which $f \circ \mathrm{id}_{\mathbb{C}} = \mathrm{id}_{\mathbb{C}} \circ f = f$ whenever $f \in \mathrm{Sym}(P)$.

4. Finally, if $f \in \mathrm{Sym}(P)$, then $f^{-1} \in \mathrm{Sym}(P)$.

   This holds since (i) by Proposition I.4.6 the inverse of an isometry is an isometry; and (ii) since f sends P onto itself, so does f^{-1}.

Thus, we see that $\mathrm{Sym}(P)$ has the structure given by the following theorem.

Theorem I.7.2 ($\mathrm{Sym}(P)$ **is a group under composition)** *If P is a subset of $\mathbb{C}$, then the set $\mathrm{Sym}(P)$, together with the operation $\circ$ of composition, satisfies the following four properties:*

1. **Closure:** *For all $f, g \in \mathrm{Sym}(P)$, $\quad f \circ g \in \mathrm{Sym}(P)$.*
2. **Associativity:** *For all $f, g, h \in \mathrm{Sym}(P)$, $\quad (f \circ g) \circ h = f \circ (g \circ h)$.*
3. **Identity:** *There exists an element $id = id_{\mathbb{C}} \in \mathrm{Sym}(P)$ such that for all $f \in \mathrm{Sym}(P)$, $f \circ id = f = id \circ f$.*
4. **Inverses:** *For each $f \in \mathrm{Sym}(P)$, there exists an element, denoted by f^{-1}, for which $f \circ f^{-1} = f^{-1} \circ f = id$.* ∎

The properties (1)-(4) are satisfied no matter what P is and are, in fact, common to many other diverse areas of mathematics, where the set $\mathrm{Sym}(P)$ is replaced by some

Figure 9. Symmetries of an equilateral triangle.

other kind of set, and the operation $\circ$ is replaced by some other kind of operation. In the next chapter, we will formally abstract these four properties, and use them in defining a *group*. Though $\mathrm{Sym}(P)$ will serve as a motivating example, it will be important to keep in mind that there are many other examples as well.

Exercises I.7

1. Let P be an equilateral triangle in the plane $\mathbb{C}$ with center O. (See Figure 9.)

   Let M_0, M_1, M_2 denote the indicated reflections across the lines making angles $0, \pi/3$ and $2\pi/3$ radians with the x-axis and R_1, R_2 the rotations by $2\pi/3$ and $-2\pi/3$ radians. Together with the identity, we have

   $$\mathrm{Sym}(P) = \{\mathrm{id}, R_1, R_2, M_0, M_1, M_2\}.$$

   Complete the following "multiplication table" for how the elements of Sym(P) combine with one another.[1] *Hint:* Keep track of where the vertices go.

[1] The entry in the *row* labelled M_1 and the *column* labelled M_2 is the composition $M_1 \circ M_2$, etc.

	id	R_1	R_2	M_0	M_1	M_2
id	id	R_1	R_2	M_0	M_1	M_2
R_1	R_1		id			
R_2						
M_0						
M_1						R_2
M_2						

What kinds of properties do you notice about this table?

2. Suppose, in the exercise above, that O is the origin in the complex plane and let the vertex labelled "A" be the complex number 1. Find the vertices labelled "B" and "C" as complex numbers. Find formulas for R_1, R_2 of the form $z \mapsto az$ and formulas for M_0, M_1, M_2 of the form $z \mapsto a\overline{z}$, where $a = e^{i\eta}$ for some η.

3. Suppose that P is the circle of radius one centered at the origin.

   (a) Exhibit two elements $g, h \in \mathrm{Sym(P)}$ such that

   $$\text{a) } g \neq h, \quad \text{and} \quad \text{b) } g \neq \mathrm{id},\ h \neq \mathrm{id}, \quad \text{but} \quad \text{c) } g \circ g = h \circ h = \mathrm{id}$$

   (b) Exhibit an element $g \in \mathrm{Sym(P)}$ such that $g \neq \mathrm{id}$, but $g \circ g \circ g = \mathrm{id}$

   (c) Exhibit an element $h \in \mathrm{Sym(P)}$ such that $h \neq \mathrm{id}$, and for every positive integer n

   $$\underbrace{h \circ h \circ h \ldots \circ h}_{n \text{ times}} \neq \mathrm{id}.$$

4. Let $\mathbb{R} \subset \mathbb{C}$ be the real line (the x-axis) in the plane $\mathbb{C}$. Let $g \in \mathrm{Sym}(\mathbb{R})$ with $g \neq \mathrm{id}$.

   (a) If $g(0) < g(1)$ find a formula for $g(x),\ x \in \mathbb{R}$. Is there an element $x \in \mathbb{R}$ with $g(x) = x$? (Such an element is called a *fixed point* of g.)

   (b) If $g(0) > g(1)$ find a formula for $g(x),\ x \in \mathbb{R}$. Does g have a fixed point?

   (c) Name precisely all the isometries of $\mathbb{C}$ which belong to $\mathrm{Sym}(\mathbb{R})$.

5. Prove or give a counterexample: If $g,\ h \in \mathrm{Sym}(\mathbb{R})$ then $g \circ h = h \circ g$.

n

Chapter II

Group theory: the beginnings

II.1 Definition and numerical examples

We finished our discussion of isometries in Chapter I by noting that in the set of symmetries of a subset P of the plane – denoted $\mathrm{Sym}(P)$ – we can compose any two elements f, g to get a new symmetry $f \circ g$, and that this rule for combining elements of $\mathrm{Sym}(P)$ satisfies four key properties (Theorem I.7.2). This now leads us into a study of *group theory*. The groups $\mathrm{Sym}(P)$ and, in particular, $\mathrm{Sym}(\mathbb{C}) = \mathrm{Isom}(\mathbb{C})$ are the first and most important examples of groups in this book.

A rule for combining the elements of a set G to get an element of G is called a *binary operation* on G. Here is the formal definition:

Definition II.1.1 (Binary operation) *A* **binary operation on** G *is a function* $B : G \times G \to G$.

By the definition of "function", a binary operation G gives, for each ordered pair $(a, b) \in G \times G$, a single result $B(a, b)$. Typically $B(a, b)$ is shortened to one of the following forms:

$$ab, \quad a \cdot b, \quad a + b \quad a * b, \quad a \circ b.$$

There are many important situations in which we run across a binary operation which satisfies the same four properties as composition in Isom($\mathbb{C}$). We single these properties out as "axioms", which may or may not hold in a given situation, and when they do hold we say that we have a group:

Definition II.1.2 (Group) *A* group *consists of a non-empty set G and a binary operation, written here as $(a, b) \mapsto ab$, which satisfies the following four axioms:*

- **Closure:** [1] If a, $b \in G$ then $ab \in G$.
- **Associativity:** If a, b, $c \in G$ then $(ab)c = a(bc)$.
- **Identity:** *There is an element $e \in G$ such that $eg = ge = g$ for all $g \in G$.*
- **Inverses:** *If $g \in G$ then there is an element $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = e$ (where e is an identity, as above).*

Convention: The statement that "G is a group" will carry with it the understanding that the operation is written as $(a, b) \mapsto ab$, unless a statement to the contrary is made.

Since a group consists of a set G *and* a binary operation, a group is formally an ordered pair (set, operation). For example we refer in Example 1.4 below to "the group $(\mathbb{R}, +)$, where $\mathbb{R}$ is the set of real numbers and '+' is the operation of ordinary addition".

Before proceeding with a large number of examples, it is useful to prove the following lemma, which says that there is only one identity element in a group, and that each element of the group has only one inverse. Note that, since our axioms do not assert these facts explicitly, they cannot be assumed to be true without proof. Indeed, many "obvious" facts about familiar number systems (like the "commutative law" $ab = ba$ discussed in the next section) do not hold in every group.

[1] This is redundant, following from our assumption that we have a binary operation on G, but we state it for future reference.

Lemma II.1.3 (Uniquess of identity and inverses) *Suppose that G is a group.*

1. *If $e_1 \in G$, and $e_2 \in G$ both satisfy the Identity axiom then $e_1 = e_2$.*
   Thus there is only one identity element in G. We denote this by e.
2. *Suppose that $g \in G$ and that $g^{-1} \in G$ is an inverse of G, as given by the Inverses axiom. If $\bar{g}$ is an element of G such that $g\bar{g} = e$ or $\bar{g}g = e$ then $\bar{g} = g^{-1}$.*
   Thus g has only one inverse in G. We denote it by g^{-1}.

Proof:[2]

1. e_1e_2 is a single element of G, since our multiplication is a binary operation – a function – which only gives one element of its range G for any element of its domain $G \times G$ (in particular for the element $(e_1, e_2) \in G \times G$). Now we have

$$e_1 = e_1e_2 = e_2.$$

The first equality is due to the fact that e_2 is an identity element, and the second equality is due to the fact that e_1 is an identity element. Therefore $e_1 = e_2$.

2. Suppose that $g\bar{g} = e$. Then, multiplying by g^{-1} we get:

$$\begin{array}{rrcll} & g\bar{g} & = & e & \\ \Longrightarrow & g^{-1}(g\bar{g}) & = & g^{-1}e & \text{(since a binary operation is a function)} \\ \Longrightarrow & (g^{-1}g)\bar{g} & = & g^{-1} & \text{(Associativity axiom, property of } e\text{)} \\ \Longrightarrow & e\bar{g} & = & g^{-1} & \text{(property of } g^{-1}\text{)} \\ \Longrightarrow & \bar{g} & = & g^{-1} & \text{(identity property of } e\text{.)} \end{array}$$

The proof in the case that $\bar{g}g = e$ is the same, except that one writes g^{-1} on the right rather than on the left in these implications. ∎

[2]When a proof is given, the best procedure for the reader is to start by closing the book. Ask yourself, "What is being claimed?", "What is the relevance of each hypothesis?", "How would I try to prove this?" Only after a personal attempt to grapple with the situation has succeeded or failed will it be useful and meaningful to read the proof presented in the book. Beginning group theory is an especially good subject in which to begin to develop this habit.

Examples

We start by considering familiar numerical systems. In the following sections we will move on to more exotic examples (like Isom $\mathbb{C}$ which motivated our definition of a group), culminating in the discussion of *transformation groups* in II.4.

Example II.1.4 (Reals under addition form group) $(\mathbb{R}, +)$ *is a group, where* $\mathbb{R}$ *is the set of real numbers and '+' is the operation of addition. The identity element is the real number* 0 *and the inverse of an element* $x \in \mathbb{R}$ *is* $-x$.

Verification of the axioms follows by appealing to the familiar properties of the arithmetic of real numbers. Similarly we have the following examples:

Example II.1.5 (Groups from arithmetic) $(\mathbb{C}, +)$, $(\mathbb{Q}, +)$, and $(\mathbb{Z}, +)$ *are groups, where* $\mathbb{C}$ *is the set of complex numbers,* $\mathbb{Q}$ *is the set of rationals, and* $\mathbb{Z}$ *is the set of integers.*

In each case $+$ is the operation of ordinary addition. Note that $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$, and in each case the binary operation on the subset is the restriction of that on the bigger set. (As a function, it has a smaller domain, but is given by the same rule – ordinary addition.) This illustrates one basic way in which new examples of groups arise; viz., as *subgroups* of known groups.

Definition II.1.6 (Subgroup) *If* G *is a group then a subset* H *of* G *is a* subgroup of G *if* H *becomes a group when the elements of* H *are combined by the same rule by which they are combined in* G. *If* H *is a subgroup of* G, *we denote this by* $H < G$.

More technically: If G is a group with binary operation $B : G \times G \longrightarrow G$ then H is a subroup of G if $B(H \times H) \subset H$ and if, for the binary operation B' on H given by the restriction

$$B' \;=\; B|H \times H : H \times H \longrightarrow H,$$

the pair $(H,\ B')$ satisfies the group axioms given in Definition II.1.2. See Appendix D for more details on restrictions.

For example, with the operation of addition,

$$\mathbb{Z} < \mathbb{Q} < \mathbb{R} < \mathbb{C}.$$

Another important example is the *circle group*

$$S^1 = \{z \mid\ |z| = 1\} = \{z \mid\ z = e^{i\theta},\ \text{for some } \theta \in \mathbb{R}\} \ < \ \mathbb{C} - \{0\},$$

where the operation is multiplication.

Example II.1.7 (proper and improper, trivial and non-trivial subgroups) . *The sets* $\{e\}$ *and* G*, with the binary operation inherited from* G*, are subgroups of the group* G*. These are called the* improper subgroups *of* G*. All other subgroups of* G *are called* proper subgroups*. The subgroup* $\{e\}$ *is called the* trivial subgroup *and any other subgroup is called* non-trivial*.*

Continuing with our numerical examples, we give another example, sometimes called "clock arithmetic".

Example II.1.8 (Addition modulo n) *Let* n *be a positive integer. Let*

$$\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$$

and define a binary operation $+_n$*, called* addition modulo n*, on* $\mathbb{Z}_n$ *by*

$$a +_n b = \begin{cases} a + b & \text{if } \ 0 \le a + b < n, \\ a + b - n & \text{otherwise .} \end{cases}$$

Then $(\mathbb{Z}_n, +_n)$ *is a group.*

The proof is Exercise 4. The identity element is 0, and the inverse of a is $n - a$.

Note: If $x,\ c \in \mathbb{Z}$, we say that $x = c$ (modulo n) if $x = qn + c$ for some $q \in \mathbb{Z}$, i.e. if the remainder upon division by n is c.

Turning to ordinary numerical multiplication, we have the following results.

Example II.1.9 ($(\mathbb{R}, \cdot)$ is not a group) $\mathbb{R}$ *is* not *a group under the operation of ordinary multiplication.*

There are many, many possible proofs.

Proof: Suppose to the contrary that it is a group. Then it has an identity element, call it e.

Since $1 \cdot 5 = 5$, Lemma II.1.3 part 1, applied with $g = 5$ and $e_1 = 1$ implies that $1 = e$.

Since $7 \cdot 0 = 0$, Lemma II.1.3 part 1, applied with $g = 0$ and $e_1 = 7$ implies that $7 = e$. Hence $1 = e = 7$, which is impossible. ∎

Example II.1.10 ($(\mathbb{R} - \{0\}, \cdot)$ is a group) *Let* $G = \mathbb{R} - \{0\}$. *Then* G *is a group under the ordinary operation of multiplication. The idenity element is* 1 *and the inverse of* $x \in \mathbb{R} - \{0\}$ *is its reciprocal* $1/x$.

Again, this follows from properties of arithmetic.

Similarly, $\mathbb{C} - \{0\}$ and $\mathbb{Q} - \{0\}$ are groups under multiplication, and we have

$$\mathbb{Q} - \{0\} < \mathbb{R} - \{0\} < \mathbb{C} - \{0\} \quad \text{and} \quad S^1 < \mathbb{C} - \{0\}$$

Note, however, that $\mathbb{Z} - \{0\}$ is not a group under multiplication.

Although $\mathbb{Z} - \{0\}$ is not a group under ordinary multiplication, we have the following fact (hints for the proof are given in Exercise **??**.), wherein we consider multiplication modulo p in the set $\mathbb{J}_p = \mathbb{Z}_p - \{0\}$. (E.g., $3 \cdot 4 = 5$ in $\mathbb{J}_7$.)

Example II.1.11 (Multiplication modulo p a prime) *When* p *is a positive prime number, let* $\mathbb{J}_p = \{1, 2, ..., p-1\}$ *with the operation of multiplication modulo* p. *Then* $\mathbb{J}_p$ *is a group.*

Exercises II.1

1. The following "obvious" statement depends on an underlying assumption that we are taught to make in childhood. It can be proved using the definition II.1.2 of a group. Prove it.

   If a, b, c, are elements of the group G, and if $b = c$ then $a\,b = a\,c$.

2. **Axioms.** Decide whether G is a group in each situation:

   (a) G = the set of all integers with the operation of ordinary subtraction.

   (b) $G = \{1, 3, 5\}$ under the operation of multiplication modulo 6.

   (c) $G = \{0, 3, 6\}$ under the operation of addition modulo 9.

   (d) $G = \{1, i, -1, -i\}$ under the operation of multiplication of complex numbers.

3. **Cancellation laws.** If a, b belong to the group G,

   - prove that the equations

     $$ax = b \qquad \text{and} \qquad ya = b$$

     are each satisfied by one and only one x or $y \in G$.
   - Is there any reason to think that $x = y$?

4. If x, y, x_1, x_2, ..., x_n are elements of a group, prove or give a counterexample to each of statements (a) and (b). Complete (c).

   (a) $(xy)^{-1} = x^{-1}y^{-1}$

   (b) $(xy)^{-1} = y^{-1}x^{-1}$
   Hint: Complete the sentence: "The reverse of putting on your socks and then putting on your shoes is ...."[3]

   (c) $(x_1x_2 \dots x_n)^{-1} = ?$

5. Prove each of the following assertions:

   (a) $\mathbb{Z}_n$ is a group (under the operation of addition modulo n).

[3]For example, if A, B are matrices in O_2, then $(AB)^{-1} = B^{-1}A^{-1}$.

(b) $\mathbb{Z} - \{0\}$ is not a group under multiplication.

(c) If p is a prime number then $\mathbb{J}_p = \{1, 2, ..., p-1\}$ is a group under multiplication modulo p.

Hints: Use **i)** the fact that $\mathbb{Z}$ is associative under multiplication, **ii)** every integer n can be written as $n = ap + r$ with $a, r \in \mathbb{Z}$, $0 \leq r < p$ and **iii)** if n and m are relatively prime then there exist integers a, b with $am + bn = 1$. (iii) will be proved as Example II.6.12)

6. List all the subgroups of the following groups, equipped with the usual operations of addition:

(a) $\mathbb{Z}$

(b) $\mathbb{Z}_3$

(c) $\mathbb{Z}_6$

Hint: In each case, if G is a subgroup other than $\{0\}$ then G contains a smallest positive element. Use this element to decide what else is in G.

II.2 Isomorphic groups

A rose by any other name would smell as sweet [4]

Having defined the concept of a group, we now consider a question which occurs when a new concept is defined in any field of mathematics:

When shall we view two objects as really being the same?

Frequently we come across two groups G and H which are made of different sets with operations on those sets but for which we think:

Those are really the same group under different names.

Example II.2.1 (Rotations by multiples of $2\pi/n$ and $\mathbb{Z}_n$)

Let $n \geq 2$ be a positive integer. Let

$$\mathcal{R}_n \;=\; \{R_0,\; R_{\frac{2\pi}{n}},\; R_{\frac{4\pi}{n}}, \ldots R_{\frac{2\pi(n-1)}{n}}\};$$

i.e. $\mathcal{R}_n$ is the set of all rotations about the origin by integer multiples of $2\pi/n$, with the operation of composition. One can check that $\mathcal{R}_n$ is a group (Exercise 1).

If we stop and think about it, we see that in some sense, this group is the same as $\mathbb{Z}_n$: The number j plays the same role in $\mathbb{Z}_n$ as $R_{\frac{2\pi j}{n}}$ does in Isom($\mathbb{C}$). Addition modulo n of i and j to get $i +_n j = (i + j)$ (modulo n) is analogous to composition of rotations $(R_{\frac{2\pi i}{n}}) \circ (R_{\frac{2\pi j}{n}}) = R_{\frac{2\pi(i+j)}{n}}$. It is as though the number $j \in \mathbb{Z}_n$ simply got renamed to $R_{\frac{2\pi j}{n}}$ and "$+_n$" got renamed to "$\circ$".

[4] **Romeo and Juliet,**, Act 2, Scene 2:
Juliet: "What's in a name? that which we call a rose
By any other name would smell as sweet."

Example II.2.2 (Logarithms) *Of great importance in the history of science was Napier's discovery (1614-15) that the group* $(\mathbb{R}_+, \cdot)$ *of positive real numbers under multiplication could be matched up in a bijective way with the group* $(\mathbb{R}, +)$ *of all real numbers under addition by using the logarithm map* $\log : \mathbb{R}_+ \longrightarrow \mathbb{R}$*, with the result that* $\log(x \cdot y) = \log(x) + \log(y)$*. While these groups feel very different to us psychologically, they "are really the same" in how they behave.*

We formalize this notion of "sameness" with the notion of an *isomorphism*, which has two parts to it.

Definition II.2.3 (homomorphism and isomorphism) *Suppose that* G *and* H *are groups (say, both with operation written multiplicatively).*

- *A function* $\phi : G \longrightarrow H$ *is a* **homomorphism** *if*

$$\phi(a \cdot b) = \phi(a) \cdot \phi(b) \text{ for all } a, b \in G.$$

- *A function* $\phi : G \longrightarrow H$ *is an* **isomorphism** *if* ϕ *is a homomorphism which is also a bijection.*
- *The groups* G *and* H *are said to be* **isomorphic***, and we write* $G \cong H$ *if there exists an isomorphism* $\phi : G \longrightarrow H$*.*

If there is a homomorphism from G onto H, then intuitively the multiplication in H reflects some aspect of the multiplication in G. Homomorphisms are of extreme importance and we will study them in Chapter IV.

The existence of an isomorphism from G to H indicates that the elements of the two groups can be perfectly matched up so that the operation in one exactly copies the operation in the other. Thus the function ϕ given by $\phi(j) = R_{\frac{2\pi j}{n}}$ in Example II.2.1, and the log function in Example II.2.2 are examples of isomorphisms. (If the operation in G is written multiplicatively and that in H is written additively, then $\phi : G \longrightarrow H$ is a homomorphism if $\phi(ab) = \phi(a) + \phi(b)$. The fact that $\log(xy) = \log(x) + \log(y)$ thus shows that bijective function $\log : \mathbb{R}_+ \longrightarrow \mathbb{R}$ is indeed an isomorphism.)

The matching of elements in one group with elements in another which "act the same" should at the least require that the identity elements correspond and that if g gets matched with h then g^{-1} gets matched with h^{-1}. This is the content of the following proposition, which does not require the full strength of an isomorphism, but only the property of being a homomorphism.

Proposition II.2.4 *Suppose that G and H are groups with identity elements e_G and e_H respectively. If $\phi : G \longrightarrow H$ is a homomorphism then*

1. $\phi(e_G) = e_H$.

2. $\phi(g^{-1}) = (\phi(g))^{-1}$ *for all* $g \in G$.

Proof: 1. Let $g \in G$ be arbitrary. Then

$$\begin{array}{rcll} \phi(g) & = & \phi(e_G g) & \text{since } e_G g = g \\ & = & \phi(e_G)\phi(g) & \text{since } \phi \text{ is a homomorphism.} \end{array}$$

Now multiply both sides of the equation $\phi(g) = \phi(e_G)\phi(g)$ on the right by $(\phi(g))^{-1}$ to get
$e_H = \phi(e_G)$.

2. Let $g \in G$. Then $e_H = \phi(e_G) = \phi(gg^{-1}) = \phi(g)\,\phi(g^{-1})$. Hence the product of $\phi(g)$ and $\phi(g^{-1})$ is the identity element of H. By Lemma II.1.3 this implies $\phi(g^{-1}) = \phi(g)^{-1}$. ■

Finally, we point out that the assertion that two groups are really the "same" suggests implicitly that "sameness" has certain properties:

- **Reflexivity:** A thing is the same as itself
- **Symmetry:** If A is the same as B, then B is the same as A.
- **Transitivity:** If A is the same as B, and B is the same as C, then A is the same as C.

These three conditions are the axioms for an *equivalence relation*; see Appendix C for the precise definition. There will be several places in this book where they arise.

Proposition II.2.5 (Isomorphism is an equivalence relation) *The relation* $\cong$ *of isomorphism is an equivalence relation on any set of groups.*

The proof is Exercise 6.

Exercises II.2

1. In Example II.2.1, show that $G = \mathcal{R}_n$ is a group.
2. What well-known properties of logarithms are consequences of Proposition II.2.4?
3. (a) Let $G = \{1,\ i,\ -1,\ -i\}$ under the operation of ordinary multiplication of complex numbers. Prove that $G \cong \mathbb{Z}_4$.

   Note: The definition of *isomorphism* in II.2.3 is interpreted here to mean a bijection $\phi : G \longrightarrow \mathbb{Z}_4$ which satisfies $\phi(a \cdot b) = \phi(a) + \phi(b)$, since the operation in the range is written additively. (This also occurred with the log function in Example II.2.2.)

   (b) If you hadn't concluded in II.1, Exercise 2d) that G is a group, how could you prove this from the bijection you constructed in part a) ?
4. Show that the group $\mathbb{C} - \{0\}$ (under ordinary multiplication) is isomorphic to the group of 2×2 matrices of the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, a, b real, $a^2 + b^2 \neq 0$, under the operation of matrix multiplication. (See also Prop. I.3.1.)

5. Let h be a nonzero complex number. Let G be the group whose underlying set is $\mathbb{C} - \{-1/h\}$ and for which the operation is defined by $z \cdot w = z + w + zwh$. Prove that G is a group. What is the identity element of G? To which more common group is G isomorphic?

6. (a) Suppose that $\phi : G_1 \longrightarrow G_2$ is an isomorphism of groups. Since ϕ is, in particular, a bijection, there is an inverse function $\phi^{-1} : G_2 \longrightarrow G_1$. Is ϕ^{-1} also an isomorphism?

   (b) Prove that the composition of two isomorphisms is an isomorphism.

   (c) Prove Proposition II.2.5. In other words, prove that:

   $\bullet\ G \cong G. \qquad \bullet\ G \cong H \Longrightarrow H \cong G. \qquad \bullet\ [G \cong H \text{ and } H \cong K] \Longrightarrow G \cong K.$

   The definition of equivalence relation is discussed in Appendix C. Note that you will need parts (a) and (b) to prove that being isomorphic is symmetric and transitive.

II.3 Abelian and Non-Abelian groups

All of the examples in §II.1 are special in that they are *abelian* or *commutative* groups.

Definition II.3.1 (Abelian group) *A group G is* abelian *(synonymously,* commutative*) if $ab = ba$ for all $a, b \in G$.*

For a finite group, the property of being abelian is reflected in its *multiplication table*, which gives the values of the binary operation vv(function) $G \times G \longrightarrow G$.

Definition II.3.2 (Multiplication table) *If $G = \{e = g_1, g_2, \dots g_n\}$ is a finite group with elements listed in a given order, then the* multiplication table of G *is the $n \times n$ table in which the entry in the ith row and jth column is the element $g_i g_j$.*

	e	g_2	$\dots$	g_i	$\dots$	g_j	$\dots$	g_n
e	e	g_2		g_i		g_j		g_n
g_2	g_2	g_2g_2		g_2g_i		g_2g_j		g_2g_n
$\dots$								
g_i	g_i	g_ig_2		g_ig_i		g_ig_j		g_ig_n
$\dots$								
g_j	g_j	g_jg_2		g_jg_i		g_jg_j		g_jg_n
$\dots$								
g_n	g_n	g_ng_2		g_ng_i		g_ng_j		g_ng_n

A group is abelian if and only if its multiplication table is symmetric across the diagonal; i.e., the (i, j) entry $g_i g_j$ equals the (j, i) entry $g_j g_i$.

Cyclic groups

The simplest examples of abelian groups are cyclic groups:

Definition II.3.3 (Cyclic group) *A group G is* cyclic *if there is an element $g \in G$ such that*

$$G = \{ g^n \mid n \in \mathbb{Z} \}. \tag{II.1}$$

The element g is called a generator *of G.*

Here we have used the notation:

$$g^n = \begin{cases} e & \text{(the identity element of } G\text{)} \quad \text{if } n = 0 \\ g \cdot g \cdots g & (n \text{ terms}) \quad \text{if } n > 0 \\ g^{-1} \cdot g^{-1} \cdots g^{-1} & (-n \text{ terms}) \quad \text{if } n < 0. \end{cases}$$

Note: (Exercise 2): If g is an element of any group G and if $m, n \in \mathbb{Z}$, we have

$$g^m g^n = g^{m+n}, \qquad (g^m)^{-1} = (g^{-1})^m, \quad \text{and} \quad g^{mn} = (g^m)^n.$$

Using this fact, we have that in particular, for any two elements g^m and g^n of a cyclic group, $g^m g^n = g^n g^m$. **Thus cyclic groups are abelian**.

When the operation in an abelian group G is written additively[5] (using "+"), we write the inverse of an element g as $-g$. Thus, if G is cyclic, we may replace the multiplicative notation in Equation II.1 with the additive notation

$$G = \{ ng \mid n \in \mathbb{Z} \}.$$

By ng we mean

$$ng = \begin{cases} 0 & \text{(the identity element of } G\text{)} \quad \text{if } n = 0 \\ g + g + ... + g & (n \text{ terms}) \quad \text{if } n > 0 \\ -g + -g + ... + -g & (-n \text{ terms}) \quad \text{if } n < 0. \end{cases}$$

This is just a change in notation; the underlying ideas are the same.

Definition II.3.4 (Order of an element) *If g is an element of the group G then the* order of g, *denoted* $\text{ord}(g)$, *is the least positive integer n such that $g^n = e$. If there is no such integer, then g is said to have* infinite order.

[5] In practice, abelian groups are probably written additively at least as often as they are written multiplicatively. However, non-abelian groups are rarely written additively.

Example II.3.5

1. *The group* $(\mathbb{Z}, +)$ *is cyclic with generator* 1 *, which is an element of infinite order. (In the future we shall simply write* $\mathbb{Z}$ *for this abelian group, and the additive operation will be understood.)*

2. *The group* $(\mathbb{Z}_n, +_n)$ *is cyclic with generator* 1*, which is an element of order* n*. (In the future we shall simply write* $\mathbb{Z}_n$ *for this abelian group,and the additive operation will be understood.)*

Note that these *generators are not unique.* The group $\mathbb{Z}$ also has the infinite order generator -1, and $\mathbb{Z}_n$ can have many generators. For example, $\mathbb{Z}_9$ is cyclic with any of the elements of order nine — 1, 2, 4, 5, 7, or 8 — as generator. On the hand, 0 has order 1 and 3 and 6 have order 3.

Definition II.3.6 (Subgroup generated by an element) *If* G *is any group and* $g \in G$ *then* $H = \{g^n \mid n \in \mathbb{Z}\}$ *is called* the cyclic subgroup of G generated by g*, and is denoted by* $H = \langle g \rangle$*.*

Thus cyclic groups appear everywhere — anytime we have an element g of a group..

The statement of Definition II.3.6 contains a statement which requires proof, namely, that $\langle\, g \,\rangle$ is indeed a subgroup.

Proposition II.3.7 ($\langle\, g \,\rangle$ is a subgroup) *If* G *is a group and* $g \in G$ *let* $H = \langle\, g \,\rangle$ *= "the cyclic subgroup generated by* g*". Then* H *is a subgroup of* G*, and the number of elements in* H *is equal to the order of* g*.*

Proof: (The general problem of recognizing whether a given subset H of G is a subgroup will be analyzed in §II.5. At this point, the reader might either prove that $H = \langle\, g \,\rangle$ is a subgroup or read the following paragraph, as a warmup for that discussion.)

H is closed under the operation since, for any two elements $g^n, g^m \in H$ their product g^{n+m} is also in H. The associative law holds for any three elements in H because it holds for these elements as elements of G, and the product in H is the same as that in G. The element $g^0 = e$ of H satisfies the requirements of an identity, since it acts as the identity element in G. Finally, for any element $g^n \in H$ the element $g^{-n} \in H$ has $g^{-n}g^n = e$, so the Inverses axiom is satisfied. Thus H is a subgroup.

If g is of infinite order, then for any two distinct integers $m, n,$ we have $m - n \neq 0$, so $g^m g^{-n} = g^{m-n} \neq e$ and therefore $g^m \neq g^n$. Hence H is infinite.

If g has (finite) order m, then we first claim that the cyclic subgroup generated by g, which is defined as

$$H = \{..., g^{-2}, g^{-1}, e, g^1, g^2, ...\} = \{\ g^n \mid n \in \mathbb{Z}\ \},$$

is actually equal to

$$\{e, g, g^2, ..., g^{m-1}\}.$$

Obviously, the latter set is contained in H, so it is enough to prove that this set contains H. To see this, let n be any integer. Then n may be written as $n = qm + i, i \in \{0, 1, ..., m-1\}$. So $g^n = g^{qm+i} = (g^m)^q g^i = e^q g^i = g^i$, so this set contains H.[6] Finally, it remains to show that H has exactly m elements, i.e. if $0 \leq i < j \leq m - 1$, then $g^i \neq g^j$. For suppose otherwise, i.e. that $g^j = g^i$. Then $e = g^j g^{-i} = g^{j-i}$. But $0 \leq j - i < m$, contradicting the fact that g was assumed to have order m, and the order is the smallest positive integer for which $g^m = 1$. ∎

Note that in the above proof, although there are infinitely many *symbols* of the form $g^n,\ n \in \mathbb{Z}$, these symbols only describe finitely many *elements* of the group G if g has finite order.

The preceding proposition sheds light on another use of the word *order*:

Definition II.3.8 (Order of a group) *If G is a group then the* **order of** *G is the number of elements in G. If G has infinitely many elements we say that G is* **of infinite order**.

[6] At this point, one is tempted to call oneself done. However, we have not yet used one property of our hypotheses, namely, that the order is the *smallest* positive integer m such that $g^m = e$.

Thus Proposition II.3.7 may be restated as

If $g \in G$ and $H = \langle g \rangle$ then H is a subgroup of G and order g = order H.

Note: We shall further see (II.7.3), that if g is an element of a finite group G then the order of g divides the order of G.

Non-cyclic abelian groups

While cyclic groups are abelian, not all abelian groups are cyclic.

Example II.3.9 (Klein 4-group) *Let*

$$G = \{(0,0), (1,0), (0,1), (1,1)\}$$

and define a binary operation on G by

$$(a,b) + (c,d) = (a +_2 c, b +_2 d).$$

Then G is an abelian group of order four in which every element other than the identity has order two. In particular, G is abelian, but not cyclic.

The proof is Exercise 6.

Non-abelian groups

Not all groups are abelian. Moving away from familiar number systems towards other extremely important examples, we have the following.

Example II.3.10 Isom $\mathbb{C}$ *is not abelian. (*Isom $\mathbb{C}$ *is a group by Theorem I.7.2.)*

Proof: If T_1 is translation by 1, ($T_1(z) = z+1$) and R_π is rotation about the origin by π radians ($R_\pi(z) = -z$), then $T_1R_\pi(z) = -z+1$ while $R_\pi T_1(z) = -z-1$. In particular $T_1R_\pi(0) = 1$ while $R_\pi T_1(0) = -1$. Hence $T_1R_\pi \neq R_\pi T_1$. Thus the group Isom $\mathbb{C}$ is not abelian. ∎

Example II.3.11 ($\mathbf{GL_2}(\mathbb{R})$) *The* general linear group in dimension 2*, denoted* $GL_2(\mathbb{R})$*, is the group of* 2×2 *invertible matrices* $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *with* a*,* b*,* c*,* $d \in \mathbb{R}$*, equipped with the operation of matrix multiplication. This is a noncommutative group.*

Proof(that $\mathbf{GL_2}(\mathbb{R})$ is a group. The Closure and Associativity axioms can be verified by straightforward (and tedious) computation. A more conceptual proof is given in Appendix F on linear algebra. The identity element is the identity matrix $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. By doing the multiplication we find that the inverse of $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is given by

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

∎

Notice that the determinant appears in the denominator of the factor in front of the matrix giving the inverse: in general, an n-by-n matrix is invertible if and only if its determinant is nonzero. See Appendix F for this and related results.

However, $GL_2(\mathbb{R})$ is not abelian! Choose almost any two randomly chosen matrices A, B and you will find that $AB \neq BA$ (Exercise 8a). Indeed, abelian groups are very special and should be viewed as the exception, rather than the rule.

The non-abelian group $GL_2(\mathbb{R})$ contains several important subgroups.

- The *orthogonal group* O_2 was defined (I.6.3) as the set of matrices whose corresponding linear transformation is a rotation or reflection about the origin. One can verify that this is a group by taking compositions and inverses of rotations and reflections. Alternatively, this is the set of 2×2 matrices A with $AA^t = I$, where A^t denotes the transpose of A. Verification that this is a group also follows from the facts that $(A^t)^t = A$ and $(AB)^t = B^tA^t$.

- The *special linear group in dimension 2*, denoted $SL_2(\mathbb{R})$, is defined[7] as the group of all two-by-two matrices of determinant 1. Verification that this is indeed a group follows easily from the fact that $\det(AB) = \det(A)\det(B)$.

- These groups have important and similarly defined generalizations for all positive integers n, yielding groups

$$\mathrm{O}_n < GL_n(\mathbb{R}), \quad \mathrm{SL}_n(\mathbb{R}) < GL_n(\mathbb{R}) \quad \text{and} \quad \mathrm{SO}_n = \mathrm{O}_n \cap \mathrm{SL}_n(\mathbb{R}).$$

  SO_n is called *the special orthogonal group.*

  There are also corresponding groups where the entries are restricted to $\mathbb{Z}, \mathbb{Q}$ or $\mathbb{C}$.

Classification of groups

We've now seen many different groups from diverse areas of mathematics, and some adjectives describing different properties of groups. One might wonder just how many kinds of groups there are, say, of a given size.

As a first step, Exercise 10 asks the reader to prove a *classification theorem.* In general this consists of:

- A set of objects which we wish to classify
  (– in this exercise, all groups of order ≤ 4).
- An equivalence relation (— in this exercise, isomorphism).
- A list of objects (— in this exercise, $\{0\}$, $\mathbb{Z}_2$, $\mathbb{Z}_3$, Z_4, $\mathbb{Z}_2 \oplus \mathbb{Z}_2$).
- A proof that each object in the set is equivalent to one and only one object in the list.

However, given different equivalence relations on the same set of objects, we would expect different classification theorems. For example, consider the problem of classifying subgroups of order two in $O_2(\mathbb{R})$. Any such subgroup is isomorphic to $\mathbb{Z}_2$, hence

[7] If you knew everything there is to know about $SL_2(\mathbb{R})$ you would win every major prize in mathematics. See LANG, $\mathrm{SL}_2\mathrm{R}$, Graduate texts in Mathematics, Springer-Verlag 1975.

any two are isomorphic to each other. Might there not nevertheless be some reason to distinguish between a certain pair of such subgroups for, say, geometric reasons?

We shall return to this point in §III.4 and Chapter V.

Exercises II.3

1. Prove that a group G is abelian if and only if $(ab)^{-1} = a^{-1}b^{-1}$ for all $a, b \in G$.

2. If g is an element of a group G and if $m, n \in \mathbb{Z}$, prove that

$$g^m g^n = g^{m+n}, \qquad (g^m)^{-1} = (g^{-1})^m, \qquad \text{and} \qquad g^{mn} = (g^m)^n.$$

3. Prove that

   (a) Any two infinite cyclic groups are isomorphic.

   (b) Two finite cyclic groups are isomorphic if and only if they have the same number of elements.

4. (a) For each element $g \in \mathbb{Z}_{10}$, write down the sequence

$$g, \quad g+g, \quad ..., \quad g+g+g \quad \ldots g \quad (10 \quad \text{terms}).$$

   (b) List the generators of $\mathbb{Z}_{10}$.

5. Prove or disprove: The group $\mathbb{Q}$ (under addition) is cyclic.

6. Prove the statement in Example II.3.9, concerning the Klein 4-group.

7. The Klein 4-group is an example of the general construction of the **direct product** $G_1 \times G_2$ of two groups G_1 and G_2. This is the group whose set is the cartesian product and whose operation is given by:

$$\begin{array}{rcll} (a_1, a_2)(b_1, b_2) & = & (a_1b_1, a_2b_2) & \text{if written mvultiplicatively,} \\ (a_1, a_2) + (b_1, b_2) & = & (a_1 + b_1, a_2 + b_2) & \text{if written additively.} \end{array}$$

   Note: The direct product $G_1 \times G_2$ is often denoted $G_1 \oplus G_2$, and is referred to as the **direct sum**, especially when G_1 and G_2 are abelian groups being written additively.

(a) Show that the group $\mathbb{Z}_6$ is isomorphic to $\mathbb{Z}_3 \oplus \mathbb{Z}_2$.

(b) When is $\mathbb{Z}_m \oplus \mathbb{Z}_n \cong \mathbb{Z}_{mn}$?

8. **Center of a group.** The *center* $C(G)$ of a group G is defined as

$$C(G) = \{c \in G \mid cg = gc \quad \text{for all } g \in G\}.$$

Thus, $C(G) = G$ if and only if G is abelian.

(a) Show by example that $GL_2(\mathbb{R})$ is nonabelian.

(b) Find the center of $GL_2(\mathbb{R})$.

9. Show that under matrix multiplication,

$$O_n = \left\{\text{n-by-n matrices} A \mid AA^T = I\right\}$$

forms a non-abelian group , and that the determinant of any such matrix is ± 1. (O_2 was discussed in Sec. I.6, Exercises 3, 4.) Relevant facts about transposition of matrices and determinants are:

(a) $(A^T)^T = A$.

(b) $(AB)^T = B^T A^T$.

(c) $\det(A) = \det(A^T)$.

10. **Classification.**

(a) Show that any group of order 4 or less is abelian.

(b) Indeed, show that every group of order 4 or less is isomorphic to exactly one of the following five groups:
The trivial group $\{e\}$, $\mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_4$, $V_4 =$ the Klein 4-group.

11. (a) What would a classification of the set of all rectangles in the plane look like? What equivalence relation would you use? Can you determine an (infinite) list of objects (lets call them *models*) such that every rectangle is equivalent to one and only one of these models?

(b) Suppose you did this with the set of squares in the plane. Could you give an equivalence relation in which one model would suffice? An equivalence relation which would need infinitely many models?

II.4 Transformation groups and group actions

Historically, the axioms for a group did not come first. Examples came first, as they usually do in good mathematics. The first examples of interest were not the familiar numerical examples above, but were *transformation groups*, upon which much of modern mathematics has since been built:

Definition II.4.1 *A* **permutation** *of a set is a bijection from the set to itself.*

Definition II.4.2 (Transformation group) *If G is a set of permutations of a non-empty set X which forms a group under the operation of composition, then we call G a* **transformation group** *and we say that G* **acts on** *X or that we are given* **an action of** *G* **on** *X.*[8]

The first explicit use of group theory was by Évariste Galois (born 1811, killed in a duel 1832.). Given a polynomial $p(x) = a_0 + a_1x + ...a_nx^n$, Galois used the properties of a certain group of permutations of the roots of $p(x)$ to determine whether there was a formula for the roots in terms of radicals and the given coefficients. (See [Birkhoff and MacLane, Chapter 15].) For much of the nineteenth century examples of transformation groups arising in algebra, geometry, and even differential equations were investigated (e.g. Klein wrote an entire book on the icosahedron). Not until 1882 was the modern axiomatic definition which we gave in II.1 presented by Dyck as a way to model some of the key features which the examples had in common.

Permutation groups

Usually we will be interested in permutations of a set X which preserve some geometric or algebraic structure on X. We will discuss examples of such groups — Sym P and Aut G — below. But at the outset we deal with plain, unadorned sets.

[8]This terminology is often generalized to the following: an action of a group H on a set X is a homomorphism from H to the group G of bijections of X. Thus to each $h \in H$ a bijection of X is asociated. By contrast, in the definition we give, each h **is** a bijection.

In the context of set theory, as opposed to geometry, transformation groups are often called *permutation groups*. These are the subgroups of *symmetric groups*:

Definition II.4.3 (Symmetric group) *If X is a non-empty set then* **the symmetric group** *on X , denoted S_X, is the set of* all *permutations of X under the operation of composition.*

The fact that S_X is indeed a group (Exercise 1) follows readily from general properties of functions outlined in Appendix D. The identity in the group S_X is the identity function id_X since $\mathrm{id}_X \circ f = f \circ \mathrm{id}_X = f$ for all $f \in S_X$, and the inverse function f^{-1} plays the role of the inverse of f in the group since $f \circ f^{-1} = f^{-1} \circ f = \mathrm{id}_X$ for all $f \in S_X$.[9]

Finite permutation groups arise naturally in many applied problems of a combinatorial nature. If $X_n = \{1, 2 \ldots n\}$ and $S_n \stackrel{\text{def}}{=} S_{X_n}$, then S_n is a group of order $n!$. We will discuss S_n and its subgroups in detail in §II.7.

From the definitions above we see that,

> *A transformation group, or permutation group, is, by definition, a subgroup of a symmetric group S_X for some set X.*

From this, we get the following useful lemma.

Lemma II.4.4 (Recognizing transformation groups) *A nonempty subset G of S_X is a transformation group if it satisfies*

- *the closure axiom*
- *If $f \in G$ then the inverse function f^{-1} also belongs to G.*

Proof: Closure is satisfied by hypothesis. The associative law is satisfied for $f, g, h \in G$ because $(f \circ g) \circ h = f \circ (g \circ h)$ for any functions for which the image of each

[9]The subtle point here is to realize that there are really two different meanings of the words "identity" and "inverse"' which must be shown to coincide.

function is contained in the domain of the next. In this case, image and domain are always equal to X. Choosing any $f \in G$ we can compose it with its inverse function, which is in G by hypothesis, to get id_X. The latter is in G by the closure assumption. As noted above, the identity function and inverse functions satisfy the conditions of identity and inverse elements. ∎

The symmetry group of a plane figure

We have already met the most important type of transformation group for our study of the geometry of plane point sets:

Example II.4.5 (Sym P) *If $P \subset \mathbb{C}$ then $\mathrm{Sym}(P)$ is a transformation group, where $\mathrm{Sym}(P)$ is the symmetry group of P – the group of all isometries of $\mathbb{C}$ taking P onto itself. (See I.7.2.)*

Instead of considering all isometries of P, we may take a subgroup consisting of a subset of $\mathrm{Sym}(P)$ which itself forms a group under the operation of composition. Here are some subgroups of $\mathrm{Isom}(\mathbb{C}) = \mathrm{Sym}(\mathbb{C})$ which arise if we restrict our attention to translations, rotations and reflections.

Example II.4.6 (Translations form a group) *Let $\mathcal{T} \subset \mathrm{Isom}(\mathbb{C})$ consist of the set of translations $T_b : \mathbb{C} \to \mathbb{C}$, where $T_b(z) = z + b$. Then $\mathcal{T}$ is an abelian subgroup of $\mathrm{Isom}(\mathbb{C})$, and the function $\mathcal{T} \to \mathbb{C}$ given by $T_b \mapsto b$ is is an isomorphism.*

Proof: If $b_1, b_2 \in \mathbb{C}$, then for all $z \in \mathbb{C}$ we have

$$T_{b_1} \circ T_{b_2}(z) = (z + b_2) + b_1 = z + (b_2 + b_1) = T_{b_2+b_1}(z).$$

Hence
$$T_{b_1} \circ T_{b_2} = T_{b_2+b_1} = T_{b_1+b_2} = T_{b_2} \circ T_{b_1}.$$

Therefore $\mathcal{T}$ is closed under composition and this operation is commutative. Further, given $T_b \in \mathcal{T}$, its inverse is the translation T_{-b} since $T_{-b} \circ T_b(z) = z = \mathrm{id}_{\mathbb{C}}(z)$ for all $z \in \mathbb{C}$. Hence $\mathcal{T}$ is a subgroup of $\mathrm{Isom}(\mathbb{C})$ by Lemma II.4.4.

The claim that we have an isomorphism $\mathcal{T} \to \mathbb{C}$ is left to the reader as Exercise 2. ∎

Example II.4.7 (Rotations with a given center form a group) *For $a \in \mathbb{C}$, let $\mathcal{R}_a$ denote the set of all rotations $R_{a,\eta}$ of the plane with center a. Then $\mathcal{R}_a$ is an abelian subgroup of* $\mathrm{Isom}(\mathbb{C})$*, and the function $e^{i\theta} \mapsto R_{a,\theta}$ is an isomorphism between the circle group S^1 and the group $\mathcal{R}_a$.*

Proof: That $\mathcal{R}_a$ is a subgroup follows from reasoning similar to that of Example II.4.6. The last claim in the case $a = 0$ — the fact that $S^1 \cong \mathcal{R}_0$ — follows from Prop. I.4.10. To prove the claim for general a, it suffices to prove that the function $\phi : \mathcal{R}_0 \to \mathcal{R}_a$ given by $\phi(R_\eta) = R_{a,\eta}$ is an isomorphism. It is clearly a bijection. Given η_1, η_2, we have

$$\phi(R_{0,\eta_1} \circ R_{0,\eta_2}) = \phi(R_{0,\eta_1+\eta_2}) = R_{a,\eta_1+\eta_2}.$$

This in turn is equal to

$$R_{a,\eta_1} \circ R_{a,\eta_2} = \phi(R_{0,\eta_1}) \circ \phi(R_{0,\eta_2}).$$

■

Example II.4.8 (Reflection in a line generates a group) *Let L be a line in the plane and let M_L be reflection across the line L, defined in I.4.19. Then $\{id_{\mathbb{C}}, M_L\}$ is a subgroup of* $\mathrm{Isom}(\mathbb{C})$ *consisting of exactly two elements.*

The proof is left as part of Exercise 4. ■

Definition II.4.9 (The dihedral group D_n) *Suppose that n is a positive integer. Let $R = R_{\frac{2\pi}{n}}$, rotation about the origin by an angle of $\frac{2\pi}{n}$ radians. Let $M = M_0$ be reflection across the x-axis. The* dihedral group D_n *is defined as*

$$D_n = \{id_{\mathbb{C}} = R^0,\ R,\ R^2,\ ..., R^{n-1},\ M,\ RM,\ R^2M,\ ...,\ R^{n-1}M\}$$

The dihedral group will be central in our investigation of wallpaper groups in Chapter V. The following lemma expands on the definition, implies that D_n is indeed a group and geometrically specifies what group it is.

Lemma II.4.10 *Let the rotation R and the reflection M be as in the definition of the dihedral group D_n.*

1. *R^j is rotation about the origin by $\frac{2\pi j}{n}$ radians.*
2. *$R^j M = MR^{-j}$*
   $=$ reflection in the line through the origin at angle $\frac{\pi j}{n}$ with the x-axis.
3. *Suppose $n \geq 2$. Let P_n be the regular n-gon in the plane[10] with vertices $\{e^{2\pi ik/n} \mid k = 0, 1, ..., n-1\}$. Then D_n is the symmetry group of P_n:*
   $D_n = \mathrm{Sym}(P_n)$ (See Figure 9, Section I.7 for the case $n = 3$.)

Proof: Recall that $R_\eta(z) = e^{i\eta}z$ and $M(z) = \bar{z}$. We leave it to the reader (Exercise 5) to use this to prove i) and ii) and the fact that D_n is closed under composition and inverses.

To see that $D_n = P_n$, note that

$$R^j(e^{\frac{2\pi k}{n}}) = e^{\frac{2\pi(j+k)}{n}} \qquad \text{and} \qquad MR^j(e^{\frac{2\pi k}{n}}) = e^{-\frac{2\pi(j+k)}{n}}.$$

The result of this is that D_n takes the set of vertices of P_n onto itself, taking adjacent vertices (say) $e^{\frac{2\pi k}{n}}$, $e^{\frac{2\pi(k+1)}{n}}$, onto adjacent vertices $e^{\frac{2\pi k'}{n}}, e^{\frac{2\pi(k'\pm1)}{n}}$. (The sign is "+" for rotations, which preserve the clockwise order around the circle, and "−" for reflections, which reverse this order.)

Moreover, the line segment $[v, w]$ connecting two adjacent vertices v, w, is the set of all points $z \in \mathbb{C}$ such that $|v - z| + |z - w| = |v - w|$. (Compare Proposition I.2.19.) Since the elements of D_n are isometries, they thus take the line segments joining adjacent vertices to the line segments adjoining the images of these vertices. Hence, each element of D_n takes P_n onto itself, and we see that $D_n \subset \mathrm{Sym}(P_n)$.

But we claim also, that $\mathrm{Sym}(P_n) \subset D_n$. We prove this when $n \geq 3$ and leave it to the reader for $n = 2$. (Exercise 6.) With $n \geq 3$ we denote the vertices by

$$u_k = e^{\frac{2\pi ik}{n}}, \qquad n \in \mathbb{Z}.$$

[10] When $n = 2$ we take P_2 to be the interval $[-1, 1]$.

There are n vertices, $1 = u_0, u_1, \dots u_{n-1}$. It is useful to use this redundant notation, with $u_k = u_{k\pm n}$.

Now suppose $f \in \mathrm{Sym}(P_n)$. We use the fact (outlined in Exercise 7) that, since f is an isometry, f takes the set of vertices $\{u_0, u_1, \dots, u_n\}$ bijectively onto itself. Thus $f(1) = u_\ell$ for some integer ℓ. The two closest (and equally close) vertices to $u_0 = 1$ are u_{-1} and u_1. Again because f is an isometry, these vertices must go the the closest vertices to u_ℓ. Hence

$$f\left(\{u_0, u_{-1}, u_1\}\right) = \{u_\ell, u_{\ell-1}, u_{\ell+1}\}.$$

But these are sets of three distinct points, since $n \geq 3$. If $f(u_1) = u_{\ell+1}$ and $f(u_{-1}) = u_{\ell-1}$ then the same effect can be achieved on these points by the isometry $R^\ell \in D_n$. However three points determine an isometry! Thus $f = R^\ell \in D_n$.

Similarly, if $f(u_1) = u_{\ell-1}$ and $f(u_{-1} = u_{\ell+1}$ then the same effect can be achieved by $R^\ell M \in D_n$. So $f = R^\ell M$. Therefore $\mathrm{Sym}(P_n) \subset D_n$ as desired. ∎

D_n is not an abelian group if $n \geq 3$ (Exercise 8), although the cyclic group of rotations $G = \langle R \rangle$ is a subgroup of n elements comprising half of the dihedral group D_n.

Other transformation groups acting on the plane

The general linear group $GL_2(\mathbb{R})$ was defined (Example II.3.11) as the group of nonsingular (= invertible) 2×2 matrices with real entries. This group may be viewed as a group of permutations of the plane — nonsingular linear transformations — with the understanding that $A : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ is given by

$$A(\vec{z}) = A\vec{z} = A \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix} \quad \text{if } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \vec{z} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

The subgroup $\mathrm{O}_2 < GL_2(\mathbb{R})$ is then also viewed as a group of transformations of the plane, where each matrix A acts by this rule. Now recall Theorem I.6.4, which asserts that every isometry F of the plane may be written (in vector notation) in one and only one way as

$$(*) \qquad F(\vec{z}) = A\vec{z} + \vec{b} \quad \text{where} \quad A \in O_2,\ \vec{b} \in \mathbb{R}^2.$$

Since $F(0) = \vec{b}$, we see that $F(0) = 0 \iff F(z) = Az$ for all $z \in \mathbb{R}^2$. Thus the isometries F with $F(0) = 0$ are uniquely expressible as $F = A$ when the matrix A is thought of as a transformation. This shows us that O_2 is the subgroup of the transformation group Isom $\mathbb{C}$ consisting of those isometries F with $F(0) = 0$. This group is also called $\mathrm{Isom}_0(\mathbb{C})$. We summarize this discussion with the following lemma.

Lemma II.4.11 *If the matrices of the orthogonal group O_2 are identified with the linear transformations they determine, then $O_2 = Isom_0(\mathbb{C})$..*

If we follow the same pattern and consider transformations given by the equation (*) above, but we allow any matrix $A \in GL_2(\mathbb{R})$ then we get the *affine transformations.*

Definition II.4.12 *An* **affine transformation** *of the plane is a function $F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ which has a formula of the form*

$$F(\vec{z}) = A\vec{z} + \vec{b}, \qquad A \in GL_2(\mathbb{R}),\ \vec{b} \in \mathbb{R}^2$$

The **affine group**, *Aff_2, is the group of all affine transformations of the plane.*

We leave it as an exercise (Exercise 9) to verify that Aff_2 is a group and that $F^{-1}(\vec{z}) = A^{-1} \cdot (\vec{z} - \vec{b})$.

One last important group of transformations of the plane is the group of *similarities.* These transformations preserve angles, but not lengths (as when one talks of *similar triangles*).

Definition II.4.13 *A* **similarity** *is a function $f : \mathbb{C} \longrightarrow \mathbb{C}$ with the property that there exists a real number $\lambda \neq 0$ such that*

$$|f(z) - f(w)| = \lambda|z - w| \quad \textit{for all } z, w \in \mathbb{C}.$$

*The***similarity group** *Sim($\mathbb{C}$) is the group of all similarity transformations of $\mathbb{C}$.*

Note that: (Exercise 10)

- If f is a similarity using the real number $\lambda \neq 0$ and if the function m_λ (called a **dilation**) is given by $m_\lambda(z) = \lambda z$ for all $z \in \mathbb{C}$ then the transformation given by $g = m_\lambda^{-1} f$ is an isometry.
- The set of all similarities $\mathrm{Sim}(\mathbb{C})$ is a group under composition.
- We have the sequence of subgroups: $\mathrm{Isom}\ \mathbb{C} < \mathrm{Sim}(\mathbb{C}) < \mathrm{Aff}_2$.

Automorphism groups

Just as we study symmetries of a geometric object, we might wish to study the symmetries of a group - the permutations of the group which preserve its group structure. These are the *automorphisms* of the group.

Definition II.4.14 (Automorphism) *An* **automorphism** *of a group G is an isomorphism $\phi : G \longrightarrow G$ of G to itself*

Proposition II.4.15 *The set of all automorphisms of a group G form a group under the operation of composition. (This group is called the* **automorphism group** *of G and is denoted* $\mathrm{Aut}\ G$*.)*

The proof is left to the reader as Exercise 12. ∎

Example II.4.16 *If G is an abelian group then the function ϕ given by $\phi(g) = -g$ is an automorphism of G.*

Proof: ϕ is a bijection. (A fancy reason is that $\phi \circ \phi = \mathrm{id}_G$, so ϕ has both a right inverse and a left inverse – namely itself.) It is a homomorphism because, using commutativity[11]

$$\phi(a+b) = -(a+b) = (-b) + (-a) = (-a) + (-b) = \phi(a) + \phi(b). \qquad \blacksquare$$

[11] Without commutativity, all we can say in a group written multiplicatively is that , $(ab)^{-1} = b^{-1}a^{-1}$, or written additively, that $-(a+b) = (-b) + (-a)$.

Example II.4.17 *Aut $\mathbb{Z}$ is isomorphic to $\mathbb{Z}_2$.*

Proof: Let $\rho : \mathbb{Z} \longrightarrow \mathbb{Z}$ by $\rho(n) = -n$. This is an automorphism by the previous example. We claim that Aut $\mathbb{Z} = \{\mathrm{id}_G, \rho\}$ — that there can be no other automorphisms of $\mathbb{Z}$. This is because any homomorphism $\phi : \mathbb{Z} \longrightarrow \mathbb{Z}$ is determined by the image of 1:

$$\begin{aligned} \phi(\mathbb{Z}) &= \{\phi(n \cdot 1) \mid n \in \mathbb{Z}\} \\ &= \{n \cdot \phi(1) \mid n \in \mathbb{Z}\} \qquad \text{[since } \phi \text{ is a homomorphism]}. \end{aligned}$$

But then ϕ is one-one precisely when $\phi(1) \neq 0$. It is onto if and only if $\phi(1) = \pm 1$, Therefore we see that ϕ is an automorphism if and only if $\phi(n) = n$ for all n or $\phi(n) = -n$ for all n. So Aut $\mathbb{Z}$ contains exactly these two elements and is isomorphic to $\mathbb{Z}_2$. ∎

Cayley's Theorem

We have emphasized that the earliest examples of groups studied were transformation groups and that transformation groups are a central part of modern mathematics. But one can go even further. The following theorem shows that, in theory, *all* groups are really permutation groups.

Theorem II.4.18 (Cayley's Theorem)
Every group G is isomorphic to a permutation group.

Proof: We must show that G is isomorphic to a subgroup of S_X for some set X.

Now, G consists of a set with a binary operation on it. We take our set X to be the set of elements of the group G itself (!). We will write G to denote both the set X and the group.

To $g \in G$ (the *group*), we associate a permutation ℓ_g of G (the *set*) as follows:

$$\ell_g : G \longrightarrow G \text{ by the rule that } \ell_g(x) = gx \text{ for all } x \in G.$$

Claim 1: ℓ_g is a bijection of G.
It is one-to-one because

$$\ell_g(x) = \ell_g(y) \quad \Longrightarrow \quad gx = gy \quad \Longrightarrow \quad g^{-1}gx = g^{-1}gy \quad \Longrightarrow \quad x = y.$$

It is onto because given any $x \in G$, we may write

$$x = gg^{-1}x = \ell_g(g^{-1}x).$$

Define

$$\ell_G \;=\; \{\ell_g \;\big|\; g \in G\}.$$

By Claim 1, the *elements* of ℓ_G are *bijections* from G to itself.

Claim 2: ℓ_G is a subgroup of S_G.
For, by Claim 1, ℓ_G is a subset of S_G. The associative law holds for any three elements of ℓ_G because these are elements of the group S_G using the same operation of composition. The element $\ell_e \in \ell_G$ is equal to 1_G since $\ell_e(x) = ex = x$ for all $x \in G$. Hence ℓ_G has an identity element. ℓ_G is closed under composition because, for any g_1, $g_2 \in G$ we have

$$\ell_{g_1}\ell_{g_2}(x) \;=\; g_1 \cdot (g_2x) \;=\; g_1g_2x \;=\; \ell_{g_1g_2}(x). \tag{II.2}$$

So $\ell_{g_1}\ell_{g_2}$ is of the form ℓ_g with $g = g_1g_2$ and therefore $\ell_{g_1}\ell_{g_2} \in \ell_G$. Finally, ℓ_G is closed under taking inverses because (Exercise 6) $\ell_{g^{-1}} = \ell_{g^{-1}} \in \ell_G$.

Claim 3: The function $\phi : G \longrightarrow \ell_G$ given by $\phi(g) = \ell_g$ is an isomorphism.

The function ϕ is one-one because: $\phi(g) = \phi(h) \quad \Longrightarrow \quad \ell_g = \ell_h$. That is, ℓ_g and ℓ_h are the same *as functions*. Thus in particular, their values at the identity are the same: $\ell_g = \ell_h \;\Longrightarrow\; \ell_g(e) = \ell_h(e) \;\Longrightarrow\; ge = he \;\Longrightarrow\; g = h$.

ϕ is onto because given any $\alpha \in \ell_G$, we have by the definition of l_G that $\alpha = \ell_g$ for some $g \in G \;\Longrightarrow\; \alpha = \phi(g)$ for some $g \in G$.

Finally, by Equation II.2 in Claim 2,, $\phi(g_1)\phi(g_2) = \ell_{g_1}\ell_{g_2} = \ell_{g_1g_2} = \phi(g_1g_2)$

Therefore ϕ is an isomorphism from G to ℓ_G. ∎

Although in principle Cayley's Theorem reduces the study of any group to the study of a permutation group, in practice this is useless. One reason for this is that if G is finite and of order n, then S_G has order $n! = n(n-1)(n-2)...3 \cdot 2 \cdot 1$, which is very large even for modest values of n. More importantly, however, the process of viewing G as a subgroup of S_G obscures the origin of G, which is often as important as knowledge of G itself. For example, consider D_4. This group has eight elements. Thinking of D_4 as the symmetry group of a square allows us to much more easily understand this group than thinking of D_4 as a certain subgroup of order eight of the group S_G, whose elements are permutations of a set of size eight and which has $8! = 40,320$ elements.

Finite permutation groups, however, are important in their own right; we will discuss their structure in Section II.7.

Exercises II.4

1. Prove: If X is a non-empty set then S_X is a group.

2. Prove that $b \mapsto T_b$ is an isomorphism between $(\mathbb{C}, +)$ and $\mathcal{T}$, the group of all translations in Isom($\mathbb{C}$).

3. Prove that the set of all rotations $\mathcal{R}_a$ of Example II.4.7 is a group under the operation of composition. (Hint: Note that $R_{a,\theta} \circ R_{a,\eta} = R_{a,\theta+\eta}$.)

4. Prove that $\{\mathrm{id}_{\mathbb{C}},\ M_L\}$ and $\{\mathrm{id}_{\mathbb{C}},\ R_\pi\}$ are subgroups of Isom($\mathbb{C}$).
   Are these groups isomorphic? How do they compare with each other in their geometric behavior?

5. Prove parts 1. and 2. of Lemma II.4.10, and prove that D_n is a group.

6. Complete the proof that $\mathrm{Sym}(P_2) = D_2$ by proving that $\mathrm{Sym}(P_2) \subset D_2$. (**Hint:** Any isometry of $\mathbb{C}$ is determined by what it does to three non-collinear points. Concentrate on the points $\{1, -1, i\}$.)

7. We denote

$$\begin{aligned} u_k &= e^{\frac{2\pi i k}{n}} \qquad (n \in \mathbb{Z}) \\ v_k &= u_{k+\frac{n}{2}}, \quad (n \text{ even}) \\ w_k &= u_{k+\frac{n-1}{2}}, \quad z_k = u_{k+\frac{n+1}{2}} \quad (n \text{ odd}). \end{aligned}$$

Prove the following for the regular polygon P_n.

(a) If n is even then the maximal distance between all possible pairs of points $z, w \in P_n$ occurs and only occurs between pairs of diametrically opposed vertices u_k, v_k. (Note that $u_k = -v_k$ and the distance between them is 2.)

(b) If n is odd then the maximal distance between all possible pairs of points $z, w \in P_n$ occurs and only occurs between pairs of vertices of the form u_k, w_k and of the form u_k, z_k. (See Figure 9, Chapter 1.)

(c) Suppose $f \in P_n$. If n is even then f takes each set $\{u_k, v_k\}$ onto a set of the form $\{u_\ell, v_\ell\}$. If f is odd then f takes each set $\{u_k, w_k\}$ onto a set of the form $\{u_\ell, w_\ell\}$ or $\{u_\ell, z_\ell\}$. In any case, f takes each vertex u_k to another vertex u_ℓ.

(d) f takes the set of vertices of P_n bijectively onto itself.

8. Show that if $n \geq 3$, then D_n is non-abelian, while if $n = 2$, then $D_n \cong V_4$, the Klein four-group (Example II.3.9). (Note that as a set, $D_2 = \{\mathrm{id}_{\mathbb{C}}, R_\pi, M_0, M_{\pi/2}\}$.)

9. Prove that Aff_2 is a group. If $F \in \mathrm{Aff}_2$ is given by $F(\vec{z}) = A\vec{z} + \vec{b}$, with $A \in GL_2(\mathbb{R})$, how is the fact that $A \in GL_2(\mathbb{R})$ used in this proof?

10. (a) If f is a similarity using the real number $\lambda \neq 0$ and if the function m_λ (called a **dilation**) is given by $m_\lambda(z) = \lambda z$ for all $z \in \mathbb{C}$ then the transformation given by $g = m_\lambda^{-1} f$ is an isometry.

(b) The set of all similarities $\mathrm{Sim}(\mathbb{C})$ is a group under composition.

(c) We have the sequence of subgroups: $\mathrm{Isom}\ \mathbb{C} < \mathrm{Sim}(\mathbb{C}) < \mathrm{Aff}_2$.

11. Show that the set G of functions $\{f^n : \mathbb{C} \to \mathbb{C}\}$ given by $f^n(z) = 2^n z$ $(n \in \mathbb{Z})$ is a subgroup of $\mathrm{Sim}(\mathbb{C})$.

12. Prove that the automorphism group $\mathrm{Aut}\ G$ of a group G is indeed a group.

13. Suppose that $G = \langle g \rangle$ is a cyclic group with generator g.

(a) If ϕ is an automorphism of G and $\phi(g) = h$, prove that $G = \langle h \rangle$ (*i.e.*, that h is a generator of G.

(b) Conversely, suppose that $h \in G$ is a generator of G. Prove that the function $\phi : G \longrightarrow G$ given by

$$\phi(g^j) = h^j, \quad \text{for all } j \in \mathbb{Z}$$

is an automorphism of G. (Note: The proof should start by showing that ϕ is a well-defined function; *i.e.*, $g^j = g^k \implies h^j = h^k$.)

14. Find $\text{Aut } \mathbb{Z}_n$, when $n = 2, 3, 4, \text{ or } 5$.

15. Show that, in the proof of Cayley's Theorem (Thm. II.4.18), $(\ell_g)^{-1} = \ell_{g^{-1}}$.

II.5 The dynamics of a group action

Transformation groups are best pictured in terms of their *dynamics* —how the group elements do and do not move points and subsets around. Here, we give the vocabulary which captures some key dynamical concepts. We use the following transformation groups $G_1,\ G_2,\ G_3$, each acting on $\mathbb{C}$, to illustrate this vocabulary.

- $G_1 = \{T_b \mid b \in \mathbb{R}\}$ = the group of all translations in horizontal directions.
- $G_2 = \{\mathrm{id}_{\mathbb{C}}, M\}$, where M is reflection across the x-axis: $M(z) = \overline{z}$.
- $G_3 = \{f_{\theta,n} \mid \theta \in \mathbb{R},\ n \in \mathbb{Z}\}$, where $f_{\theta,n} : \mathbb{C} \longrightarrow \mathbb{C}$ is defined by $f_{\theta,n}(z) = 2^n R_\theta(z)$ where $R_\theta(z) = e^{i\theta}(z)$

The groups G_1 and G_2 are subgroups of $\mathrm{Isom}(\mathbb{C})$. G_3 is not, since distances are distorted. In particular, $f_{\theta,n}$ dilates the plane by a factor of 2^n about the origin and rotates about the origin by an angle θ. To see that G_3 is indeed a group notice first that the closure axiom is satisfied since

$$f_{\theta,n} \circ f_{\psi,m} = f_{\theta+\psi,n+m}.$$

The function $f_{0,0} = \mathrm{id}_{\mathbb{C}} \in G_3$ is the identity. Also, $f_{\theta,n}^{-1} = f_{-\theta,-n}$, so that the inverse of $f \in G_3$ is again in G_3. Finally, associativity follows since, in general, composition of functions is associative.

We suppose in the definitions below that we are given a group G and a non-empty set X on which it acts.

Definition II.5.1 (Fixed point sets)

1. *For $g \in G$, the* **fixed point set of** *g, denoted* $\mathrm{Fix}(g)$*, is the subset of X defined by*
$$\mathrm{Fix}(g) \equiv \{x \in X \mid g(x) = x\} \subset X.$$

2. *For a subgroup H of G, the* **fixed point set of** H, *denoted* $\mathrm{Fix}(H)$, *is the subset of X defined by*

$$\mathrm{Fix}(H) \equiv \{x \in X \mid g(x) = x \text{ for all } g \in H\} \subset X.$$

Thus, the *fixed point set of an element g* is all points in X which are not moved by g, and the *fixed point set of H* is all points in X which are are unmoved by *all* elements of H.

For example:

- In any transformation group G acting on a set X, we have $\mathrm{Fix}(\mathrm{id}_X) = X$.
- For G_1, which consists of horizontal translations, we have $\mathrm{Fix}(g) = \emptyset$ for all $g \in G_1$ with $g \neq \mathrm{id}_{\mathbb{C}}$ and $\mathrm{Fix}(G_1) = \emptyset$. (In fact, notice (Exercise 4.3c)) that, if $g \in H \subset G$ then $\mathrm{Fix}(G) \subset \mathrm{Fix}(H) \subset \mathrm{Fix}(g)$.)
- For G_2, the group generated by a single reflection in the x-axis, $\mathrm{Fix}(M) = \mathbb{R}$ and $\mathrm{Fix}(G_2) = \mathbb{R}$.
- For G_3, which consists of certain scalings and rotations fixing the origin, $\mathrm{Fix}(f_{\theta,n}) = \{0\}$ for all $f_{\theta,n} \neq \mathrm{id}_{\mathbb{C}}$, and then $\mathrm{Fix}(G_3) = \{0\}$.

Definition II.5.2 (Stabilizer) *For an element $x \in X$, the* **stabilizer of x in G,** *denoted* $\mathrm{Stab}_G(x)$, *is the subgroup of G defined by*

$$\mathrm{Stab}_G(x) \equiv \{g \in G \mid g(x) = x\} \subset G.$$

For example, the actions of G_1, G_2, G_3 above give

$\mathrm{Stab}_{G_1}(0) = \{\mathrm{id}_{\mathbb{C}}\}$	$\mathrm{Stab}_{G_2}(0) = G_2$	$\mathrm{Stab}_{G_3}(0) = G_3$
$\mathrm{Stab}_{G_1}(3) = \{\mathrm{id}_{\mathbb{C}}\}$	$\mathrm{Stab}_{G_2}(3) = G_2$	$\mathrm{Stab}_{G_3}(3) = \{\mathrm{id}_{\mathbb{C}}\}$.

So the fixed-point set is a *subset* of X, whereas the stabilizer is a *subgroup* of G. In Exercise 3a, the reader is asked to show that $\mathrm{Stab}_G(x)$ is indeed a subgroup of G.

Definition II.5.3 (Orbit) *For $x \in X$, the* **orbit of x under G,** *denoted by* $\mathrm{Orbit}_G(x)$ *or $G \cdot x$, is the subset of X defined by*

$$\mathrm{Orbit}_G(x) \;=\; G \cdot x \;\equiv\; \{g(x) \mid g \in G\} \subset X.$$

In other words, the orbit of a point is all points in X to which you can get by starting from X and applying an element of G. (We often write $\mathrm{Orbit}(x)$ instead of $\mathrm{Orbit}_G(x)$ when the group G is understood.)

For example:

$$\begin{aligned}
\mathrm{Orbit}_{G_1}(i) &= \{x + i \mid x \in \mathbb{R}\},\\
\mathrm{Orbit}_{G_2}(i) &= \{i, -i\},\\
\mathrm{Orbit}_{G_3}(i) &= \bigcup_{n\in\mathbb{Z}} \text{circle of radius } 2^n \text{ about } 0.
\end{aligned}$$

The last statement is true because a point z is in this of circles if and only if it can be written as $z = 2^n e^{i\alpha}$ for some $n \in \mathbb{Z}$ and $\alpha \in \mathbb{R}$. This means

$$z = 2^n e^{i(\alpha - \frac{\pi}{2})} e^{i\frac{\pi}{2}} = 2^n e^{i(\alpha-\frac{\pi}{2})} i = f_{\theta,n}(i) \quad \text{where } \theta = \alpha - \frac{\pi}{2}.$$

Notice that this orbit is the figure drawn in Figure 5 of Section I.1.

Definition II.5.4 (Free action) *The action of G on X is called* free *if, for all $g \in G$ not equal to the identity,* $\mathrm{Fix}(g) = \emptyset$.

In other words, G acts freely on X if every element of G, other than the identity, moves every point of X. For example, the action of G_1 on $\mathbb{C}$ above is free, but the actions of G_2 and G_3 are not. The adjective "free" should be thought of as "fixed-point free".

The property of acting freely is a property of the *pair* G and X, taken together.

The following proposition illustrates how the concepts introduced can interact in the simple situation of a free action. It also illustrates some points of logic which are

useful to have at one's command. Above the list of statements in the proposition, the phrase "the following are equivalent" means that any one of the statements implies any other one. Notice, then, that to prove this, it is enough to prove that (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (1). Tactically, the difficulty of doing this will usually depend on the order in which the statements are listed. See Appendix A for further discussion.

Proposition II.5.5 (Characterizations of free actions) *Suppose that the group G acts on the set X. Then the following are equivalent:*

1. *The group acts freely on X.*
2. *For each $x \in X$, $\mathrm{Stab}_G(x) = \{id_X\}$.*
3. *For each $g \in G$ with $g \neq id_X$, $\mathrm{Fix}(g) = \emptyset$.*
4. *For each $x \in X$, the function $G \longrightarrow G \cdot x = \mathrm{Orbit}_G(x)$, given by the rule $g \mapsto g(x) \in \mathrm{Orbit}(x)$, is a bijection.*

Proof: We prove that (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (1).

The fact that (1) $\Longrightarrow$ (2) and that (2) $\Longrightarrow$ (3) are direct exercises in the definitions and are left to the reader (Exercise 6).

(3) $\Longrightarrow$ (4): Let $E_x : G \longrightarrow G \cdot x$ be the function given in (4), with $E_x(g) = g(x)$. (E_x is called the "evaluation map": we evaluate each g on x to get the image of g under E_x.) We must show that E_x is one-one and onto. It is certainly onto since each element y of the orbit of x is of the form $y = g(x) = E_x(g)$ for some $g \in G$.

To see that E_x is one-one, suppose that $E_x(g) = E_x(h)$. Then $g(x) = h(x)$, so that $g^{-1}h(x) = x$. Thus $x \in \mathrm{Fix}(g^{-1}h)$. By (3) this fixed point set can only be non-empty if $g^{-1}h = \mathrm{id}$. Hence $g = h$, and we see that E_x is one-one.

(4) $\Longrightarrow$ (1): Suppose that $x \in X$ and $g \in G$ with $g \neq \mathrm{id}$. Since, by (4), E_X is one-one, $E_x(g) \neq E_x(\mathrm{id})$. In other words, $g(x) \neq \mathrm{id}(x) = x$. Hence the action of G on X is free. ∎

Corollary II.5.6 (Orbits of a free action) *If the group G acts freely on X then all orbits have the same cardinality*[12] *— namely the cardinality of the group G.*

Proof: By Proposition II.5.5, there is a bijection of the group G to each orbit $G \cdot x$ given by given by $g \mapsto g(x)$. Then the composition $g(x) \mapsto g \mapsto g(y)$ gives a bijection of the orbit $G \cdot x$ to the orbit $G \cdot y$. Hence G, $G \cdot x$ and $G \cdot y$ all have the same cardinality. ∎

Example II.5.7 (Free action of a subgroup on the big group)

An important example of a free action comes when the set X is the underlying set of a group G. (We will write G to denote both the set X and the group.) Suppose that H is a subgroup of G. As in the proof of Cayley's Theorem (Thm. II.4.18), for each $h \in H$ we let ℓ_h be the bijection of G which is given by left multiplication:

$$\ell_h : G \longrightarrow G \text{ by the rule that } \ell_h(x) = hx \text{ for all } x \in G.$$

Let $\ell_H = \{\ell_h \mid h \in H\}$.

As in the proof of Cayley's Theorem, ℓ_H is a group of bijections of G. In this situation, the action of ℓ_H on G is free. To see this, suppose that $x \in G$ and $\ell_h \neq 1_G$. Then

$$\ell_h \neq 1_G \implies h \neq e \implies hx \neq x \implies \ell_h(x) \neq x,$$

proving that the action is free.

Notice that the group ℓ_H is actually isomorphic to H. (Exercise 7.)

The partition of X into G-orbits

Finally, we wish to point out how the orbits given by the action of G on X break X into pieces. In our three examples with $X = \mathbb{C}$, the actions give the following decompositions of the plane into orbits:

[12] By definition, two sets have the same cardinality if there is a bijection from one to the other.

- G_1: the plane is the union of the family of all parallel lines;
- G_2: the plane is the union of the family of sets $\{ \{z, \overline{z}\} \mid z \notin \mathbb{R}\} \cup \{ \{z\} \mid z \in \mathbb{R}\}$;
- G_3: the plane is the union of the family of all circles about 0, and the set $\{0\}$ itself.

These decompositions are examples of *partitions* (see Definition C.3), wherein a set X is decomposed into a family of disjoint non-empty subsets:

Proposition II.5.8 (Orbits form a partition) *If the group G acts on the set X then the family $\mathcal{F}$ of orbits under this action, $\mathcal{F} = \{ G \cdot x \mid x \in X \}$, is a partition of the set X.*

Proof: Each element $x \in X$ belongs to an orbit. Indeed, $x = \mathrm{id}_X(x) \in G \cdot x$. Thus the union of the orbits is X.

If a single element $h(x)$ in the orbit of x were in the orbit $G \cdot y$ of y, we would have $h(x) = k(y)$ for some $k \in G$. It would then follow for *every* $g \in G$ that

$$g(x) = g\left(h^{-1}k(y)\right) = \left(g \circ h^{-1} \circ k\right)(y) \in G \cdot y.$$

Thus the entire orbit $G \cdot x \subset G.y$. By the same reasoning, $G \cdot y \subset G \cdot x$. Therefore the two orbits would be equal if they had a single element in common. If the orbits are not equal we conclude that they are disjoint. Thus the orbits form a partition of X, as claimed. ∎

Exercises II.5

1. Let $G = \mathcal{R}_0 = \{R_{0,\theta} \mid \theta \in \mathbb{R}\} =$ the group of all rotations of $\mathbb{C}$ about the origin.
   (a) For each $z \in \mathbb{C}$, find $\mathrm{Stab}_G(z)$.
   (b) For each $g \in \mathcal{R}_0$, find $\mathrm{Fix}(g)$.
   (c) For each $z \in \mathbb{C}$, find the orbit of z.
   Show that the set of orbits forms a partition of $\mathbb{C}$.

2. If $t \in \mathbb{R}$ let $f_t : \mathbb{C} - \{0\} \longrightarrow \mathbb{C} - \{0\}$ by $f_t(z) = \frac{e^{2\pi i t}}{2^t} z$.
Let $G = \{f_t \mid t \in \mathbb{R}\}$.

   (a) Show that G is a transformation group.

   (b) Prove that G acts freely on $\mathbb{C} - \{0\}$.

   (c) Draw the orbit of $z = 3$.
   Label the images of $f_t(3)$ on this orbit for $t = -2, -1, 0, 1, 2$.

3. Suppose that G acts on the set X. Prove the following statements:

   (a) If $x \in X$ then $\mathrm{Stab}_G(x)$ is a subgroup of G.

   (b) If $x \in X$ and $g \in G$ then $\mathrm{Orbit}_G(x) = \mathrm{Orbit}_G(g(x))$.

   (c) If H is a subgroup of G then $\mathrm{Fix}(H) \supset \mathrm{Fix}(G)$. This follows from the fact that
   $$\mathrm{Fix}(H) = \bigcap \{\mathrm{Fix}(g) | g \in H\}.$$

   (d) If $g \in G,\ x \in X$ then
   $$\mathrm{Stab}_G(g(x)) = g\,\mathrm{Stab}_G(x)\,g^{-1} \overset{def.}{=} \{ghg^{-1} \mid h \in \mathrm{Stab}_G(x)\}.$$

4. In each of the following cases a group G and a subgroup $H < G$ are given. In each case, list the orbits under ℓ_H, where ℓ_H is defined as in Example II.5.7.

   (a) $G = \mathbb{Z}$ and $H = \{\, 6k \,|\, k \in \mathbb{Z}\}$.

   (b) $G = \mathbb{C}$ (as an additive group) and $H = \{t(1 + 3i) \mid t \in \mathbb{R}\}$.

   Note: The orbits under ℓ_H are called the **right cosets** of H in G. (See Definition II.7.5.)

5. Suppose that the finite group H of k elements acts freely on the finite set X of n elements. How many orbits are there under this action?

6. Prove that (1) $\Longrightarrow$ (2) and that (2) $\Longrightarrow$ (3) in Proposition II.5.5.

7. Prove that $\ell_H \cong H$ in Example II.5.7.
Hint: Compare this with Claim 3 in the proof of Cayley's Theorem, II.4.18.

II.6 How do we recognize and generate subgroups?

Recall (Definition II.1.6) that H is a *subgroup* of the group G if $H \subset G$ and if H itself becomes a group when the elements of H are combined by the same rule used to combine them as elements of G. For example, recalling the groups $\mathbb{Z}_n$ defined in Example II.1.8, we have

$$H \;=\; \{0, 2, 4\} \;\subset\; \mathbb{Z}_6 \;=\; \{\; 0, 1, 2, 3, 4, 5 \;\}$$

and we can check that $(H, +)$ satisfies the axioms of a group if we define "+" in H to have the same values as "$+_6$" in $\mathbb{Z}_6$:

$$2 + 2 = 4, \qquad 2 + 4 = 4 + 2 = 0, \qquad 0 + 2 = 2 + 0 = 0, \qquad 4 + 4 = 2, \;\text{ etc.}$$

Thus H is a subgroup of $\mathbb{Z}_6$. However $K = \{0, 1, 3, 4, 5\}$ is not a subgroup of $\mathbb{Z}_6$ because $3,\, 5 \in K$ while $3 + 5 = 2 \notin K$. Thus $(K, +)$ does not satisfy the closure axiom for a group. (Alternatively, K is not a group because 4 has no inverse in K, the identity element of K being 0.)

We now ask:

- *What do we know automatically about a non-empty subset H of a given group G if multiplication in H is defined using the binary operation from G?*
- *What further do we need to check in order to conclude that H is a subgroup of G?*
- *If H is not a subgroup of G, what is the smallest subgroup of G which contains H? (Is there a smallest subgroup of G which contains H?)*

What we know automatically about the subset H

Lemma II.6.1 *Multiplication is associative: if $a, b, c \in H$ then $a(bc) = (ab)c$.*

Proof: a, b, c are elements of G since $H \subset G$. The associative law holds in G, and the rule of multiplication in H is just the rule in G applied to the elements of H. So the answers are the same (though the answers may end up in G outside of H). ∎

Lemma II.6.2 (Identity and inverses coincide) *1. If H has a multiplicative identity element e_H then $e_H = e$, and in particular $e \in H$, where e is the identity element of the group G.*

2. If H has a multiplicative identity element e_H and if $a \in H$ has an inverse $\bar{a}$ in H (i.e., $a\bar{a} = e_H$), then $\bar{a} = a^{-1}$, where a^{-1} is the inverse of a in G.

Proof: To prove (1), let $a \in H$. Then $ae_H = a$, since e_H is an identity element in H. But a and e_H are elements of G. Thus

$$e_H = ee_H = (a^{-1}a)e_H = a^{-1}(ae_H) = a^{-1}a = e.$$

To prove (2), suppose that $a \in H$ and $\bar{a} \in H$ with $a\bar{a} = e_H$. Since part (a) tells us that $e_H = e$, we have $a\bar{a} = e$. Thus $\bar{a} = a^{-1}$, by uniqueness of inverses in G (Lemma II.1.3(2))). ∎

How do we check whether H is a subgroup of G?

Just by virtue of being a subset of G with the inherited operation, we have seen above that we already have substantial information about H. The result is that we don't have to check all of the group axioms in order to know that H is a group.

Lemma II.6.3 (Conditions for a subgroup) *If H is a nonempty subset of the group G with the operation inherited from G, then H is a subgroup of G $\iff$*
(i) *H is closed under the operation, i.e.: $a, b \in H \implies ab \in H$; and*
(ii) *H is closed under the taking of inverses, i.e.: $a \in H \implies a^{-1} \in H$ (where a^{-1} is the inverse of a in G).*

Proof: $\implies$: If H is a subgroup, then it must satisfy the axioms for a group. Thus it satisfies the Closure axiom (i). Also, each element $a \in H$ must have an inverse in H. By Lemma II.6.2, this is then the inverse of a in G, and we see that H also satisfies (ii).

$\impliedby$: We check that H is a subgroup if we assume (i) and (ii):
H satisfies the Closure axiom by hypothesis (i) and H satisfies the Associative axiom

by Lemma II.6.1. Let $a \in H$. (Remember we are assuming that H is nonempty.[13]) By hypothesis (ii), $a^{-1} \in H$. By hypothesis (i) we then have that $e = aa^{-1} \in H$, so H satisfies the Identity axiom. Finally, H satisfies the Inverses axiom by assumption (ii). ■

When the subset H is finite, the situation iseven simpler. We need only check closure:

Lemma II.6.4 (Conditions for a finite subgroup) *If H is a nonempty finite subset of the group G with the operation inherited from G, then H is a subgroup of G $\iff$ H is closed under the operation: $a, b \in H \implies ab \in H$.*

Note: If the hypothesis that H is finite is dropped, however, there are counterexamples (Exercise 7).

Proof: Closure is clearly necessary for H to be a group. To prove that it is sufficient, we need, by Lemma II.6.3, only prove that H is closed under the taking of inverses. As a first step in proving this we will show that $e \in H$.

Let $H = \{a_1, a_2, \ldots, a_n\}$. If we multiply every element in G on the left by a_1 then we get a bijection (Example II.5.7) $\ell_{a_1} : G \longrightarrow G$. By our assumption that H is closed under multiplication, we have $\ell_{a_1}(H) \subset H$. So ℓ_{a_1} maps H into H in a one-one manner. But, by hypothesis, H is finite! Hence ℓ_{a_1} must map H *onto* H. (Exercise 2). Therefore

$$H = \{\ a_1a_1,\ a_1a_2,\ \ldots,\ a_1a_n\ \}.$$

Since $a_1 \in H$ we must then have for some i that $a_1 = a_1a_i$. It follows then that $a_i = e$, and we have shown that $e \in H$, as desired.

Now, to show that H is closed under the taking of inverses, consider any element $a_j \in H$. Multiplying each element of H on the left by a_j we get a bijection of H with itself as above. But we know that $e \in H$. Hence, for some k, $a_ja_k = e$. Then $a_k = a_j^{-1}$, and we have shown that for each element $a_j \in H$, its inverse is also in H.

[13]When reading a proof, and especially at the end of reading the proof, you can check your understanding by asking yourself where each hypothesis was used. This insures that the proof is actively grasped, rather than passively accepted.

■

The smallest subgroup of G containing a subset S

If S is a nonempty subset of the group G which is not itself a group, then, according to Lemma II.6.3, S fails to be a group because it is "missing some elements". Either there is a pair of elements of S whose product is not in S, or there is an element of S whose inverse is not in S. Can we add some elements of G to S to get a subgroup H of G containing S? The process might be delicate, because if we add a missing product ab where $a \in S,\ b \in S$, we might then find that the inverse of the new element we added $(ab)^{-1} = b^{-1}a^{-1}$ is missing. Moreover if we add this inverse to our set, we then have to worry about whether its product with the things already in the set is in the set. Etc. Will this ever end? Can we find a subgroup of G containing S at all?

Of course the answer to this is "yes", since G itself is a subgroup containing S. But can we be more efficient in filling out S to get a group? For example, if we take the subset S of the group $\mathbb{Z}$ (with the operation of addition) given by

$$S \ = \ \{6,\ -6,\ 9,\ -9, \dots \} \ = \ \{\ 3n |\ n \in \mathbb{Z},\ n > 1\}$$

then we need only add the elements $0,\ 3,\ -3$ to S to get a group, and the process does stop with a proper subgroup of $\mathbb{Z}$:

$$S \ \subset \ S \cup \{0\ , 3\ , -3\} \ = \ H \ = \ \{\ 3n \mid n \in \mathbb{Z}\ \} \ \subset \ \mathbb{Z}.$$

In the following discussion we show that there is in fact a smallest subgroup of G containing a given subset S. Two useful descriptions of this group are given in II.6.9 II.6.6 below.

Lemma II.6.5 (Intersection of subgroups is subgroup) *The intersection of any family of subgroups of a group G is itself a subgroup of G.*

Proof: Let $\{\ H_\alpha \mid \alpha \in A\ \}$ be any family of subgroups of G, indexed, say, by a set A. Let

$$H_0 \ = \ \bigcap_{\alpha \in A} H_\alpha.$$

We shall use Lemma II.6.3 to show that H_0 is a subgroup of G.

H_0 is nonempty because $e \in H_\alpha$ for all $\alpha \in A$, and so e is an element of the intersection H_0 of these sets.

If a and b are elements of H_0, then $a, b \in H_\alpha$ for each $\alpha \in A$. So, H_α being a group, $ab \in H_\alpha$ for each $\alpha \in A$. Hence $ab \in H_0$, and H_0 satisfies the Closure axiom.

Similarly, if $a \in H_0$ then $a \in H_\alpha$ for each $\alpha \in A$. Then, since H_α is a subgroup of G, we have $a^{-1} \in H_\alpha$ for each $\alpha \in A$. Hence $a^{-1} \in H_0$, and H_0 satisfies the Inverses axiom.

■

Corollary II.6.6 (Subgroup generated by a set) *If S is a subset of the group G then there is a unique subgroup $\langle S \rangle$ of G such that*

$$S \subset \langle S \rangle \subset K$$

for every subgroup K of G which contains the set S.

Proof: Set

$$\langle S \rangle = \bigcap \{H_\alpha \mid H_\alpha \text{ is a subgroup of } G \text{ containing } S\}.$$

Clearly $\langle S \rangle \supset S$ since each $H_\alpha \supset S$. Further, $\langle S \rangle$ is a subgroup by Lemma II.6.5. If K is any subgroup of G which contains S then K is one of the H_α. Hence $K \supset \langle S \rangle$.

To see that the subgroup $\langle S \rangle$ is unique, suppose another subgroup H also contains S and is contained in any subgroup containing S. Then $\langle S \rangle \subset H$ since H is a subgroup containing S and, similarly, $H \subset \langle S \rangle$. Hence $H = \langle S \rangle$, proving uniqueness.

■

Definition II.6.7 (Subgroup generated by a set) *If $\langle S \rangle$ is the unique subgroup given in Corollary II.6.6, we define*

$$\langle S \rangle \ = \ \textbf{the subgroup of } G \textbf{ generated by } S.$$

When $S = \{g_1, g_2, \ldots, g_n\}$ is finite , then we refer to $\langle S \rangle$ simply as

$$\langle g_1, g_2, \ldots, g_n \rangle \ = \ \textbf{the subgroup of } G \textbf{ generated by } g_1, g_2, \ldots, g_n.$$

Definition II.6.8 (Generating set) *If $\langle S \rangle = G$ (or $\langle g_1, g_2, \ldots, g_n \rangle = G$) we say that S* **generates** G **(or** $g_1, g_2, \ldots, g_n$ **generate** G**)**

Corollary II.6.6 is an *existence statement* — it tells us that a certain subgroup exists — and, theoretically, the proof tells us how to find this subgroup. (Just take the intersection of all the subgroups containing S.) But, in practice, this does not tell us which elements are in $\langle S \rangle$. The following Proposition supplies this information. It also clears up a possible ambiguity: In §II.2 we gave a different definition of $\langle g \rangle$. It was defined as the set of all powers of g. In fact we show that this is the same as the smallest subgroup containing $\{g\}$.

Proposition II.6.9 (Generators for subgroups)

1. *If G is a group and $g \in G$ then*

   $$\langle \{g\} \rangle \ = \ \{g^n \ \big| \ g \in \mathbb{Z}\}$$ (≡ "the cyclic subgroup generated by g" – see II.3.7).

2. *For an arbitrary nonempty subset S of G,*

   $$\langle S \rangle \ = \ \{ \ g_1^{n_1} g_2^{n_2} \cdots g_q^{n_q} \ \big| \ q \geq 1, \ \ g_i \in S, \ \ n_i \in \mathbb{Z} \ \ (i = 1, \ldots, q) \ \}$$

Remark: In the light of this Proposition, the intuitive meaning of the statement that *"S generates G"* is that *G is the set of all possible finite products of elements of S and their inverses.* A warning is in order: though the set of *symbols* inside the

set brackets on the right-hand side of the expression for $\langle S \rangle$ is infinite, it is entirely possible that the resulting set of *elements* of G is finite. We've already encountered this in the case of finite cyclic groups; see also the Examples at the end of this section.

Proof: 1. Since $\langle g \rangle$ is a subgroup of G, it must contain the identity element $e = g^0$. Since $\langle g \rangle$ satisfies the Closure axiom and contains g, it must, by mathematical induction, then contain $g \cdot g \cdots g = g^n$ for any integer $n > 0$. As a subgroup, it must also then contain $(g^n)^{-1} = g^{-n}$ for any $n > 0$. Thus $\langle g \rangle$ contains the cyclic subgroup generated by g. On the other hand (using Proposition II.3.7 or Lemma II.6.3 above) the latter *is* a subgroup containing g, and thus it must contain $\langle g \rangle$. Therefore $\langle g \rangle$ is precisely equal to the cyclic subgroup generated by g.

2. This is a slight elaboration of the proof of 1. We leave this to the reader (Exercise 5). ∎

Proposition II.6.9 is often used to construct examples in conjunction with the following simple lemma.

Lemma II.6.10 *1. If A and B are subsets of the group G such that $\langle A \rangle \supset B$ then $\langle A \rangle \supset \langle B \rangle$.*

2. If B generates G and $\langle A \rangle \supset B$ then A generates G.

Proof: 1. $\langle B \rangle$ is contained in any group which contains B, and $\langle A \rangle$ is given to be such a group. So $\langle B \rangle \subset \langle A \rangle$.

2. We are given that $\langle A \rangle \supset B$ and B generates G. Therefore by a), $\langle A \rangle \supset \langle B \rangle = G$. But $\langle A \rangle$ is a subgroup of G. Therefore $\langle A \rangle = G$. ∎

Example II.6.11 *If $m_1, m_2, \ldots m_q$ are non-zero integers and $d = g.c.d.\{m_1, m_2, \ldots m_q\}$. (i.e., d is the greatest positive integer which divides all of the m_i.). Then*

- *$\langle m_1, m_2 \ldots m_q \rangle$ is the cyclic subgroup of $\mathbb{Z}$ generated by d.*

- $d = a_1 m_1 + \ldots a_q m_q$ *for some integers* $a_1, a_2, \ldots a_q$.

Proof: Let $H = \langle\, m_1, m_2, \ldots m_q \,\rangle$. In the next section we prove that every subgroup of a cyclic group is cyclic (see Theorem II.7.1) . So $H = \langle\, d_0 \,\rangle$ for some $d_0 \in H$. We may assume that d_0 is positive.

On the other hand, by Proposition II.6.9 (written in additive notation) we have

$$H = \langle\, m_1, m_2, \ldots m_q \,\rangle = \{a_1 m_1 + a_2 m_2 + \ldots a_q m_q \mid a_i \in \mathbb{Z}\}.$$

Since the greatest common divisor d divides each m_i, it divides each element of H. In particular d divides d_0. But $d_0 \leq d$. This is because d_0 is a common divisor of the m_i since $m_i \in \langle\, d_0 \,\rangle = H$ and d is the greatest common divisor of the m_i. Hence $d = d_0 \in H$ and both claims follow immediately. ∎

When we have just two integers and they are relatively prime (*i.e.,* their greatest common divisor is 1), then, since $\langle\, 1 \,\rangle = \mathbb{Z}$, the previous example becomes

Example II.6.12 (Relatively prime integers)
Suppose that $m, n \in \mathbb{Z}$ *are relatively prime integers. Then*

- $\mathbb{Z} = \langle\, m, n \,\rangle$.
- *There are integers* a, b *such that* $1 = am + bn$. ∎

Example II.6.13 (Generators for $\mathbb{Z}_n$)
If $m \in \mathbb{Z}_n$ *and* m *is relatively prime to* n*, then* m *generates* $\mathbb{Z}_n$.

Proof: As just noted, there exist integers a and b such that $am + bn = 1$. Thus

$$1 \in \langle\, m \,\rangle = \{am \in \mathbb{Z}_n \;\big|\; a \in \mathbb{Z}\}.$$

Since 1 generates $\mathbb{Z}_n$, Lemma II.6.10 implies that m generates $\mathbb{Z}_n$. ∎

Example II.6.14 (Generators for D_n) *The dihedral group*

$$D_n \;=\; \langle R,\; M \;\rangle$$

where R is rotation by $2\pi/n$ about the origin and M is reflection in the x-axis.

Proof:

In Proposition II.6.9, take $S = \{R, M\}$ and $G = \mathrm{Isom}(\mathbb{C})$. Recall that

$$D_n \;=\; \{\mathrm{id}_{\mathbb{C}} = R^0,\; R,\; R^2,\; \ldots R^{n-1},\; M,\; RM,\; R^2M,\; \ldots,\; R^{n-1}M\}.$$

So on the one hand, by part 2 of Proposition II.6.9, $D_n \subset \langle R, M\rangle$. On the other hand, D_n is a *subgroup* of $G = \mathrm{Isom}(\mathbb{C})$ containing $S = \{R, M\}$. So by Corollary II.6.6, $D_n \supset \langle R, M\rangle$.

■

The group D_n has many other generating sets. Here are three of them:

1. If m is relatively prime to n then (as in Example II.6.13), then $(R^m)^a = R^{am} = R^1$ for some $a \in \mathbb{Z}$. Hence $\langle\{R^m,\; M\}\rangle$ contains the generating set $\{R,\; M\}$. Therefore the set $\{R^m,\; M\}$ also generates D_n. So D_n is also generated by the set of two isometries: rotation by $\frac{2\pi m}{n}$ radians and reflection across the x-axis.

2. Also, the group $\langle\{R^m,\; R^jM\}\rangle$ contains the generating set $\{R^m,\; M\}$ for any m relatively prime to n and any integer j. This is because, for some $a \in \mathbb{Z}$, $(R^m)^a = R^1 = R$, so that

   $$M \;=\; MR^{-j}R^j \;=\; (R^jM)^{-1}(R^{am})^j \;=\; (R^jM)^{-1}(R^m)^{aj} \in \langle\, R^jM, R^m \,\rangle$$

   Therefore M (as well as R^m) belongs to $\langle\{R^m,\; R^jM\}\rangle$

   Recall from Lemma II.4.10 that R^jM is reflection in the line through the origin at angle $\frac{\pi j}{n}$ radians with the x-axis. Thus, D_n is generated by the two isometries: rotation by $\frac{2\pi m}{n}$ radians (where m is relatively prime to n) and reflection in the line at angle $\frac{\pi j}{n}$ radians with the x-axis, for any integer j.

3. The preceding fact implies that **two reflections R^jM and R^kM in D_n generate D_n if and only if the angle between their axes , $\frac{\pi j - \pi k}{n}$, is half the size of some generating rotation: i.e., iff $2(j-k)$ is relatively prime to n.** (Exercise 2)

Exercises II.6

1. (a) Is $\{\mathrm{id}_{\mathbb{C}},\ R^2,\ R^4, M,\ R^2M,\ R^4M\}$ a subgroup of the dihedral group D_6 (cf. Definition II.4.9)?

   (b) Is $\{\, T_a \,|\, a \in \mathbb{R},\ a > 0\,\}$ a subgroup of the group $\mathcal{T}$ of all translations of the plane (defined in Example II.4.6).

2. Prove that two reflections R^jM and R^kM in D_n generate D_n if and only if the angle between their axes , $\frac{\pi j - \pi k}{n}$, is half the size of some generating rotation: i.e., iff $2(j-k)$ is relatively prime to n.

3. (a) If H is a finite set and $\ell : H \longrightarrow H$ is a one-one function, prove that ℓ is onto.

   (b) Give an example of an infinite set H and a one-one function $\ell : H \longrightarrow H$ which is not onto.

4. (a) List the elements of $\langle\, 2, i\,\rangle \subset \mathbb{C} - \{0\}$ where $\mathbb{C} - \{0\}$ is viewed as a group under multiplication.

   (b) Is $\mathbb{Z}_{14} = \langle\, 4,\, 7\,\rangle$?

   (c) Is $\mathbb{Z} = \langle\, 15,\, 33,\, 55\,\rangle$?

   (d) Is $D_8 = \langle\, M, R^3\,\rangle$?

5. Let Δ be an equilateral triangle in the plane with sides a, b, c opposite vertices A, $B.C$. Let $\alpha, \beta, \gamma \in \mathbb{C}$ be reflections in the lines containing a, b, c, respectfully. Draw Δ and the images of Δ under all elements of $\langle\, \alpha, \beta, \gamma\,\rangle$ of length two or less.

   [**Definition**. The length of $g_1^{n_1} g_2^{n_2} \cdots g_q^{n_q}$ is defined to be $|n_1| + |n_2| + \ldots + |n_q|$. Thus, in this problem, α^{-1} has length 1, β^2 has length 2, and $\alpha^{-1}\beta\alpha$ has length 3.]

6. Prove Proposition II.6.9, Part 2.

7. A group G is *finitely generated* if there exists a finite set $S \subset G$ such that $G = \langle S \rangle$.

   (a) Prove that the group $\mathbb{Q}$ is not finitely generated.

   (b) Prove that the direct sum $\mathbb{Z} \oplus \mathbb{Z}$ is finitely generated,

   (c) Prove that any finitely generated group is countable.
   (This means that it has a bijection to a finite set or to the set of positive integers. This problem assumes some experience with countable and uncountable sets.)

8. Find an example of an infinite subset H of an infinite group G such that H is closed under the group operation, but is nonetheless not a subgroup of G.

   (This shows that the hypothesis that H is finite in Lemma II.6.4 is necessary.)

II.7 The number of elements in a subgroup

If H is a subgroup of the group G then certain properties of G are obviously inherited by H. For example, if G is finite then H is finite. If G is abelian then H is abelian. The following is not quite as immediate:

Proposition II.7.1 *Every subgroup of a cyclic group is cyclic.*

Proof: Suppose that H is a non-trivial subgroup of the cyclic group $G = \langle g \rangle$. (The trivial subgroup $\{e\}$ is cyclic with generator e, by definition.) We claim that H is cyclic with generator g^m for for some positive integer m.

Indeed, let m be the least positive integer such that $g^m \in H$. (There is such an m because, since G is cyclic, H contains some g^j with $j \neq 0$. Being a subgroup, H also contains $(g^j)^{-1} = g^{-j}$. So by considering g^{-j} if necessary, we see that H contains positive powers of g. We let m be the least integer in this set of positive exponents.)

Since H is a subgroup, it then contains (see Proposition II.6.9,1)

$$\langle g^m \rangle = \{ (g^m)^k \mid k \in \mathbb{Z}\} = \{ g^{km} \mid k \in \mathbb{Z}\}.$$

If there were another element $g^q \in H$which is not in $\langle g^m \rangle$, we could write $q = km+r$ for some r with $0 < r < m$. But then, since H is a subgroup, $(g^m)^{-k} = g^{-mk} \in H$ so that $g^{-mk}g^q = g^{-mk}(g^{mk}g^r) = g^r \in H$. This would contradict the choice of m as being the smallest such integer. Therefore $H = \langle g^m \rangle$. ∎.

Note: It follows (Exercise 1) from the above proof that we can be more precise:

1. If $G = \langle g \rangle$ is cyclic and g has infinite order then the non-trivial subgroup H of G is cyclic with generator of infinite order.
2. If $G = \langle g \rangle$ is cyclic and g has finite order n — so that $G = \{e, g, g^2, \ldots g^{n-1}\}$ — then $H = \langle h \rangle$ where $h = g^m$ has order k and $mk = n$.

For example, the only proper subgroups of $\mathbb{Z}_{10}$ are $H_1 = \{0, 5\}$ and $H_2 = \{0, 2, 4, 6, 8\}$ since the only divisors of 10 are 2 and 5.

Recall (Definition II.3.8) that the *order of a group* G is the number of elements in G. If G contains infinitely many elements then we say that G *is of infinite order*.

The order of a cyclic group is equal to the order of any one of its generators. However, a non-cyclic group will have order greater than the order of any of its elements. For example the dihedral group $D_n = \{\text{id}, R, R^2, \ldots, R^{n-1}, M, RM, \ldots R^{n-1}M\}$ (II.4.9) is a group of order $2n$ which has an element e of order 1, $n-1$ elements $R, \ldots, R^{n-1}$ whose orders divide n (by the note above) and n elements $M, RM, \ldots R^{n-1}M$ each of which has order 2 (since they are reflections).

This discussion gives evidence for the following striking and important theorem:

Theorem II.7.2 (LaGrange's Theorem)
If H is a subgroup of the finite group G then the order of H divides the order of G.

We will give the proof of this theorem after we give some corollaries which give some idea of its significance and after we introduce the fundamental concept of a *coset*.

Corollary II.7.3 (Order of g vs. order of G) *If G is a finite group and $g \in G$ then*

1. *the order of g divides the order of G*
2. $g^{\text{ord}(G)} = e$.

Corollary II.7.4 (Groups of prime order are cyclic) *A group of prime order is cyclic.*

Proof of Corollary II.7.3: By Proposition II.3.7, the order of g equals the order of the subgroup $\langle g \rangle$. By Lagrange's theorem, the latter divides the order of G, proving 1. Further, since $\text{ord}(G) = \text{ord}(g) \cdot k$ for some integer k, we have $g^{\text{ord}(G)} = (g^{\text{ord}(g)})^k = e^k = e$, proving 2. ∎

Proof of Corollary II.7.4: Let G be a group of prime order p. Let $g \in G$ with $g \neq e$. By LaGrange's Theorem, the order of $\langle g \rangle$ divides the prime p. Since the order of $\langle g \rangle$ is not equal to 1, the order of $\langle g \rangle$ is equal to p. Since $\langle g \rangle \subset G$ and G also has p elements in it, $G = \langle g \rangle$. ∎

Definition II.7.5 (Coset) *If H is a subgroup of G then* right cosets *and* left cosets *of H in G are subsets of G of the form*

$$\begin{array}{rll} \text{right coset of } H\text{:} & Ha = \{ha \,|\, h \in H\} & \text{for some } a \in G \\ \text{left coset of } H\text{:} & aH = \{ah \,|\, h \in H\} & \text{for some } a \in G \end{array}$$

Notes:

1. For any $h \in H$, $H = Hh = hH$ is both a right coset and a left coset of H. But, if G is not abelian, other right cosets are often not left cosets. (Try Exercise 6).

2. Given a coset of the form aH (or Ha), we will call the element $a \in G$ a *label* of the coset aH (or Ha). It is crucial to realize that, by definition, a coset is a *subset* of G. Thus it is reasonable to ask, *when do two labels define the same coset?* The next Proposition provides an answer.

Proposition II.7.6 (When do cosets coincide?) *If $H < G$ and $a, b \in G$, then the following are equivalent:*

1. $aH = bH$
2. $a^{-1}b \in H$
3. $a \in bH$
4. $aH \cap bH \neq \emptyset$

The analagous set of equivalences holds for right cosets.

The proof is Exercise 10. As a consequence, *the set of possible labels for a coset is precisely the set of elements of the coset itself.*

Example II.7.7 (Even and odd integers as cosets) *Suppose that H is the subgroup of $G = \mathbb{Z}$ consisting of the even integers. Then the cosets of H consist of H itself and the set of odd integers:*

$$\begin{aligned}\mathbb{Z} &= \{\textit{even integers}\} \sqcup \{\textit{odd integers}\} \\ &= \{\#\textit{s of form } 0+ \textit{ (even integer)}\} \sqcup \{\#\textit{s of form } 1+ \textit{ (even integer)}\} \\ &= 0 + H \sqcup 1 + H.\end{aligned}$$

Here the symbol $\sqcup$ denotes **disjoint union** *– the union of disjoint sets.*

Proof of LaGrange's Theorem, II.7.2:

The outline of the proof using right cosets is this: (the Theorem can similarly be proved using left cosets.)

1. There is a bijection from H to any one of its right cosets Ha, given by the correspondence $h \mapsto ha$. So all right cosets have the same number of elements as H.
2. The right cosets form a partition of G: They are pairwise disjoint and their union is all of G.
3. If there are m right cosets and the order of H is k, then (1) and (2) imply that the order of G is $m \cdot k$.

(1) and (2) are fundamental facts about cosets. *They are true whether G is finite or not.* The reader should verify (i) and (ii) directly. But in fact, we can see that we have already proved them if we think in terms of group actions as in Example II.5.7.

Let $\ell_H = \{\ell_h \mid h \in H\}$ where

$$\ell_h : G \longrightarrow G \text{ by the rule that } \ell_h(x) = hx \text{ for all } x \in G.$$

The group ℓ_H is shown to act freely on G in Example II.5.7. The orbit of $a \in G$ is just the right coset Ha. By Proposition II.5.8 the orbits form a partition of G, and by II.5.5 all these orbits have the same cardinality as H. ∎

Definition II.7.8 (Index) *If H is a subgroup of G then* the index of H in G — *denoted $[G : H]$ is the number of left (or right) cosets of H in G. H is of* infinite index *if there are infinitely many cosets of H in G.*

Note:

- There is a bijection between the set of left cosets of H in G and the set of right cosets of H in G given by (Exercise 8) $aH \leftrightarrow Ha^{-1}$.
- By LaGrange's Theorem, $[G : H] = \frac{\text{order of } G}{\text{order of } H}$ is a positive integer when G is a finite group.

Example II.7.7 shows that the subgroup of even integers has index 2 in $\mathbb{Z}$.

The following example uses several of the concepts and results introduced thus far. We exhibit a group in which one subgroup has index 2 while another has infinite index.

Example II.7.9

Let $G = \langle T, M \rangle$ where T and M are the isometries of $\mathbb{C}$ given by $T(z) = z+1$, a horizontal translation by one unit, and $M(z) = \overline{z}$, reflection in the x-axis. Consider the subgroups

$$\langle T \rangle = \{T^n \mid n \in \mathbb{Z}\} \quad \text{and} \quad \langle M \rangle = \{\text{id}_{\mathbb{C}}, M\}$$

where $T^n(z) = z + n$.

By Proposition II.6.9, every element of $g \in G$ is of the form

$$g = M^{n_1} T^{m_1} M^{n_2} T^{m_2} \ldots M^{n_k} T^{m_k}$$

where the n_i and m_i are integers. (Note that n_1 or m_k might be 0, so that this might really start with a power of T or end with a power of M.) However, M is a reflection, so that

$$M^{\text{even power}} = \text{id}_{\mathbb{C}} \qquad \text{and} \qquad M^{\text{odd power}} = M.$$

Further,

$$MT(z) \;=\; \overline{z+1} \;=\; \overline{z}+1 \;=\; TM(z),$$

so that $TM = MT$ and G is abelian. It follows from these observations that, if we let $m = m_1 + m_2 + \ldots + m_k$, then every element $g \in G$ can be written in the simplified form

$$g = \begin{cases} MT^m = T^m M, & \text{if } n_1 + \ldots n_k \text{ is odd,} \\ T^m, & \text{if } n_1 + \ldots n_k \text{ is even.} \end{cases}$$

Notice that $M^\epsilon T^a = M^\delta T^b$ $(\epsilon, \delta \in \{0,1\},\ a, b \in \mathbb{Z})$ if and only if $\epsilon = \delta$ and $a = b$. (We can tell that different choices of the ordered pair (ϵ, δ) yield different functions $M^\epsilon T^\delta$ by just checking what the functions do to the complex number i: *Exercise!*)

We may therefore decompose the elements of G first as

$$G = \{...T^{-2}, T^{-1}, \mathrm{id}, T^1, T^2, ...\} \sqcup \{...T^{-2}M, T^{-1}M, M, T^1M, T^2M, ...\}$$

i.e.

$$G = \langle\, T \,\rangle \sqcup \langle\, T \,\rangle M$$

and so the subgroup $\langle\, T \,\rangle$ has index two in G. But we may also decompose the elements of G as

$$G = ... \sqcup \{T^{-2}, MT^{-2}\} \sqcup \{T^{-1}, MT^{-1}\} \sqcup \{\mathrm{id}, M\} \sqcup \{T^1, MT^1\} \sqcup \{T^2, MT^2\} \sqcup ...$$

i.e.

$$G = ... \sqcup \langle\, M \,\rangle T^{-2} \sqcup \langle\, M \,\rangle T^{-1} \sqcup \langle\, M \,\rangle \sqcup \langle\, M \,\rangle T^2 \sqcup \langle\, M \,\rangle T^2 \sqcup ...$$

and so the subgroup $\langle\, M \,\rangle$ has infinite index in G. ■

Exercises II.7

1. Prove the following assertions:

   (a) If $G = \langle\, g \,\rangle$ is cyclic and g has infinite order then any non-trivial subgroup H of G is cyclic with generator of infinite order.

(b) If $G = \langle g \rangle$ is cyclic and g has finite order n — so that $G = \{e, g, g^2, \ldots g^{n-1}\}$ — then any subgroup H of G is of the form $H = \langle h \rangle$ where $h = g^m$ has order k and $mk = n$.

2. List all possible orders of elements of $\mathbb{Z}_{24}$, and exhibit an element of each order. Can two different elements have the same order?

3. Is there a group of order 8 which contains no element of order 4?
Why doesn't your answer contradict Lagrange's Theorem?

4. Suppose that p is a prime number and G_1 and G_2 are both groups of order p. Prove that $G_1 \cong G_2$.

5. Let $D_3 = \langle R, M \rangle$ be the dihedral group of order 6 (see Definition II.4.9). Let H be the subgroup $H = \langle M \rangle$. Is the right coset HR equal to a left coset xH for some $x \in D_3$?

6. Consider the additive groups $\mathbb{Z} \subset \mathbb{R} \subset \mathbb{C}$.
What are the cosets of $\mathbb{Z}$ in $\mathbb{R}$? of $\mathbb{R}$ in $\mathbb{C}$? of $\mathbb{Z}$ in $\mathbb{C}$?
Draw a few cosets in each case.

7. Show that the line $L_{2+3i} = \{ t(2+3i) \mid t \in \mathbb{R} \}$ is a subgroup of the additive group $\mathbb{C}$. Draw this subgroup and its cosets.

8. Prove that the number of left cosets of the subgroup H of G is equal to the number of right cosets.

9. Prove **(Fermat's Little Theorem:**

If p is a prime number and a is any non-zero integer, then $a^p \equiv a$ (modulo p).

Hint: Suppose first that $a \in \mathbb{J}_p = \{1, 2, \ldots, , p-1\}$ (as in Section II.1, Exercise 5c), and consider the order of a.

10. (a) Show that the group G of Example II.7.9 is equal to $\mathrm{Sym}(P)$ where P is the subset of the plane consisting of the union of the x-axis and the line segments of length 1 which stick out at angles ± 45 degrees from the integer points.

(b) Show that the group G of Example II.7.9 is isomorphic to $\mathbb{Z} \oplus \mathbb{Z}_2$

11. Prove the equivalences in Proposition **??**. Furthermore, show that $aH = bH$ if and only if $b^{-1}a \in H$. Finally, formulate precisely and prove an analogous result for right cosets.

II.8 Finite permutation groups

Recall that S_X denotes the symmetric group on the set X — the group of all permutations of X. Suppose $X = X_n$, where $X_n = \{1, 2, \ldots, n\}$. Then we denote $S_X = S_{\{1,2,\ldots,n\}}$ as simply S_n. In Exercise 1 you will show that whenever X has n elements, S_X is isomorphic to S_n. In other words, it is the number of elements (and not what we choose to call them) that determines the group of permutations. The group S_n is called the *symmetric group on n elements*.

We have seen (Proposition II.3.8, Cayley's Theorem) that any group G is isomorphic to a subgroup of the group S_G, where G is viewed as a set. Thus any finite group of order n is isomorphic to a subgroup of S_n. In this section we analyze the elements of S_n and introduce the subgroup of S_n consisting of the *even permutations*, the *alternating group* $A_n \subset S_n$.

Definition II.8.1 (Array notation for permutations) *If $g \in S_n$, we denote by*

$$\begin{bmatrix} 1 & 2 & \ldots & n \\ g(1) & g(2) & \ldots & g(n) \end{bmatrix}$$

the permutation sending $1 \mapsto g(1), \quad 2 \mapsto g(2), \ldots, \quad n \mapsto g(n)$.

Thus, for example,

$$g = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 3 & 1 & 7 & 2 & 6 & 4 \end{bmatrix}$$

is the permutation with

$$1 \mapsto 5, \qquad 2 \mapsto 3, \qquad 3 \mapsto 1, \qquad 4 \mapsto 7, \qquad 5 \mapsto 2, \qquad 6 \mapsto 6, \qquad 7 \mapsto 4$$

(here, $1 \mapsto 5$ means, "g sends 1 to 5").

Cycles

Array notation can be cumbersome. For example, suppose that X has ten elements $\{1, 2, \ldots, 10\}$ and that the permutation we are interested in interchanges 1 and 2

and fixes the remaining elements. To describe this simple permutation using array notation requires ten columns! Composing permutations is also quite cumbersome.

An alternative, the cycle notation, is quite compact and easy to work with, but requires a bit of getting used to. We first illustrate this with the above example. If we "follow the effect" of the permutation g above, we can see that it breaks up into *cycles*:

$$1 \mapsto 5 \mapsto 2 \mapsto 3 \mapsto 1, \qquad 4 \mapsto 7 \mapsto 4 \qquad \text{and} \qquad 6 \mapsto 6$$

We encode the permutation of g by writing

$$g = (1523)(47)(6)$$

or, for short, we suppress mention of the elements which are fixed under g and we write

$$g = (1523)(47).$$

Some features of this notation:

- **Non-uniqueness.** A given permutation g may have many representations in cycle notation. For example, the same permutation g can be encoded by a symbol in which the order of the stuff in parentheses is changed:

  $$g = (47)(1523)(6) = (6)(1523)(47).$$

  We can also start with, say, 2 and 7 instead of 1 and 4 and follow these to get e.g.

  $$g = (2315)(74)(6).$$

- **Ambiguity using shorthand.** Consider e.g. (123). It could represent the element

  $$(123) \in S_3 \qquad \text{or } (123)(4)(5)(6)(7) \in S_7.$$

So, we will have to be careful when using this notation.

Notice that, if $g \in S_n$ then $\langle g \rangle = \{g^m \mid m \in \mathbb{Z}\}$ is a subgroup of S_n. So $\langle g \rangle$ is a group of permutations of X_n — it acts on this set. We use this to define the idea of a *cycle*, which is a special kind of permutation.

Definition II.8.2 (Cycle) *Let S_n be the symmetric group on $X_n = \{1, 2, \ldots, n\}$. An element $g \in S_n$ is called a* cycle *if, under the action of the group $\langle\, g \,\rangle < S_n$ on X_n, there is exactly one orbit (called the* underlying set *of the cycle) which has more than one point. The* length *of a cycle is the number of elements in this orbit.*

In S_7, consider the permutation given in cycle notation by $g = (1523)(47)(6)$. As you can check, there are three orbits:

$$\langle\, g \,\rangle.1 = \langle\, g \,\rangle.5 = \langle\, g \,\rangle.2 = \langle\, g \,\rangle.3 = \{1, 2, 3, 5\},$$

$$\langle\, g \,\rangle.4 = \langle\, g \,\rangle.7 = \{4, 7\},$$

and

$$\langle\, g \,\rangle.6 = \{6\}.$$

Since two of these orbits consist of more than one point, the permutation g is *not* a cycle.

However, in S_7 again, the two permutations g_1, g_2 given in (shorthand) cycle notation by

$$g_1 = (1523) \qquad \text{and} \qquad g_2 = (47)$$

are both cycles with underlying sets $\{1, 2, 3, 5\}$ and $\{4, 7\}$, respectively. Moreover, $g = g_1 \circ g_2 = g_2 \circ g_1$ (check this!) So, even though g itself is not a cycle, it is a product of cycles in the group S_7, and the order of the terms in the product is irrelevant.

Definition II.8.3 (Disjoint cycles) *Two cycles* $c_1 = (i_1\, i_2\, \ldots\, i_k)$ *and* $c_2 = (j_1\, j_2\, \ldots\, j_\ell)$ *are* disjoint cycles *if* $\{i_1, i_2, \ldots\, i_k\} \cap \{j_1, j_2, \ldots\, j_\ell\} = \emptyset$.

Thus, in S_7, the cycles (47) and (1523) are disjoint, while (1523) and (36) are not disjoint.

Lemma II.8.4 (Basic facts about cycles)

1. $(i_1\, i_2\, \ldots\, i_k) = (i_k\, i_1\, \ldots\, i_{k-1}) = (i_{k-1}\, i_k\, i_1\, \ldots\, i_{k-2}) = \ldots = (i_2 i_3 \ldots i_k i_1)$

2. $(i_1\, i_2\, \ldots\, i_k)^{-1} \;=\; (i_k\, i_{k-1}\, \ldots\, i_2\, i_1)$

3. *If* c_1 *and* c_2 *are disjoint cycles then* $c_1c_2 \;=\; c_2c_1$.

Proof: The proof is left to the reader (Exercise 3).

The next Proposition makes precise what we did with the example: "following the effect" of a permutation allows us to write any permutation as a product of cycles.

Proposition II.8.5 (Permutations are products of cycles) *Let* g *be any non-identity element of* S_n. *Then there is a unique set* $C = \{c_1, c_2, ...c_s\}$ *of disjoint cycles of length* ≥ 2 *such that*

1. g *is the product of the elements of* C, *taken in any order;*
2. *any two different cycles in* C *are disjoint;*
3. *the non-trivial orbits (= those orbits consisting of more than one point) of the action of the cyclic subgroup* $\langle\, g\,\rangle$ *on* $X_n = \{1, 2, \ldots, n\}$. *are the underlying sets of the cycles in* C.

Proof: By Proposition II.5.8, the orbits of the group $\langle\, g\,\rangle$ acting on S_n are disjoint. Let

$$O = \{O_1, O_2, ..., O_s\}$$

be the set of those orbits of $\langle\, g\,\rangle$ which contain at least two points. Then O is nonempty since g is not the identity element.

If $j \in O_1$, then

$$O_1 = \{g^m(j) \;\mid\; m \in \mathbb{Z}\} \;=\; \{\, j,\, g(j),\, g^2(j),\, \ldots\,\} \cup \{\, g^{-1}(j), g^{-2}(j), \ldots\,\}.$$

We're going to extract a cycle from this orbit. Since X_n is finite, there must be a first time when a repetition occurs in the list $j,\, g(j),\, g^2(j),\, \ldots$. Suppose that $g^k(j) = g^\ell(j), \quad 0 \leq k < \ell,$ and that there are no repetitions between these two elements of the list. Let $q = \ell - k$. Then

$$j \;=\; g^{-k}g^k(j) \;=\; g^{-k}g^\ell(j) \;=\; g^q(j)$$

and there are no repetitions in the list $j, g(j), \ldots, g^{q-1}(j)$. Further, $g^{-q}(j) = g^{-q}g^q(j) = j$, and it follows, for any integers a, b, with $0 \leq b < q$, that $g^{aq+b}(j) = g^b(j)$. Hence

$$[\, O_1 = \text{the orbit of } j \text{ under } \langle\, g \,\rangle] \;=\; \{j, g(j), \ldots, g^{q-1}(j)\}.$$

Let c_1 be the permutation given in cycle notation by

$$c_1 = (j\, g(j)\, \ldots\, g^{q-1}(j)).$$

The underlying set of the cycle c_1 is the orbit O_1. By construction, the permutation c_1 and the original permutation g agree on O_1. Moreover, off of O_1, the permutation c_1 is the identity. The length of c_1 is $l \geq 2$.

Now repeat the procedure for the remaining orbits $O_2, O_3, ...$ to obtain cycles $c_2, c_3, ...$ and put $C = \{c_1, c_2, ..., c_s\}$. Note that

$$c_i(j) = \begin{cases} g(j) & \text{if } j \in O_i \\ j & \text{otherwise.} \end{cases} \qquad (*)$$

Since the orbits are disjoint, any pair of different cycles in C is disjoint, proving (2). Claim (3) follows by construction.

To prove (1), let $h =$ the product of the elements in O. Disjoint cycles commute, by Lemma II.8.4, so the product h of the elements of O is the same no matter what order we take in the product. We will be done if we show that $h = g$.

To show $h = g$, we must show that for all $j \in X_n$, we have $h(j) = g(j)$. So let $j \in X_n$. If j is *not* in any of the orbits O_i in O, then $g(j) = j$ and $c_i(j) = j$ for all i. Since h is the product (which in this case means composition) of the cycles c_i, $h(j) = j$. So $g(j) = j = h(j)$ in this case. Now suppose j *is* in one of the orbits, say O_i. The permutation h is the product of the cycles in C and is the same no matter what order the product is written. For convenience let's write

$$h = c_i c_1 c_2 ... \widehat{c_i} ... c_s$$

where the notation $\widehat{c_i}$ means that the term c_i is *missing* in the expression $c_1 c_2 ... c_s$. Since $j \in O_i$, the underlying set of the cycle c_i, and since O_i is disjoint from the underlying sets of the remaining cycles, we have by (*)

$$c_1 c_2 ... \widehat{c_i} ... c_s(j) = j.$$

So

$$h(j) = c_i c_1 c_2 ... \widehat{c_i} ... c_s(j) = c_i(j) = g(j)$$

where the last equality follows by (*). Hence $h = g$ and we are done.

■

Corollary II.8.6 (S_n is generated by cycles) *S_n is generated by the set of cycles it contains.*

Proof: According to the proposition above, each $g \in S_n$ is a product of cycles. ■

The preceding Proposition gives each element in a very nice way — as a product of disjoint cycles, unique except for the order in which the cycles are written. However, it is often useful to sacrifice the disjointedness in order to use the shortest possible non-identity cycles.

Definition II.8.7 (Transposition) *A* **transposition** *is a cycle (ij) of length 2. (We emphasize that $i \neq j$).*

By definition, $g = (ij)$ interchanges ("transposes") i and j and fixes every $k \neq i, j$. From the last Corollary we get the following.

Corollary II.8.8 (S_n generated by transpositions)

- *Every cycle of length k is the product of $(k-1)$ transpositions.*
- *S_n is generated by the set of transpositions it contains.*

Proof: To prove the first statement, suppose that $c = (j_1 \dots j_k)$ is a cycle in S_n. For simplicity of notation assume that $c = (12 \dots k)$: $1 \mapsto 2 \mapsto 3 \mapsto \cdots \mapsto k \mapsto 1)$. The statement follows from that observation that

$$(12 \cdots k) \;=\; (1k)(1\,k-1) \dots (14)(13)(12).$$

The equality holds because the functions on both sides of the equation have the same effect on each number j, $1 \leq j \leq n$. [Remember that transpositions, being functions, are composed from right to left.]

The statement that transpositions generate now follow from the fact (II.8.6) that the cycles generate, and each cycle is a product of transpositions (with the convention that any cycle of length 1 (= identity map) is the product of zero transpositions). ■

Even and odd permutations

Notice than an element $g \in S_n$ can be written in very different ways as a product of transpositions:

$$(12)(34)(42)(12)(34) \;=\; (13)$$

However we have the following remarkable fact:

Theorem II.8.9 (Even and odd) *If $g \in S_n$ is written as a product of transpositions in two different ways*

$$g \;=\; t_1 t_2 \dots t_p \;=\; t'_1 t'_2 \dots t'_q$$

then p and q are both odd or both even.

Definition II.8.10 (Even and odd permutations) *A permutation $g \in S_n$ is called* **even (respectively, odd)** *if g is the product of an even (respectively, odd) number of transpositions.*

Here's another interpretation of the content of Theorem II.8.9, *Even and Odd*. The definition of even/odd permutations is not automatically *well-defined*. A given permutation may be written as a product of transpositions in many different ways. So, conceivably, it is possible for a given permutation g to be written as, say, the product of three transpositions, and also the product of eight transpositions. If this were the case, then we would not know whether to call g even or odd. The Theorem, however, says that this is impossible, and so another way of stating the theorem is this: *the concept of even/odd permutation given above is well-defined.*

Before proving Theorem II.8.9 we introduce the *crossing pairs* for g and the sign $\sigma(g)$ of a permutation and we prove Lemma II.8.13 concerning the sign of a composition.

We analyze g in terms of what it does to 2-element subsets $\{i, j\}$ of X_n. Since g is one-one, $\{g(i), g(j)\}$ is also a two element subset. Notice (Exercise 9) that if $\mathcal{T}_0$ is the set of all $2-$element subsets of X_n, then the function

$$\Theta_g : \mathcal{T}_0 \longrightarrow \mathcal{T}_0 \quad \text{given by} \quad \Theta_g(\{i, j\}) = \{g(i), g(j)\}$$

is a bijection.

Definition II.8.11 (Crossing pair and number)

1. *We say that* $\{i, j\}$ *is a* crossing pair for g *if* $i < j$ *while* $g(i) > g(j)$ *or if* $i > j$ *while* $g(i) < g(j)$.

   Pictorially, if we draw a line segment from $(0, k)$ *to* $(1, g(k))$, $1 \leq k \leq n$,

   then the line segment $\overline{(0, i), (1, g(i))}$ *crosses the line segment* $\overline{(0, j), (1, g(j))}$.

2. *We define the* crossing number $c(g)$ *and the* sign $\sigma(g)$ of g *by*

$$c(g) = \text{the number of crossing pairs for } g$$

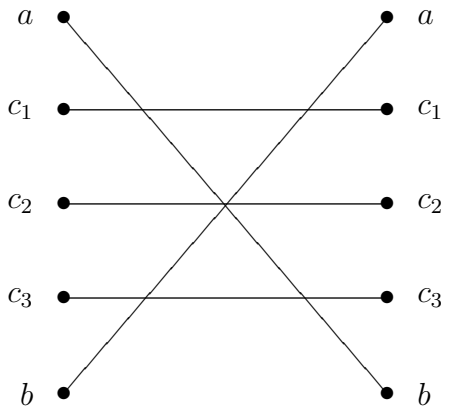
$$\sigma(g) = (-1)^{c(g)} = \begin{cases} +1 & \text{if the number of crossing pairs is even,} \\ -1 & \text{if the number of crossing pairs is odd.} \end{cases}$$

Lemma II.8.12 (Sign of transposition is -1)

If $t = (ab)$ *is a transposition* $(1 \leq a < b \leq n)$ *then* $\sigma(t) = -1$.

Proof:

The only non-horizontal lines are the lines from a to b and from b to a. These cross each other and each crosses every horizontal line from c to c for each integer c with $a < c < b$. Therefore $c(t) = 2(b-a-1)+1$ and $\sigma(t) = -1$.



■

Lemma II.8.13 (Signs multiply under composition) *For all* $g, h \in S_n$ *the function* $\sigma : S_n \longrightarrow \{+1, -1\}$ *satisfies*

$$\sigma(gh) = \sigma(g)\sigma(h).$$

Proof: If $\{i, j\}$ is a crossing pair for h and then its image $\{h(i), h(j)\}$ is a crossing pair for g, it follows that $\{i, j\}$ is *not* a crossing pair for gh. Therefore $\{i, j\}$ is a crossing pair for $gh \iff$ either

(i) $\{i, j\}$ is a crossing pair for h while $\{h(i), h(j)\}$ is not a crossing pair for g **or**

(ii) $\{i, j\}$ is not a crossing pair for h but $\{h(i), h(j)\}$ is a crossing pair for g.

Suppose that m of the crossing pairs $\{i, j\}$ for h have $\{h(i), h(j)\}$ as a crossing pair for g. Hence $c(gh) = (c(g) - m) + (c(h) - m)$ so that

$$\sigma(gh) = (-1)^{c(g)+c(h)-2m} = (-1)^{c(g)}(-1)^{c(h)} = \sigma(g)\sigma(h).$$

■

Proof of Theorem II.8.9 Suppose that g is written as a product of transpositions in two ways

$$g \;=\; t_1 t_2 \dots t_p \;=\; t'_1 t'_2 \dots t'_q$$

By Lemma II.8.12, $\sigma(t_i) = \sigma(t'_j) = -1$ for each of the transpositions appearing. Thus, repeatedly using the fact II.8.13 that the signs multiply, we have

$$(-1)^p \;=\; \sigma(g) \;=\; (-1)^q. \qquad \blacksquare$$

From the proof of Theorem II.8.9 we get the following corollary.

Corollary II.8.14 (Even iff positive sign) *The permutation g is even (i.e., a product of an even number of transpositions) $\iff$ $\sigma(g) = +1$ (i.e., its crossing number is even).*

Example II.8.15 How to determine if a permutation is even or odd: *Write the given permutation as a product of disjoint cycles. By II.8.8 a cycle of length k is even $\iff$ the integer k is odd. Then use the fact that the signs multiply.*

For example,

$$\begin{aligned} g \;&=\; \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 5 & 6 & 1 & 7 & 8 & 2 & 9 & 3 \end{bmatrix} \\ &=\; (14)(257)(3680) \\ &=\; \textit{odd} \cdot \textit{even} \cdot \textit{odd} \\ &=\; \textit{even permutation.} \end{aligned}$$

Definition II.8.16 (Alternating group) *The* **alternating group** *A_n is the subgroup of the symmetric group S_n consisting of all even permutations.*

Exercises II.8

1. "The symmetric group on a set really only depends on the number of elements in the set and not on the particular nature of those elements." To justify this

statement and make it precise, prove the following proposition:

If $F : X \longrightarrow Y$ is a bijection then the groups S_X and S_Y are isomorphic under the isomorphism $\phi : S_X \longrightarrow S_Y$ given by $\phi(g) = F \circ g \circ F^{-1}$ for all $g \in S_X$.

2. List all the cyclic subgroups of S_3.

3. Prove Lemma II.8.4..

4. State and prove a version of Proposition II.8.5 for a permutation g of an infinite set X. (In this case, you will have to allow infinite "cycles", which don't ever really come back to where they start.)

5. Let
$$g = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 3 & 6 & 2 & 1 \end{bmatrix} \in S_6.$$
   (a) Write g as a product of disjoint cycles.
   (b) Write g as a product of transpositions.
   (c) Find the crossing number $c(g)$ and the sign $\sigma(g)$ of g.
   (d) What is the order of g?
   (e) What are the orbits under the action of $\langle g \rangle$ on X_6?

6. Suppose that $g = c_1 c_2 \dots c_q$ is a product of disjoint cycles of lengths $\ell_1, \ell_2, \dots, \ell_q$ respectively. Prove that
$$\text{ord}(g) = \text{l.c.m.}(\ell_1, \ell_2, \dots, \ell_q) \qquad [\text{l.c.m. = least common multiple}]$$

7. Show that the subgroup of S_4 which is generated by $\{(13), (1234)\}$ is isomorphic to the dihedral group D_4. (Hint: Think of D_4 as the symmetry group of a square in the plane centered at the origin, with vertices labelled counterclockwise as 1, 2, 3, 4.)

8. (a) Prove: If $g \in S_n$ and $h = (i_1 i_2 \dots i_k)$ then
$$ghg^{-1} = \big(g(i_1)g(i_2)\dots g(i_k)\big)$$
   (b) Use a) to quickly calculate $(135)(124)(3)(56)(531)$.

9. If $g \in S_n$ and if $\mathcal{T}_0$ is the set of all $2-$element subsets of X_n, prove that the function $\Theta_g : \mathcal{T}_0 \longrightarrow \mathcal{T}_0$ given by $\Theta_g(\{i, j\}) = \{g(i), g(j)\}$ is a bijection.

10. Let A_n (*the alternating group on n elements*) consist of the set of all even permutations in S_n. Prove that A_n is indeed a subgroup of S_n and that it has index two in S_n and has order $n!/2$.

11. (a) Prove or disprove: $A_3 \cong \mathbb{Z}_3$.
    (b) Explain why the subgroup $G = \langle (123), (234) \rangle$ of S_4 is contained in A_4. List all of its elements. (Lagrange's Theorem might save you some work!)

12. In S_{20} is the following permutation g even or odd?

$$g = (1\,3\,\widehat{13})\,(2\,8)\,(\widehat{20}\,\widehat{17}\,\widehat{19}\,7)(4\,\widehat{14})\,(\widehat{18}\,\widehat{16}\,\widehat{15}\,9\,6\,5)\,(\widehat{11}\,\widehat{12}\,\widehat{10})$$

    Here we denote (for example) the number sixteen by $\widehat{16}$.

13. Exhibit the elements of the cosets of the following subgroups of the indicated group.

    a) $\langle (13) \rangle \subset S_3$ **b)** $\langle (123) \rangle \subset S_3$ **c)** $\langle (123) \rangle \subset A_4$
    d) the cosets of $A_4 \cap \langle (1234) \rangle$ in A_4 where $\langle (1234) \rangle$ is taken in S_4.

14. Exhibit the elements of the cosets of the subgroup H of S_{n+1} where

$$H = \{g \in S_{n+1} \mid g(n+1) = n+1\}.$$

    (This is indeed a subgroup, since it is $\mathrm{Stab}_{S_{n+1}}(n+1)$.)
    Hint: Consider the transpositions $(k\,(n+1))$, $1 \le k \le n$.

15. According to Cayley's Theorem (II.3.8), every group G is isomorphic to a group ℓ_G of permutations of the set G. Let

$$G = D_2 = \{id,\ R_\pi,\ M,\ M_{\pi/2}\}$$

    Let $F : D_2 \longrightarrow \{1,\ 2,\ 3,\ 4\}$ be given by

$$id \mapsto 1, \qquad R_\pi \mapsto 2, \qquad M \mapsto 3, \qquad M_{\pi/2} \mapsto 4.$$

    As in Exercise 1 , we get an isomorphism $\phi : S_{D_2} \longrightarrow S_4$, and thus we have a subgroup $\phi(\ell_{D_2})$ of S_4 which is isomorphic to D_2.
    List the elements of the group $\phi(\ell_{D_2})$, and draw its multiplication table. (It should mimic the multiplication table of D_2!)

Chapter III

The Isometry Group of the Plane

In Chapter I we introduced and classified the individual isometries of the plane. In this chapter we study the *group of all such isometries*, Isom($\mathbb{C}$), and its subgroups

In Section III.1 we apply the classification of isometries to reveal a rich view of the group Isom($\mathbb{C}$). We discuss the fact that reflections generate Isom($\mathbb{C}$); we analyze the fixed point sets of isometries and the stabilizer groups of points; and finally we discuss the dichotomy among isometries of being *direct* or *opposite* — a dichotomy which is strongly related to that between *even* and *odd* permutations of finite sets. In any group of isometries the direct isometries form a subgroup of index two.

In Section III.2 we discuss how different isometries represent congruent geometric situations. These isometries must be *conjugate*: $g_2 = f g_1 f^{-1}$ for some $f \in \text{Isom}(\mathbb{C})$. In Section III.3 we introduce an important homomorphism $\pi : \text{Isom } \mathbb{C} \longrightarrow \text{O}_2$ which associates to each isometry f an isometry which fixes the origin (an orthogonal transformation) but which has the same effect on directions as f does. For each subgroup G of Isom($\mathbb{C}$) The group $\pi(G) \subset \text{O}_2$ carries much of the information of G. Finally, in Section III.4 we list all of the finite subgroups of Isom($\mathbb{C}$). In fact, they are all subgroups of dihedral groups (groups conjugate to the groups D_n).

III.1 What the classification tells us about Isom($\mathbb{C}$)

From the classification theorems I.6.1, I.6.2, a rich view of the group Isom($\mathbb{C}$) flows forth. In this section we discuss the following topics:

- *reflections generate* Isom($\mathbb{C}$)
- *fixed point sets of isometries*
- *stabilizers of points*
- *direct and opposite isometries.*

Reflections generate

Proposition III.1.1 (Reflections generate Isom($\mathbb{C}$)**)** *The group* Isom($\mathbb{C}$) *is generated by the set of all reflections across lines in the plane.*

Proof: An isometry is either a translation, a rotation, a reflection or a glide reflection, by Theorem I.6.1. However (Exercise 1) every translation is a product of two reflections and every rotation is a product of two reflections. Finally, every glide reflection [(= (non-zero translation) $\circ$ reflection] is then a product of three reflections.

■

The preceding proof gives the stronger fact:

Proposition III.1.2 *Every isometry is the product of three or fewer reflections.* ■

The fixed point sets of isometries

Recall from Section II.4 that the fixed point set of an isometry f of $\mathbb{C}$ is given by

$$\mathrm{Fix}(f) = \{z \in \mathbb{C} \mid f(z) = z\}.$$

Since we know precisely what the isometries of $\mathbb{C}$ are, we can immediately name all the fixed point sets of elements of Isom($\mathbb{C}$):

Proposition III.1.3 *Let* $f \in \mathrm{Isom}\ \mathbb{C}$. *Then* $\mathrm{Fix}(f)$ *is*

$\mathbb{C}$ *iff* $f = id$,

L *iff* f *is reflection in the line* L,

$\{z_0\}$ *iff* f *is a rotation with center* z_0,

$\emptyset$ *iff* f *is a translation or glide reflection.* ■

The stabilizers of points

Recall from Section II.4 that the stabilizer of an element $z \in \mathbb{C}$ under the action of Isom($\mathbb{C}$) is the subgroup defined by

$$\mathrm{stab}_{\mathrm{Isom}\ \mathbb{C}}(z) \ = \ \{g \in \mathrm{Isom}(\mathbb{C}) \mid g(z) = z\}.$$

When $z = 0$, we recall Lemma II.4.11 and conclude that

$$\mathrm{stab}_{\mathrm{Isom}\ \mathbb{C}}(0) = \{g \in \mathrm{Isom}(\mathbb{C}) \mid g(0) = 0\} = \mathrm{Isom}_0(\mathbb{C}) = \mathrm{O}_2.$$

The same result, for arbitrary points $z_0 \in \mathbb{C}$, can be simply given as:

Proposition III.1.4 *If* $z_0 \in \mathbb{C}$ *then* $\mathrm{stab}_{Isom\ \mathbb{C}}(z_0)$ *is the group of all rotations about* z_0 *and reflections across lines through* z_0.

Proof: Obviously, rotations about z_0 and reflections in lines through z_0 fix z_0. These are all of the isometries fixing z_0, since rotations about other points and reflections through other lines will not fix z_0, while non-zero translations and glide reflections do not fix anything. ■

Direct and opposite isometries

A fundamental dichotomy in how isometries behave is given by whether they are *direct* or *opposite*.

Definition III.1.5 (Direct and opposite) *An isometry f of $\mathbb{C}$ is called* direct *if it is a translation or rotation, i.e. if it has formula $f(z) = az+b$. It is called* opposite *if it is a reflection or glide reflection, i.e. if it has formula $f(z) = a\overline{z} + b$. The set of direct isometries is denoted by* $\mathrm{Isom}^+(\mathbb{C})$.

The terms *orientation-preserving/reversing* are sometimes used instead of direct/opposite, respectively. Since a direct isometry $f(z) = az + b$ maybe written as $f = T_b \circ R_\theta$ (with $a = e^{i\theta}$), it preserves the orientation (clockwise or counterclockwise) of directed arcs on circles. For example the counterclockwise arc from 1 to i on the unit circle about the origin goes to a counterclockwise arc from $f(1)$ to $f(i)$. See Figure 1.

Figure 1. A direct isometry.

An opposite isometry f, on the other hand, may be written as $f = T_b R_\theta M$. Directed arcs on circles have their orientation reversed. An arc which points counterclockwise has an image which points clockwise. See Figure 2.

Isometries preserve distance, and the magnitudes of angles, though the *sense* or *direction* of angles may be reversed. Direct isometries preserve the sense of angles; opposite isometries reverse them.

Figure 2. An opposite isometry.

Proposition III.1.6 *Multiplication of isometries follows the following pattern:*

$$\begin{array}{ccccc} \textit{direct} & \circ & \textit{direct} & = & \textit{direct} \\ \textit{direct} & \circ & \textit{opposite} & = & \textit{opposite} \\ \textit{opposite} & \circ & \textit{direct} & = & \textit{opposite} \\ \textit{opposite} & \circ & \textit{opposite} & = & \textit{direct.} \end{array}$$

In particular, the inverse of a direct isometry is direct, and the inverse of an opposite isometry is opposite.

Proof: To verify the table, just check the four cases, using formulas and the definition of a direct isometry (do it!). For the last sentence of the Proposition, , suppose f is direct, and let $g = f^{-1}$. Then $f \circ g = \mathrm{id}_{\mathbb{C}}$, which is direct. Hence by the table, g cannot be opposite. The case when f is opposite is proved similarly. ■

An isometry cannot be both direct and opposite, by the classification of isometries. Thus, any subgroup G of Isom($\mathbb{C}$) is partitioned into direct and opposite isometries.

$$G = \underbrace{G \cap \mathrm{Isom}^+(\mathbb{C})}_{\text{direct}} \quad \sqcup \quad \underbrace{G \cap (\mathrm{Isom}^+(\mathbb{C}) \cdot M)}_{\text{opposite}}.$$

It is possible that $G \cap (\mathrm{Isom}^+(\mathbb{C}) \cdot M)$ is empty. However $G \cap \mathrm{Isom}^+(\mathbb{C})$ is never empty since it contains the identity element. For example:

- If g is a glide reflection then

$$\langle\, g\,\rangle = \{g^n | n \in \mathbb{Z}\} = \underbrace{\{g^n \mid n \in \mathbb{Z} \text{ even}\}}_{\text{translations}} \sqcup \underbrace{\{g^n \mid n \in \mathbb{Z} \text{ odd}\}}_{\text{glide reflections}}.$$

  Therefore the direct elements $G \cap \mathrm{Isom}^+(\mathbb{C})$ of G consist of the identity and the (positive and negative) even powers g^n, which are translations. The opposite elements $G \cap (\mathrm{Isom}^+(\mathbb{C}) \cdot M)$ consist of the (positive and negative) odd powers g^n, which are again glide-reflections.

- If g is a rotation then every element of $\langle\, g\,\rangle = \{g^n \mid n \in \mathbb{Z}\}$ apart from the identity is a rotation, and so $\langle\, g\,\rangle \cap (\mathrm{Isom}^+(\mathbb{C}) \cdot M) = \emptyset$.

These examples and Proposition III.1.6 lead us to the next proposition.

Proposition III.1.7

1. $\mathrm{Isom}^+(\mathbb{C})$ *is a subgroup of* $\mathrm{Isom}(\mathbb{C})$.
   It is a subgroup of index two and thus

$$\mathrm{Isom}(\mathbb{C}) = \mathrm{Isom}^+(\mathbb{C}) \sqcup \mathrm{Isom}^+(\mathbb{C}) \cdot M,$$

   where $M(z) = \overline{z}$ *for all* $z \in \mathbb{C}$ *and "$\sqcup$" represents disjoint union.*

2. *If* G *is any subgroup of* $\mathrm{Isom}(\mathbb{C})$, *let* $G^+ = G \cap \mathrm{Isom}^+(\mathbb{C})$. *Then* G^+ *is a subgroup of* G *of index one or two.*

Proof: The proof is left as an exercise (Exercise 2). It is useful to remember, in checking that a subgroup A has index two in a subgroup B, that $x, y \in B$ belong to the same right coset of A (*i.e.,* $Ax = Ay$) if and only if $xy^{-1} \in A$.

■

This intersection result with subgroups of index two always occurs, as does the multiplication pattern with respect to subgroups of index two given in Proposition III.1.6. (See Exercise 3.)

Since reflections are opposite isometries, we can close this section by obtaining a more precise characterization of how reflections generate Isom(ℂ):

Theorem III.1.8 *An isometry is direct if and only if it is the composition of an even number of reflections.*

Proof: Suppose $f = f_1 g_1 ... f_k g_k$ where f_i, g_i are reflections. Then $f = (f_1 g_1)...(f_k g_k) = h_1 ... h_k$, where $h_i = f_i \circ g_i$. Since h_i is the composition of two reflections, h_i is direct. Moreover, the composition of any direct isometries is again direct. So f is direct if it is the composition of an even number of reflections.

To prove necessity, suppose that f is the composition of an odd number of reflections: $f = r f_1 g_1 ... f_k g_k$ where the f_i, g_i and r are reflections. The previous argument then shows that $f = rh$, where h is direct. Therefore f is opposite if it is the composition of an odd number of reflections. ∎

This result gives another way of *defining* a direct isometry: We could have defined an isometry to be direct if it is the composition of an even number of reflections. In some ways, this is appealing, since it does not make reference to the formula for such an isometry, and in our formulas the origin is special. But then we would have to check that an isometry could not be both a composition in one way of an even number of reflections (just as we had to check that an even permutation could not also be written as a product of an odd number of transpositions in Section II.8). Ultimately, we would also have had to use the classification of isometries had we chosen this definition of direct isometry.

Exercises III.1

1. (a) Prove that every non-zero translation T_b is a composition of two reflections in lines which are perpendicular to the direction of $\vec{b}$.

   (b) Prove that every rotation about a point z_0 is the composition of two reflections in lines through z_0.

2. (a) Show that the set $\text{Isom}^+(\mathbb{C})$ of direct isometries forms a subgroup of index two in Isom(ℂ).

(b) Prove: If G is any subgroup of Isom($\mathbb{C}$), and if $G^+ = G \cap \text{Isom}^+(\mathbb{C})$ then G^+ is a subgroup of G of index one or two.

3. Suppose that K^+ is a subgroup of index two of a group K.

   (a) If $x \in K$, $x \notin K^+$, prove that $K = K^+ \sqcup K^+ \cdot x$.

   (b) Generalize Proposition III.1.6 to a proposition about multiplication in K. Apply this to even and odd permutations in the symmetric group S_n.

   (c) If G is a subgroup of K, prove that $G \cap K^+$ has index one or two in G.

4. Suppose that an isometry f is given in vector notation by the formula $f(\vec{z}) = A\vec{z} + \vec{b}$. Prove that f is direct if and only if $\det A$ (= the determinant of the 2×2 matrix A) is equal to $+1$ and f is opposite if and only if det A $= -1$.

III.2 Conjugacy and Congruence

One mark of an educated person is the ability to see how something looks from another perspective. In Section I.4 we did this in two important situations.

1. Having defined rotation R_η by angle η about the origin, we asked

   What would be the formula for the analogous isometry in which the point z_0 plays the role of the origin?

   Our answer was to define rotation by η radians about z_0 by

   $$R_{z_0,\eta} = T_{z_0} \circ R_\eta \circ T_{z_0}^{-1}.$$

   Here T_{z_0} is an isometry (a translation) which moves the origin to z_0, and $T_{z_0}^{-1}$ moves z_0 to the origin.

2. Having obtained the simplest formula for a reflection — reflection across the x-axis given by $M(z) = \overline{z}$ — we asked

   What would be the formula for the analogous isometry in which the line L plays the role of the x-axis?

   Using crucial data which determines L – a point z_0 on L and the angle η which L makes with the x-axis – we noted that $R_{-\eta} \circ T_{-z_0} = (T_{z_0} \circ R_\eta)^{-1}$ takes L onto the x-axis, and $T_{z_0} \circ R_\eta$ takes the x-axis onto L. We defined

   $$M_L = \big(T_{z_0} \circ R_\eta\big) \circ M \circ \big(T_{z_0} \circ R_\eta\big)^{-1}.$$

This leads us to the general definition

Definition III.2.1 *Let G be a group. Two elements g_1, g_2 are called* conjugate in G *if there exists an element $f \in G$ such that $g_2 = f g_1 f^{-1}$. In this case we say that f* conjugates g_1 to g_2

In our current context this becomes:

Definition III.2.2 *Two isometries* $g_1, g_2 \in \text{Isom } \mathbb{C}$ *are* conjugate *in* Isom($\mathbb{C}$) *if there exists an isometry* f *such that* $g_2 = f g_1 f^{-1}$. *In this case we say that* f conjugates g_1 to g_2.

The examples above illustrate the general principle that

> *When the isometry* f *conjugates* g_1 *to* g_2, *the action of* g_1 *on a feature of the plane is traded in for "the same action" by* g_2 *on the image of this feature under* f.

To achieve the effect of g_2:

- the plane is moved by f^{-1},
- then g_1 is applied,
- then the plane is moved back by f.

Conjugacy as an equivalence relation

In any group G, let us write, just for the moment, $g_1 \sim g_2$ if g_1 and g_2 are conjugate in G. To say that $\sim$ is an equivalence relation on G means (see Appendix C) that, for all $g, g_1, g_2, g_3 \in G$,

- $g \sim g$. [*reflexive property*]
- $g_1 \sim g_2 \implies g_2 \sim g_1$. [*symmetric property*]
- $g_1 \sim g_2$ and $g_2 \sim g_3 \implies g_1 \sim g_3$. [*transitive property*]

Lemma III.2.3 *The relation of conjugacy is an equivalence relation on* G.

Proof: If $g \in G$ then $g \sim g$ because the identity element e satisfies $ege^{-1} = g$. [We could also note that $g \sim g$ because $g\,g\,g^{-1} = g$. The existence of any f with $fg_1f^{-1} = g_2$ proves that $g_1 \sim g_2$, and there is often more than one f conjugating g_1 to g_2.]

We leave the proof of the symmetric and transitive properties to the reader (Exercise 1). ∎

Definition III.2.4 *If $g \in G$ then the* conjugacy class of g in G *is the equivalence class of g under the relation of conjugacy in G. This equivalence class is denoted by $[g]$. Thus we have*

$$[g] = \{xgx^{-1} \mid x \in G\}.$$

Example III.2.5 *If e is the identity element in the group G, then $[e] = \{e\}$.*

Proof: *$e \in [e]$ since $e \sim e$. But for any $g \in [e]$, there exists $x \in G$ with $xex^{-1} = g$. Necessarily then, $g = e$.* ∎

Example III.2.6 (and warning!) *We must be explicit about what the ambient group G is when speaking of conjugacy classes: If g_1, $g_2 \in H < G$ then it's important to be clear as to whether g_1 is conjugate to g_2 in H or in G. In other words,*

$$g_1 = xg_2x^{-1}, \quad \textit{but is } x \textit{ in } H \textit{ or merely in } G?$$

- **If they are conjugate in H then they are conjugate in G,** *since $x \in H \implies x \in G$. Thus for example:*
  If two isometries are conjugate in Isom $\mathbb{C}$, then they are conjugate in Aff_2.
- *But the converse is not necessarily true.*

  For example, let T_1 and T_2 be translations by one and two units to the right. Let $m_2(z) = 2z$. Then m_2 is not an isometry, but $m_2 \in Aff_2$ while $T_1, T_2 \in Isom\ \mathbb{C} < Aff_2$. Then

  1. *$T_2 = m_2\,T_1\,m_2^{-1}$, so $T_1 \sim T_2$ in Aff_2.*
  2. *T_2 is not conjugate to T_1 in Isom $\mathbb{C}$.*

Statement 1 is true because

$$m_2\, T_1\, m_2^{-1}(z) = \lambda_2(\frac{1}{2}z + 1) = 2(\frac{1}{2}z + 1) = z + 2 = T_2(z).$$

Statement 2 will follow from Corollary **??**.

Conjugation of isometries

We now give the basic facts about conjugation of isometries. Besides the conjugation of an isometry g by another isometry (i.e., conjugation in Isom $\mathbb{C}$), we will be interested in Chapter V in changing scale – for example, in order to guarantee that a translation we care about is translation by one unit in some direction. A change in scale by a factor of $\lambda > 0$ will be achieved by conjugation by the dilation $m_\lambda \in \mathrm{Aff}_2$.

Definition III.2.7 *Suppose* λ *is a positive real number. Then the function* $m_\lambda : \mathbb{C} \longrightarrow \mathbb{C}$, *defined by*

$$m_\lambda(z) = \lambda z$$

is called a **dilation** . *In vector notation,*

$$m_\lambda(\vec{z}) = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \vec{z}$$

Note: $m_\lambda \in GL_2(\mathbb{R})$ and $m_\lambda A = A m_\lambda$ for all $A \in GL_2(\mathbb{R})$ (as well as for all 2×2 matrices A.)

Lemma III.2.8 (Translation-conjugation Lemma)
If T_b *is a translation and* $g \in Aff_2$ *with* $g(z) = Az + c$ $(A \in GL_2(\mathbb{R}),\ z, b, c \in \mathbb{R}^2)$, *then*

$$g T_b g^{-1} = T_{Ab}.$$

If we denote $A = \pi(g)$, *as we shall in the next section, this becomes*

$$g T_b g^{-1} = T_{\pi(g)\cdot b}.$$

Proof:

$$\begin{aligned} gT_bg^{-1}(z) &= gT_b(A^{-1}(z-c)) \\ &= g(A^{-1}z - A^{-1}c \ + \ b) \\ &= A(A^{-1}z - A^{-1}c \ + \ b) + c \\ &= z - c + Ab + c \\ &= T_{Ab}(z) \end{aligned}$$ ∎

Corollary III.2.9

1. *If g is an isometry and T_b is a translation, then $gT_bg^{-1} = T_c$, where $|c| = |b|$.*
2. *If $b, c \in \mathbb{C}$ with $|b| = |c|$, then there is a rotation R with $RT_bR^{-1} = T_c$*
3. *If $b, c \in \mathbb{C}$ are non-zero numbers then there is a rotation R and a dilation m_λ such that*
$$(Rm_\lambda)\, T_b\, (Rm_\lambda)^{-1} \ = \ T_c.$$

Proof: 1. Let $c = Ab$, where $A = \pi(g)$ as in the Translation-conjugation Lemma. That lemma tells us that $gT_bg^{-1} = T_c$. But the fact that g is an isometry means that $A \in \mathrm{O}_2$ is also an isometry. Hence $|c| = |Ab| = |b|$, and the first assertion is proved.

2. First note that if $g \in \mathrm{O}_2 < \mathrm{Isom}\ \mathbb{C}$, then $\pi(g) = g$. Now if $|b| = |c|$ then there are real numbers $r \geq 0$, β, γ such that

$$b = r^{i\beta}, \qquad c = re^{i\gamma}.$$

Then the rotation R about the origin by $\gamma - \beta$ radians (which is an isometry, playing the role of $g \in \mathrm{O}_2$), given by $R = R_{\gamma-\beta}$ has

$$R(b) = c \quad \text{and so} \quad \pi(R)(b) = c. \quad \text{Thus} \quad RT_bR^{-1} = T_c.$$

3. Let $\lambda = \frac{|c|}{|b|}$. Then $|m_\lambda(b)| = |\lambda b| = |c|$ and $[m_\lambda \in GL_2(\mathbb{R}) \implies \pi(m_\lambda) = m_\lambda]$. Thus, by statement 2., there is a rotation R with

$$T_c = RT_{m_\lambda(b)}R^{-1} = RT_{\pi(m_\lambda)(b)}R^{-1} = Rm_\lambda\, T_b\, {m_\lambda}^{-1}R^{-1}.$$ ∎

Lemma III.2.10 *Any two reflections M_{L_1} and M_{L_2} are conjugate in Isom $\mathbb{C}$.*

Proof: We saw in Chapter I (I.4.17, I.4.19) that any reflection M_L can be conjugated by a rotation and translation to the reflection $M(z) = \overline{z}$. (in fact, if $g = T_{z_0}R_\eta$ takes L to the x-axis, then $M_L = gMg^{-1}$.) Therefore $M_{L_1} \sim M \sim M_{L_2}$. Since the relation of conjugacy in Isom $\mathbb{C}$ is transitive, we conclude that $M_{L_1} \sim M_{L_2}$. ∎

Lemma III.2.11 *For any $\theta \in \mathbb{R}$, $z_0, z_1 \in \mathbb{C}$ we have:*

1. *The rotations $R_{z_0,\theta}$, $R_{z_1,\theta}$ by θ radians are conjugate in Isom $\mathbb{C}$.*
2. *The rotations $R_{z_0,\theta}$ and $R_{z_1,-\theta}$ are conjugate in Isom $\mathbb{C}$.*

Proof: 1. As pointed out at the beginning of this section, our original definitions give

$$R_{z_0,\theta} \sim R_\theta \sim R_{z_1,\theta},$$

and the result follows because conjugacy is a transitive relation. Stated directly, if $g = T_{z_0-z_1}$ then $R_{z_0,\theta} = gR_{z_1,\theta}g^{-1}$. (Actually, any direct isometry with $g(z_1) = z_0$ will work as conjugating element.)

To prove 2. note that $R_\theta \sim R_{-\theta}$ by the reflection M, since

$$R_{-\theta}(z) = e^{-i\theta}z = \overline{e^{i\theta}\,\overline{z}} = M\,R_\theta\,M(z) = M\,R_\theta\,M^{-1}(z).$$

The result 2. now follows using 1. and transitivity. (Actually any opposite isometry g with $g(z_0) = z_1$ will work as conjugating element.) ∎

Conjugacy invariants

To begin to see why two isometries might *not* be conjugate, we give two lemmas. In each case we give an *invariant*, a number or property of an isometry which won't change if we conjugate it to another isometry.

The first of these two lemmas implies that *the property of an isometry being direct or opposite is an invariant of its conjugacy class in Isom $\mathbb{C}$.*:

Lemma III.2.12
If g_1 and g_2 are isometries which are conjugate in Aff_2, then g_1 is direct if and only if g_2 is direct.

Said otherwise, a rotation or translation cannot be conjugate in Aff_2 (so certainly not in Isom $\mathbb{C}$) to a reflection or glide reflection.

Proof: Suppose that, in vector notation, $g_i(\vec{z}) = A_i\vec{z} + \vec{b_i}$, $i = 1, 2$. Then

$$g_i \text{ is direct} \iff A_i \text{ is a rotation matrix (possibly } A_i = I) \iff \det(A_i) = +1.$$

Suppose that $f \in \mathrm{Aff}_2$, say with $f(\vec{z}) = C\vec{z} + \vec{d}$, and with $fg_1f^{-1} = g_2$. Then straightforeward calculation gives $fgf^{-1}(\vec{z}) = CAC^{-1}(z) + \vec{e}$ for some vector $\vec{e}$. However, the determinant function is multiplicative and $\det C \neq 0$, since [$f \in \mathrm{Aff}_2 \implies C$ is invertible]. Therefore

$$\det(ACA^{-1}) = (\det A)(\det B)(\det C)^{-1} = \det A.$$

Therefore $g_2 = fg_1f^{-1}$ is direct if and only if g_2 is direct ∎

The second of these two lemmas implies that *the shape of the fixed point set of an isometry g (and in particular the number of elements of the fixed point set) is an invariant of the conjugacy class of g in Isom $\mathbb{C}$.*

Lemma III.2.13 *If G is a transformation group (for example Isom $\mathbb{C}$), and g_1 and g_2 are conjugate in G, then there is an element $f \in G$ with $Fix(g_2) = f(Fix(g_1))$.*

Proof: If $g_1 \sim g_2$ in G, then there exists $f \in G$ with $f\, g_1\, f^{-1} = g_2$. Thus

$$\begin{aligned} z \in \mathrm{Fix}(g_1) &\iff g_1(z) = z \iff g_1\, f^{-1}\big(f(z)\big) = z \\ &\iff f\, g_1\, f^{-1}\big(f(z)\big) = f(z) \iff g_2\big(f(z)\big) = f(z) \\ &\iff f(z) \in \mathrm{Fix}(g_2). \end{aligned}$$

∎

Corollary III.2.14

1. *A non-zero translation and a non-trivial rotation cannot be conjugate in Aff_2 (and consequently certainly not in Isom $\mathbb{C}$.)*
2. *A reflection and a glide reflection cannot be conjugate in Aff_2 (and consequently certainly not in Isom $\mathbb{C}$).*

Proof: The fixed point set of a non-zero translation is the empty set, while the fixed point set of a non-trivial rotation is a set consisting of a single point. The elements of $\mathrm{Aff}_2$are bijective functions. So no $f \in \mathrm{Aff}_2$ carries a set of the form $\{z_0\}$ to the empty set. So the translation and rotation cannot be conjugate in Aff_2. Similarly, the fixed point set of a glide reflection is empty, while the fixed point set of a reflection is a line. Thus a glide reflection and a reflection cannot be conjugate in Aff_2. ∎

The Lemmas and Corollaries above supply the information by which we may now classify isometries up to conjugation. The next theorem says exactly which isometries are conjugate to each other. We leave the remaining details of the proof to the reader (Exercise 5).

Theorem III.2.15

1. *The conjugacy class in Isom $\mathbb{C}$ of a translation, rotation, reflection, or glide-reflection consists entirely of, respectively, translations, rotations, reflections, glide-reflections. Thus, isometries of different types are not conjugate in Isom $\mathbb{C}$.*
2. *(a) The identity id is conjugate only to itself.*
   (b) Two translations $g_1 = T_{b_1}$, $g_2 = T_{b_2}$ are conjugate in Isom $\mathbb{C}$ if and only if $|b_1| = |b_2|$.
   (c) Two rotations $g_1 = R_{z_1,\theta_1}$, $g_2 = R_{z_2,\theta_2}$ are conjugate in Isom $\mathbb{C}$ if and only if $\theta_1 = \pm\theta_2$.
   (d) Any two reflections $g_1 = M_{L_1}$, $g_2 = M_{L_2}$ are conjugate. in Isom $\mathbb{C}$.
   (e) Two glide-reflections $g_1 = T_{c_1} \circ M_{L_1}$, $g_2 = T_{c_2} \circ M_{L_2}$, where c_i is parallel to L_i, $(i = 1, 2)$, are conjugate in Isom $\mathbb{C}$ if and only if $|c_1| = |c_2|$.

∎

If we consider conjugacy of isometries inside Aff_2, then the corollaries above give

Theorem III.2.16 *If g_1 and g_2 are isometries which are conjugate in Aff_2 then they are of the same type (translation, rotation, reflection, glide reflection).* ■

Every dilation m_λ belongs to Aff_2. We can extend the preceding theorem to

Corollary III.2.17 *Suppose $\lambda > 0$ and g is an isometry of the plane. Then $m_\lambda g\, {m_\lambda}^{-1}$ is an isometry of the plane of the same type as g.*

Proof: If $z, w \in \mathbb{C}$ we have

$$\begin{aligned} |m_\lambda g {m_\lambda}^{-1}(z) - m_\lambda g {m_\lambda}^{-1}(w)| &= |\lambda \cdot g({m_\lambda}^{-1}(z)) - \lambda \cdot g({m_\lambda}^{-1}(w))| \\ &= \lambda |g(\lambda^{-1} z) - g(\lambda^{-1} w)| \\ &= \lambda |\lambda^{-1} z - \lambda^{-1} w| \qquad (\text{since } g \text{ is an isometry}) \\ &= \lambda \lambda^{-1} |z - w| = |z - w| \end{aligned}$$

Therefore $m_\lambda g {m_\lambda}^{-1}$ is an isometry. The result now follows from the preceding theorem. ■

Conjugate subgroups and congruent figures

The fact that conjugation by an isometry f transfers the geometric action of an isometry g on a plane figure P to a geometric action of fgf^{-1} on $f(P)$ is now brought to bear in comparing the symmetry groups of congruent figures. We have already defines what it means for two elements of a group to be conjugate. Now we are interested in the conjugacy of subgroups. The point of what follows is to note that *congruent figures have conjugate symmetry groups* and to explain the algebraic and geometric consequences of the conjugacy of two groups.

Definition III.2.18 *Two subgroups G_1, G_2 of* Isom$\mathbb{C}$ *are called* conjugate in Isom($\mathbb{C}$) *if there is an isometry f such that*

$$G_2 = fG_1f^{-1} = \{fg_1f^{-1} |\ g_1 \in G_1\}.$$

In this setting, we say that f conjugates G_1 to G_2 in Isom $\mathbb{C}$ *, or that f is a* conjugacy between G_1 and G_2.

For example, the group $\mathcal{R}_0$ of rotations about the origin is conjugate in Isom $\mathbb{C}$ to the group $\mathcal{R}_{z_0}$ of rotations about the point z_0 since the translation T_{z_0} conjugates $\mathcal{R}_0$ to $\mathcal{R}_{z_0}$.

If two isometry groups are conjugate in Isom $\mathbb{C}$ then they are algebraically isomorphic in a manner which takes particular account of the geometry of the individual isometries:

Proposition III.2.19 *Suppose that $f \in isom$ conjugates G_1 to G_2. Let the function*
$c_f : G_1 \longrightarrow G_2$ be defined by

$$c_f(g) = fgf^{-1}, \qquad \text{for all } g \in G_1$$

Then

1. *c_f is an isomorphism of groups*
2. *If $g \in G_1$ then $c_f(g)$ is an isometry of the same type (translation, rotation, reflection or glide reflection) as g is.*

Example III.2.20 *Two subgroups of* Isom($\mathbb{C}$) *can be isomorphic without the existence of an isomorphism which preserves isometry type. This is clearly the case for the groups $G_1 = \{id,\ M\}$ and $G_2 = \{id,\ R_\pi\}$, both of which are isomorphic to $\mathbb{Z}_2$. By the Proposition, these groups are isomorphic but not conjugate.*

Proof of Proposition III.2.19. To prove that c_f is an isomorphism, first note that c_f maps G_1 to G_2, and is in fact onto. This is because, by definition, the hypothesis that f conjugates G_1 to G_2 means that $c_f(G_1) = G_2$.

c_f is one-one because $c_f(g) = c_f(h) \implies fgf^{-1} = fhf^{-1} \implies g = h$.

Finally, f is an isomorphism because

$$c_f(gh) = f(gh)f^{-1} = (fgf^{-1})\,(fhf^{-1}) = c_f(g) \cdot c_f(h).$$

By Theorem III.2.15, c_f must take each isometry in G_1 to an isometry of the same type in G_2. ∎

Definition III.2.21 [1] *Two figures* P_1, P_2 *(i.e. two subsets of* $\mathbb{C}$*) are called* congruent *if there exists an isometry* $f \in \mathrm{Isom}(\mathbb{C})$ *such that* $f(\mathrm{P}_1) = \mathrm{P}_2$.

Example III.2.22 *(Left as Exercise 6.)*

1. *Any two straight lines are congruent.*
2. *Two circles are congruent if and only if they have the same radius.*
3. *Two triangles are congruent if (in some ordering) the sides of the first triangle have the same lengths as the sides of the second. (This is Euclid, Book I, Proposition 4.)*

Example III.2.23 *In Lemma III.2.13 we showed that if* g_1 *and* g_2 *are conjugate in Isom* $\mathbb{C}$ *then their fixed-point sets are congruent.*

In Chapter I, we introduced the group $\mathrm{Sym(P)}$ of a plane figure P, consisting of all elements of $\mathrm{Isom}(\mathbb{C})$ preserving the figure P. For example, the dihedral group D_n introduced in Section II.3 is the group of symmetries preserving a regular n-gon P_1 centered at the origin, with vertices at $e^{2\pi ij/n}, j = 0, ..., n-1$. Now imagine a regular n-gon P_2 congruent to P_1, but *drawn somewhere else in the plane*, say centered at $2 + 2i$ instead and rotated a bit (see Figure 3).

Obviously, the groups $\mathrm{Sym}((\mathrm{P}_1)$ and $\mathrm{Sym}(\mathrm{P}_2)$ are different subgroups of $\mathrm{Isom}(\mathbb{C})$. For example, $\mathrm{Sym}(\mathrm{P}_1)$ contains non-trivial rotations about the origin, but $\mathrm{Sym}(\mathrm{P}_1)$ does not. But certainly we feel that the *structure* of these two groups should be the same. Not only should they be isomorphic, but there should be an isomorphism

[1] This is very close to the original Euclid, in which *Common Notion 4* translates as: *Things which coincide with one another are equal to one another.* This was classically interpreted to mean that one could prove two figures equal by superimposing one on the other. See Sir Thomas Heath, *Euclid, The Elements*, 3rd Ed. (1947), p. 224.

Figure 3. A congruence f sending the square P_1 to the square P_2, and symmetries $g_1 \in \mathrm{Sym}(\mathrm{P}_1),\ g_2 = fg_1f^{-1} \in \mathrm{Sym}(\mathrm{P}_2)$.

which reflects the geometric fact that these groups come from congruent figures. For example, each rotation about the origin by $2\pi j/n$ radians in $\mathrm{Sym}(\mathrm{P}_1)$ should be matched by the isomorphism with a rotation about the point $2+2i$ by $2\pi j/n$ radians in $\mathrm{Sym}(\mathrm{P}_2)$.

Where will we get such an isomorphism?

From Proposition III.2.19 we know that we will get just such an isomorphism which respects the geometry if we can prove that the symmetry groups are conjugate. This is what we now prove:

Theorem III.2.24 *If plane figures $\mathrm{P}_1, \mathrm{P}_2$ are congruent, then their symmetry groups $\mathrm{Sym}(\mathrm{P}_1), \mathrm{Sym}(\mathrm{P}_2)$ are conjugate.*

Proof: Since P_1 and P_2 are congruent, there is an isometry $f : \mathbb{C} \to \mathbb{C}$ such that $f(\mathrm{P}_1) = \mathrm{P}_2$. Let $G_i = \mathrm{Sym}(\mathrm{P}_i),\ i = 1, 2$. Consider the function

$$c_f : G_1 \longrightarrow \mathrm{Isom}\ \mathbb{C}$$

defined by

$$c_f(g_1) = fg_1f^{-1}.$$

The function fg_1f^{-1} is an element of G_2, since f^{-1} sends P_2 to P_1, g_1 sends P_1 to itself, and f sends P_1 to P_2; see Figure 3.

Moreover, any element g_2 of G_2 is the image of some g_1 under c_f. To see this, note that the same reasoning as in the previous paragraph shows that $f^{-1}g_2f \in G_1$. Thus

$$g_2 = f(f^{-1}g_2f)f^{-1} = c_f(g_1), \text{where } \; g_1 = f^{-1}g_2f.$$

Therefore $G_2 = fG_1f^{-1}$; so f is a conjugacy between $\mathrm{Sym}(\mathrm{P}_1)$ and $\mathrm{Sym}(\mathrm{P}_2)$. ∎

Note: There are figures P_1 and P_2 whose symmetry groups are isomorphic while the groups are not conjugate. By the theorem above, this can only happen if the figures are not congruent. See Exercise 7.

Exercises III.2

1. Complete the proof that conjugacy is an equivalence relation (Lemma III.2.3).

2. Show that the translation $T_1(z) = z + 1$ is conjugate to the translation $T_i(z) = z + i$ by the rotation $R_{\pi/2}$.

3. Show that the reflection M across the x-axis and the reflection $M_{\frac{\pi}{2}}$ across the y-axis are conjugate in the group Isom $\mathbb{C}$, but are not conjugate in the subgroup $D_2 < Isom\ \mathbb{C}$.

4. Decide which of the following isometries are conjugate to each other in Isom $\mathbb{C}$:

   a) $f_1(z) = z + \sqrt{2}$ **b)** $f_2(z) = iz + 7$ **c)** $f_3(z) = i\overline{z} + 7$
   d) $f_4(z) = i\overline{z} + 7 - 7i$ **e)** $f_5(z) = z + 1 + i$ **f)** $f_6(z) = -iz + 4 - 3i$
   g) $f_7(z) = \big(\cos(\frac{\pi}{7}) + i\ \sin(\frac{\pi}{7})\big)\overline{z}$ **h)** $f_8(z) = \overline{z} - 7.$

5. Write out the complete proof of the classification of isometries by conjugation in Isom $\mathbb{C}$ (Theorem III.2.15).

6. Construct isometries which show that the congruences from classical geometry listed in Example III.2.22 do indeed hold.

7. Exhibit figures P_1 and P_2 whose symmetry groups are isomorphic while the groups are not conjugate in Isom $\mathbb{C}$.

8. **Definition:**
   If $X \subset \mathbb{C}$ *and* $g \in$ Isom $\mathbb{C}$ *then* X *is* invariant under g *if* $g(X) = X$.

(Individual elements of X need not be fixed by g. For example, the axis of a glide reflection is invariant under the glide reflection. Also $\mathrm{Fix}(g)$ is invariant under g for any $g \in \mathbb{C}$.)

Suppose that $f,\ g_1,\ g_2$ are isometries such that f conjugates g_1 to g_2. Prove that if X is invariant under g_1 then $f(X)$ is invariant under g_2. [Compare this with Lemma III.2.13.]

9. Two elements x, y of a group G *commute* if $xy = yx$.
Prove that x and y commute

$\iff$ x conjugates y to y.

$\iff$ y conjugates x to x

10. Using the definition of the preceding exercise, prove the following statements. (Exercise 8 may be useful.)

    (a) A non-zero translation T_b and a non-zero rotation $R_{z_0,\theta}$ never commute.

    (b) Two glide reflections commute if and only if their axes are the same.

    (c) A reflection and a glide reflection commute if and only if the line of reflection is equal to the glide line.

    (d) A non-trivial rotation and a reflection never commute.

11. Suppose that $f,\ g,\ g_1,\ g_2 \in \mathrm{Isom}\,\mathbb{C}$ and $z \in \mathbb{C}$. Prove that:

    (a) The g-orbit of z is invariant under g.

    (b) If f conjugates g_1 to g_2 and if $z \in \mathbb{C}$, then
    $f(\text{the } g_1\text{-orbit of } z) = \text{the } g_2\text{-orbit of } f(z)$.

    [The g-orbit of $z \equiv \{g^n(z) \,|\, n \in \mathbb{Z}\}$ = the orbit of z under $\langle g \rangle$.]

12. Let G_1 be the group generated by a nonzero translation and G_2 the group generated by a glide reflection. Show that G_1 and G_2 are isomorphic.

    For the groups $G_1 = \langle z \mapsto z + 1 \rangle$ and $G_2 = \langle z \mapsto \overline{z} + 1 \rangle$, draw the orbit of the point i. Would you say that these orbits are similar, geometrically? Are the groups G_1 and G_2 conjugate?

13. Repeat the previous problem with the same G_1, and with $G_2 = \langle R_\theta \rangle$, where $\theta = 2\pi s$ and s is an irrational number between zero and one.
(Hint: find a condition on θ for the order of $R_{0,\theta}$ to be finite.)

14. Let T_{b_1}, T_{b_2} be two nonzero translations. When is $G = \langle T_{b_1}, T_{b_2} \rangle$ isomorphic to $\mathbb{Z} \times \mathbb{Z}$, the product of two copies of $\mathbb{Z}$?

15. Let $b_1 = 1, b_2 = 1 + i$ be as in the previous problem. Show that then G is isomorphic to $\mathbb{Z} \times \mathbb{Z}$, and draw the orbit of the point 0 under G. Repeat this with $b_1 = 1, b_2 = \sqrt{2}$. Compare your findings. (You may assume that $\sqrt{2}$ is irrational.) Verify that the generators in the second case are individually conjugate to the generators in the first case. Nonetheless, these groups are very different geometrically. Why is this not a violation of the fact that two conjugate subgroups of Isom $\mathbb{C}$ are the same, just viewed in different coordinates?

16. Let $b_1,\ b_2$ be two complex numbers which, as vectors in $\mathbb{R}^2$, are linearly independent. Prove that if an isometry f commutes with *both* T_{b_1} and T_{b_2}, then f itself is a translation.

17. Prove that the subgroup of Isom($\mathbb{C}$) generated by two nontrivial rotations $R_{z_0,\alpha},\ R_{z_1,\beta}$ with different centers contains a nonzero translation.
Hint: Show that the group contains a rotation of the form $R_{z_2,-\beta}$ where $z_2 \neq z_1$.

III.3 The point map $\pi : \mathrm{Isom}(\mathbb{C}) \to O_2$

We now consider a function π from the group $\mathrm{Isom}(\mathbb{C})$ to the group O_2 which captures how isometries change the directions of lines. This will be of particular use when we classify the wallpaper groups in Chapter V. The point map π has the important property: $\pi(f \circ g) = \pi(f)\pi(g)$, i.e. is a *homomorphism* (see the Proposition below). The proof uses in a crucial way the fact that an isometry f is a *linear transformation* $\vec{z} \mapsto A\vec{z}$ followed by a translation $\vec{z} \mapsto \vec{z} + \vec{b}$.

In Chapter IV we will study homomorphisms between arbitrary groups.

Notation: If A is in O_2 let f_A denote the isometry given by $f_A(\vec{z}) = A\vec{z}$.

Proposition III.3.1 (The point map is a homomorphism)
Let the function $\pi : \mathrm{Isom}(\mathbb{C}) \to O_2$ be given by

$$\pi(f) = A \qquad \text{if} \qquad f(\vec{z}) = (A\vec{z} + \vec{b})$$

Then

- π is a surjective mapping of $\mathrm{Isom}(\mathbb{C})$ onto O_2.
- $\pi(f \circ g) = \pi(f)\pi(g)$. ($\pi$ is a *homomorphism* – see Chapter IV.)

We call π the *point map* of $\mathrm{Isom}(\mathbb{C})$, since the image group O_2 is, by Corollary **??**, isomorphic to the stabilizer $\mathrm{Isom}_0(\mathbb{C})$ of the point which is the origin. The isomorphism is given in the present notation by $A \mapsto f_A$.

Proof: Note that π is a well-defined function: Since each isometry f can be written in only one way in this form (Theorem III.1.4) there is only one possible value of $\pi(f) = A$.

π is onto since: $A \in O_2 \implies A = \pi(f_A)$.

To see that π is a homomorphism, suppose that $f(\vec{z}) = A\vec{z} + \vec{b}$ and $g(\vec{z}) = A_1\vec{z} + \vec{b}_1$. Then

$$(f \circ g)(\vec{z}) = A(A_1\vec{z} + \vec{b}_1) + \vec{b} = AA_1\vec{z} + (A\vec{b}_1 + \vec{b}).$$

Therefore

$$\pi(f \circ g) = AA_1 = \pi(f)\pi(g).$$

■

Here, we used the property that matrix multiplication acting on vectors is linear to get $A(A_1\vec{z} + \vec{b}_1) = AA_1\vec{z} + A\vec{b}_1$.

Remark III.3.2 *If we identify* O_2 *with* $\mathrm{Isom}_0(\mathbb{C})$*, using Corollary* **??***, then we may describe* π *in complex notation as the function* $\pi : \mathrm{Isom}\ \mathbb{C} \longrightarrow \mathrm{Isom}_0(\mathbb{C})$ *given by:*

$$\textit{If} \quad f(z) = az + b \quad \textit{then} \quad \pi(f) = z \mapsto az$$

$$\textit{If} \quad f(z) = a\overline{z} + b \quad \textit{then} \quad \pi(f) = z \mapsto a\overline{z}$$

However, the vector space notation and fact that π *maps into* O_2 *seems to be more traditional here.*

We now give several interesting interpretations of the the image $\pi(f)$ of an isometry $f(\vec{z}) = A\vec{z} + \vec{b}$.

1. $\pi(f)$ *is a (possibly trivial) rotation about the origin or a reflection across a line through the origin.* $\pi(f)$ *records how* directions of lines *in the plane are changed by the function* f*.*

   To be precise, $\mathrm{O}_2 = \pi(\mathrm{Isom}\ \mathbb{C})$ consists of matrices which represent such rotations or reflections.

   The quickest way to see the effect of $\pi(f)$ on directions is to note that $A = \pi(f) = T_{-\vec{b}} \circ f$, where $T_{-\vec{b}}$ is translation by the vector $-\vec{b}$. Since a translation moves every vector parallel to itself, it does not change the direction of any line. Thus $\pi(f)$ changes the direction of any given line in exactly the same way that f does. Since $\pi(f)$ is a matrix representing a linear transformation, we can check what f does to any direction by checking what $\pi(f)$ does to the line through the origin in that direction.

Figure 4. A translation seen from high above the plane. Note that the perceived effect of a translation $T_{\vec{b}}$ is negligible if the height D is large relative to $|\vec{b}|$.

2. $\pi(f)$ *represents how an observer standing arbitrarily far above the plane sees an isometry.* Imagine an observer of height 1 standing directly above the origin on a clear plastic floor parallel to the plane, and whose head is D units above the plane $\mathbb{C}$. (See Figure 4.) Let $f(\vec{z}) = A\vec{z} + \vec{b}$ be an isometry of the plane. What does she see on the clear plastic floor beneath her?

   By similar triangles, she sees any point $\vec{z}$ in the plane $\mathbb{C}$ beneath her as the point $\vec{w} = \vec{z}/D$ on the clear floor. So she sees that the point $\vec{w} = \vec{z}/D$ is sent to $(A\vec{z}+\vec{b})/D = A(\vec{z}/D)+\vec{b}/D = A\vec{w}+\vec{b}/D$. Hence, *if D is very large relative to $|\vec{b}|$, then $\vec{b}/D$ is almost $\vec{0}$ and the image of $\vec{w}$ is very nearly $A\vec{w} = \pi(f)(\vec{w})$.* See Figure 4.

3. (Optional, for those who have had advanced calculus.) Thinking of f as a differentiable function from $\mathbb{R}^2 \to \mathbb{R}^2$, $\pi(f)$ is the *derivative* or *Jacobian* of f. That is, if $\vec{v}$ is a tangent vector to some point z in the plane, then the vector $\pi(f)\vec{v}$ based at $f(z)$ is the image of $\vec{v}$ under the derivative df. The homomorphism property follows from the Chain Rule of advanced calculus.

In the first remark in III.3.2 we noted that $\pi(f) = T_{-\vec{b}} \circ f$. Then $f = T_{\vec{b}} \circ \pi(f)$ and we get the following theorem (switching from vector to complex notation).

Theorem III.3.3 *Suppose that* $f = T_b \circ \pi(f)$.

1. $\pi(f) =$ *a rotation by* $\theta \iff f =$ *a rotation by* θ.
2. *If* $\pi(f) =$ *the reflection* M_L *in a line* L *through the origin, write* $b = u + v$ *where* $u \parallel L,\ v \perp L$. *Then*
   (a) f *is a reflection* $\iff f = T_v \circ M_L$ *where* $v \perp L$
   (b) f *is a glide reflection* $\iff$ $f = T_u \circ (T_v \circ M_L)$ *where* $v \perp L$ *and* $u \parallel L,\ u \neq 0$.

Proof: *The proof is left to the reader.* ∎

Exercises III.3

1. What type of isometry is $\pi(g)$ if g is a translation? rotation? reflection? glide reflection?
2. Suppose that $G < \mathrm{Isom}\,\mathbb{C}$ is a subgroup of $\mathrm{Isom}(\mathbb{C})$ consisting totally of direct isometries (*i.e.*, a subgroup of $\mathrm{Isom}^+(\mathbb{C})$). What can you say about $\pi(G)$?
3. Suppose that g is a glide reflection. How many elements does the group $\pi(\langle\, g \,\rangle)$ have? What are these elements?
4. (a) If G is a subgroup of $\mathrm{Isom}(\mathbb{C})$ such that $\pi(G)$ contains a rotation, does it follow that G contains a rotation?
   (b) If G is a subgroup of $\mathrm{Isom}(\mathbb{C})$ such that $\pi(G)$ contains a reflection, does it follow that G contains a reflection?
   (c) Can $\pi(G)$ contain a glide reflection?
5. Give an example of a group G such that $\pi(G)$ contains
   (a) reflection in the real axis, but G does not.

(b) rotation about the origin by $\pi/2$, but G does not.

(c) a non-trivial finite group G such that $\pi(G) \cap G = \{\mathrm{id}_{\mathbb{C}}\}$.

6. If $\pi(G)$ contains an element of infinite order, must G contain one as well? If G contains an element of infinite order, must $\pi(G)$ contain one as well?

7. If G contains an element of finite order n, must $\pi(g)$ contain an element of order n?

8. Find the point groups $\pi(G_i)$ of the symmetry groups G_i of the six figures P_i of Figures I.1-I.6.
Hint: Look for generators of G_i, and find the images of these under π to find generators for $\pi(G_i)$. We will justify this in Exercise IV.I.3.)

III.4 Finite Subgroups of Isom(ℂ)

What are the finite subgroups of Isom(ℂ)?

Let G be such a group. Let us write $G = \{\underbrace{g_1}_{\text{id}}, g_2, \ldots, , g_n\}$.

Clearly G cannot contain a non-zero translation T_b. For T_b has infinite order and, by the closure axiom, no finite group can contain an element of infinite order — since it would be forced to contain all of the powers of this element.

It follows that G also could not contain a glide reflection g, since g^2 would be a translation in G.

So G contains only rotations of finite order and reflections. Thus each g_i fixes a point z_i (and even a whole line L_i if g_i is a reflection). Surprisingly, there is one point which they all fix in common. So all the rotations are rotations about the same point, and all the reflections have their axes going through this point.

We will use vector notation to prove this, because then we won't have to write direct and opposite isometries with different formulas (using z in some cases and $\overline{z}$ in others); rather, we can assume that each $g_i \in G$ has a formula of the form

$$g_i(\vec{z}) = A_i\vec{z} + \vec{b}_i, \quad A_i \in \mathrm{O}_2,\ \vec{b}_i \in \mathbb{R}^2.$$

The center of mass of an orbit

Proposition III.4.1 *Suppose that* $G = \{id, g_2, \ldots, , g_n\}$ *is a finite subgroup of* Isom(ℂ) *of order* n. *If* $\vec{a} \in \mathbb{C}$, *set*

$$\vec{a}_0 = \frac{1}{n}\sum_{i=1}^{n} g_i(\vec{a}) \qquad \textit{(the "center of mass" of the orbit of } \vec{a}\textit{).}$$

Then $g_j(\vec{a}_0) = \vec{a}_0$ *for all* $g_j \in G$.

Proof:

$$
\begin{aligned}
g_j(\vec{a}_0) &= g_j\left(\frac{1}{n}\sum_{i=1}^{n} g_i(\vec{a})\right) \\
&= A_j \cdot \left(\frac{1}{n}\sum_{i=1}^{n} g_i(\vec{a})\right) + \vec{b}_j \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} A_j \cdot g_i(\vec{a})\right) + \frac{1}{n}\underbrace{\sum_{i=1}^{n} b_j}_{n\cdot b_j} \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} A_j \cdot [\, g_i(\vec{a}) + \vec{b}_j \,]\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} g_j g_i(\vec{a})\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} g_i(\vec{a})\right) \qquad \text{since } \{g_1,\ldots,g_n\} = \{g_j g_1,\ldots,g_j g_n\}, \\
&\qquad\qquad \text{by Claim 1, proof of II.4.18} \\
&= \vec{a}_0
\end{aligned}
$$

■

The only examples

Examples of finite groups to which the above theorem applies are

1. $G = \mathcal{R}_n = \langle R \rangle$, the group of order $n \geq 1$ of rotations about the origin generated by $R = R_{2\pi/n}$.
2. $G = D_m$, the dihedral group of order $n = 2m \geq 2$. (See Section II.3)

We now show that, up to conjugacy, these are the only examples.

Theorem III.4.2 *If G is a finite subgroup of* Isom(ℂ) *of order n then*

1. *G is conjugate in* Isom(ℂ) *to $\mathcal{R}_n$ if G contains no reflections.*

2. *G is conjugate in* Isom(ℂ) *to D_m ($n = 2m$) if G contains a reflection.*

Proof: If $G = \{id\} = \mathcal{R}_0$, the theorem is trivial. Otherwise, suppose that G has order $n > 1$. By the preceding proposition there is a common fixed point z_0 of all the elements of G. Then T_{-z_0} conjugates G to $G_0 = T_{-z_0}\, G\, T_{-z_0}^{-1}$ (Definition III.2.18). By III.2.13, $g(0) = 0$ for all $g \in G_0$ and, by III.2.19, $G \cong G_0$ under an isomorphism which takes rotations to rotations and reflections to reflections.

If G, and thus G_0, contains no reflections, let R_θ be the rotation in G_0 by the smallest positive angle θ. This exists because G_0 is a finite group of order $n > 1$. If $G_0 \neq \langle\, R_\theta \,\rangle$ then there is a smallest angle η, $0 < \eta < 2\pi$, such that $R_\eta \in G_0$ but $R_\eta \notin \langle\, R_\theta \,\rangle$. If j is the largest positive integer such that $j\theta < \eta$ then $R_\eta\, R_\theta^{-j} = R_{\eta - j\theta}$ is a rotation in G_0 by a positive angle less than θ. This contradicts the choice of θ. Thus, we must in fact have G conjugate to $G_0 = \langle\, R_\theta \,\rangle = \langle\, R_{2\pi/n} \,\rangle = \mathcal{R}_n$.

If G, and thus G_0, contains at least one reflection, let $\mathcal{R}$ be the subgroup of G_0 consisting of rotations. These are all of the direct isometries in G_0, but not all of G_0. Thus, by Proposition III.1.7, $\mathcal{R}$ has index two in G_0. The argument in the preceding paragraph tells us then that $\mathcal{R} = \mathcal{R}_m = \langle\, R_{2\pi/m} \,\rangle$. Therefore, if M_L is a reflection in G_0 and if we write $R = R_{2\pi/m}$ then we have

$$G_0 = \mathcal{R}_m \sqcup \mathcal{R}_m \cdot M_L = \{\mathrm{id},\ R, R^2, \ldots, R^{m-1}, M_L, RM_L, \ldots, R^{m-1}M_L\}.$$

This looks very much like the dihedral group except that the line L (which goes through the origin because $M_L(0) = 0$) is playing the role of the x-axis.

If L makes an angle α with the x-axis, we note that

1. $R_{-\alpha}\, M_L\, R_\alpha = M$ (reflection across the x-axis).

2. $R_{-\alpha} R^j R_\alpha = R^j$

Thus, if we conjugate G_0 by $R_{-\alpha}$, we get

$$\begin{aligned}(R_{-\alpha}T_{-z_0})\,G\,(T_{z_0}R_{\alpha}) &= R_{-\alpha}\,G_0\,R_{\alpha}\\ &= \{\mathrm{id},\ R, R^2, \ldots, R^{m-1}, M, RM, \ldots, R^{n-1}M\}\\ &= D_m.\end{aligned}$$

Therefore G is conjugate to the dihedral group D_m, as claimed. ∎

Corollary III.4.3 [Finite subgroups of O_2]
If G is a finite subgroup of O_2 which contains n elements then G is conjugate in O_2 to $\mathcal{R}_n$ if G contains no reflections and is conjugate in O_2 to $D_m\,(n = 2m)$ if G contains a reflection.

Proof: Since O_2 is contained in Isom $\mathbb{C}$ the preceding theorem tells us that G is conjugate in Isom $\mathbb{C}$ to $\mathcal{R}_n$ if G contains no reflections and is conjugate to $D_m\,(n = 2m)$ if G contains a reflection. The proof of III.4.2 uses conjugation by T_{-z_0}, where z_0 is the common fixed point of all the elements of the finite group of isometries in question and (in the case where G contains a reflection) conjugation by a rotation about the origin. But here, $G < \mathrm{O}_2$, so this fixed point is $z_0 = 0$. The result is that, under our current hypothesis, the proof of III.4.2 gives the desired conjugation not only in Isom $\mathbb{C}$, but in O_2. ∎

Exercises III.4

1. Prove or disprove:
   If $G <$ Isom($\mathbb{C}$) is finite, then $\pi : G \to \pi(G)$ is an isomorphism.

2. Suppose that G_1 and G_2 are finite subgroups of Isom($\mathbb{C}$) of order n with $\pi(G_1) \cong \pi(G_2)$. For which integers n, if any, can we guarantee that G_1 and G_2 are conjugate in Isom($\mathbb{C}$)?

Chapter IV

Homomorphisms

IV.1 Homomorphisms: Definition and examples

In Chapter III we saw that the function $\pi : \mathrm{Isom}(\mathbb{C}) \to O_2$ given by

$$\pi(f) = A \qquad \text{if} \qquad f(\vec{z}) = A\vec{z} + \vec{b}$$

had the key property

$$\pi(f \circ g) = \pi(f)\pi(g).$$

One way of thinking about this is that the function π turns the group operation in $\mathrm{Isom}(\mathbb{C})$ (composition of functions) into the group operation in O_2 (matrix multiplication or, equivalently, composition of linear transformations), maintaining the pattern of multiplication. We've seen this before when we introduced the notion of *isomorphism* between two groups as a *homomorphism* which is also a bijection. However, π sends translations to the identity element, so π is not one-to-one and is therefore not an isomorphism. It is the basic example of a homomorphism which we will use in our investigation of symmetry groups.

We recall Definition II.2.3. Let G, H be groups.

Definition. *A function* $\phi : G \longrightarrow H$ *is a* **homomorphism** *if*

$$\phi(ab) = \phi(a)\phi(b) \qquad \textit{for all } \ a, b \ \in G.$$

■

A word about notation: here, for $a, b \in G$, we have written ab for the product of a and b in G, and $\phi(a)\phi(b)$ for the product of $\phi(a)$ and $\phi(b)$ in H. Just as is the case for isomorphisms, it is important to remember that the product operations in H may be quite different from those in G, though no difference in notation is used.

Examples:

1. An isomorphism $\phi : G \to H$ is a special kind of homomorphism, namely, one where the function ϕ is a bijection.

2. The determinant function $\det : \mathrm{O}_2 \to \{-1, +1\}$ is a homomorphism from the group of two-by-two orthogonal matrices to the multiplicative group $\{-1, +1\}$. Why? First, if $A \in \mathrm{O}_2$ then $AA^t = I$ (Exercise III.2.5). By properties of determinants, $\det(AB) = \det(A)\det(B)$, so $1 = \det(I) = \det(A)\det(A^t) = \det(A)^2$ and hence $\det(A) \in \{-1, +1\}$. Thus ϕ sends O_2 to $\{-1, +1\}$ and satisfies $\det(AB) = \det(A)\det(B)$.

3. The sign function $\sigma : S_n \to \{-1, +1\}$ is a homomorphism from the symmetric group on n letters to the multiplicative group $\{-1, +1\}$; this is the content of Lemma II.8.13.

4. The complex exponential function $\exp : (\mathbb{R}, +) \to (S^1, \cdot)$ given by $\exp(\theta) = e^{i\theta}$ is a homomorphism from the additive group of real numbers to the multiplicative group of unit modulus complex numbers. This follows since $e^{i(\theta_1+\theta_2)} = e^{i\theta_1}e^{i\theta_2}$ for all $\theta_1, \theta_2 \in \mathbb{R}$.

5. The function $\phi : (\mathbb{Z}, +) \to (\mathbb{Z}_n, +_n)$ given by $\phi(a) = a$ modulo n is a homomorphism from the additive group of the integers to the additive group of the integers modulo n. This follows since $a + b$ modulo n is the same as a (modulo n) $+_n$ b (modulo n); see Example II.1.8.

6. Let G be any group. The function $l : G \to S_G = H$, the set of bijections of the underlying set of G, given by $l_g(x) = gx$, is a homomorphism. This is the content of the proof of Claim 2 in the proof of Cayley's Theorem in II.4.18.

7. Suppose V, W are vector spaces and $T : V \to W$ is a linear transformation. Then in particular $T(\vec{v}_1 + \vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2)$ for all $\vec{v}_1, \vec{v}_2 \in V$. So T is a

homomorphism from the group V to the group W, where the operations in both are vector addition.

A homomomorphism $\phi : G \longrightarrow H$ preserves basic relationships. In the following proposition we recall the assertions of Proposition II.2.4 and we add the fact that $\phi(G)$ is a subgroup of H.

Proposition IV.1.1 *Suppose that G and H are groups with identity elements e_G and e_H respectively. If $\phi : G \longrightarrow H$ is a homomorphism then*

1. $\phi(e_G) = e_H$.

2. $\phi(g^{-1}) = (\phi(g))^{-1}$ *for all* $g \in G$.

3. $\phi(G) \equiv \{\phi(g) \mid g \in G\}$ *is a subgroup of* H.

The subgroup $\phi(G)$ in (3) above is called the *image* of the group G under the homomorphism ϕ.[1]

Proof: The first two assertions were proved in Proposition II.2.4. To prove the third assertion, We must show that the image $\phi(G)$ is closed under (i) the group operation in H, and (ii) taking inverses in H.

(i): Suppose $h_1, h_2 \in \phi(G)$. Then there are $g_1, g_2 \in G$ with $\phi(g_1) = h_1$ and $\phi(g_2) = h_2$. Since ϕ is a homomorphism, $h_1h_2 = \phi(g_1)\phi(g_2) = \phi(g_1g_2) \in \phi(G)$.

(ii): Suppose $h \in \phi(G)$. Then there is $g \in G$ with $h = \phi(g)$. Since G is a group, $g^{-1} \in G$, so $h^{-1} = \phi(g)^{-1} = \phi(g^{-1}) \in \phi(G)$.

■

The proof of the next proposition is Exercise 1.

[1] See Appendix D for more on the use of the term "image".

Proposition IV.1.2 *If* $\phi : G \longrightarrow H$ *and* $\psi : H \longrightarrow K$ *are homomorphisms, then* $\psi \circ \phi : G \longrightarrow K$ *is a homomorphism.*

Since homomorphisms are functions from one group to another, we can ask the same questions about homomorphisms that we do about functions: are they one-to-one? onto? bijections? A homomorphism which assumes the constant value equal to the identity is called *trivial.*

Exercises IV.1

1. Prove (Proposition IV.1.2) that the composition of two homomorphisms is a homomorphism.
2. If $\phi : G \longrightarrow H$ is a homomorphism and if $g \in G, n \in \mathbb{Z}$ then $\phi(g^n) = (\phi(g))^n$.
3. If $\phi : G \longrightarrow H$ is a homomorphism and if the order of $g \in G$ is finite, prove that the order of $\phi(g)$ divides the order of g.
4. Find all the homomorphisms from $\mathbb{Z}_5$ to S_3 and from D_5 to S_3.
5. Show that a homomorphism from a cyclic group into any group is determined by the image of a generator. [2] More generally, show that if $G = \langle g_1, ..., g_n \rangle$, then a homomorphism $\phi : G \to H$ is determined by $\phi(g_1)$, ..., $\phi(g_n)$. (State precisely what this means).
6. Apply the previous exercise to find all homomorphisms from $\mathbb{Z}$ to $\mathbb{Z}_n$ and to S_n.
7. Prove that the image of a group G under an injective homomorphism is isomorphic to G.
8. (a) Prove or disprove each of the next two statements:
   i. For every integer p, the function $\phi(z) = z^p$ is a homomorphism of $\mathbb{C}^*$ to itself .
   ii. For every integer p, and any group G, the function $\phi : G \to G$ given by $\phi(g) = g^p$ is a homomorphism of G to itself.
   iii. Generalize your results for (a),(b).

[2] In this context, this means the following: let ϕ_1, ϕ_2 be two homomorphisms from a cyclic group $G = \langle g \rangle$ to another group H. If $\phi_1(g) = \phi_2(g)$, then $\phi_1 = \phi_2$.

9. (a) Let $G = \mathrm{Sym(P)} < \mathrm{Isom}(\mathbb{C})$ where P is a square. Label the set X of vertices of the square $\{1, 2, 3, 4\}$. Show that this labelling determines a homomorphism of G to the symmetric group S_4. Is it injective? surjective? an isomorphism?

   (b) Do the same for a regular octagon P, where X is now the set of four *undirected* lines joining pairs of opposite vertices of the octagon.

IV.2 The Kernel of a Homomorphism

The elements of G that are sent to the identity element of H under a homomorphism $\phi : G \to H$ are special. We single them out by making

Definition IV.2.1 (Kernel) *If If $\phi : G \longrightarrow H$ is a homomorphism and e_H is the identity element in H, then the* **kernel** *N of ϕ is defined to be the subset of G given by*

$$N = \{g \in G \mid \phi(g) = e_H\}.$$

In fact,

Proposition IV.2.2 *The kernel of a homomorphism $\phi : G \longrightarrow H$ is a subgroup of the domain group G.*

Examples:

1. The kernel of any isomorphism, or, more generally, any injective homomorphism is just the trivial subgroup. (Proposition IV.2.3
2. The kernel of the homomorphism $\pi : \mathrm{Isom}(\mathbb{C}) \to O_2$ is the subgroup $\mathcal{T}$ of all translations. (Exercise 1)
3. The kernel of the determinant homomorphism $\det : O_2 \to \{-1, +1\}$ is the subgroup SO_2 of orthogonal matrices of determinant one. (Equivalently, it is the group of rotations fixing the origin.)
4. The kernel of the sign function $\sigma : S_n \to \{-1, +1\}$ is the alternating subgroup A_n. (Corollary II.7.14)
5. The kernel of the complex exponential function $\exp : (\mathbb{R}, +) \to (S^1, \cdot)$ given by $exp(t) = e^{2\pi it}$ is the group $\mathbb{Z}$ of integers.
6. The kernel of $\phi : (\mathbb{Z}, +) \to (\mathbb{Z}_n, +_n)$ is the subgroup $n\mathbb{Z} = \langle n \rangle$.
7. If V, W are vector spaces and $T : V \to W$ is linear, then the kernel of T (as a homomorphism from $(V, +)$ to $(W, +)$) coincides with the usual definition of kernel for a linear transformation.

What the kernel tells us about a homomorphism

The most basic fact about the kernel of a homomorphism is given by the following proposition.

Proposition IV.2.3 *A homomorphism* $\phi : G \longrightarrow H$ *is injective* $\iff$ kernel$(\phi) = \{e_G\}$ *(i.e., "*kernel*(ϕ) is trivial".)*

Proof: Suppose that ϕ is injective. Since ϕ is a homomorphism, $\phi(e_G) = e_H$. Since ϕ is injective, no other element can go onto e_H. Thus kernel$(\phi) = \{e_G\}$.

Suppose that $kernel(\phi) = \{e_G\}$. Then

$$\phi(a) = \phi(b) \implies e_H = \phi(a)(\phi(b))^{-1} = \phi(a)\phi(b^{-1}) = \phi(ab^{-1}).$$

Thus $ab^{-1} \in$ kernel$(\phi) = \{e_H\}$. So $a = b$ and we see that ϕ is one-one. ∎

More generally, the kernel tells us exactly when and how two elements $a, b \in G$ are identified under ϕ. Below, recall from Appendix D that the *preimage* $\phi^{-1}(h)$ of an element $h \in H$ is, by definition, all elements b of G for which $\phi(b) = h$.

Proposition IV.2.4 *Let* $\phi : G \to H$ *be a homomorphism with kernel* N *and let* $a \in G$. *Set* $h = \phi(a)$. *Then*

$$\phi^{-1}(h) = aN = Na.$$

Proof: Note that the two equal signs above are two separate statements requiring proof. As usual, if we want to show two sets X and Y are equal, we must show $X \subset Y$ and $Y \subset X$.

First we show aN and Na are subsets of $\phi^{-1}(h)$. Recall that by definition, $aN = \{an | n \in N\}$. Since

$$\phi(an) = \phi(a)\phi(n) = he_H = h$$

we have $aN \subset \phi^{-1}(h)$. Similarly $\phi(na) = h$ and hence $Na \subset \phi^{-1}(h)$.

Now we show $\phi^{-1}(h)$ is a subset of aN. If $b \in \phi^{-1}(h)$, then

$$\phi(a) = h = \phi(b) \implies \phi(a)^{-1}\phi(b) = e_H \implies \phi(a^{-1}b) = e_H \implies a^{-1}b \in N \implies b \in aN.$$

If we swap the order of the product $a^{-1}b$ to write instead $e_H = \phi(b)\phi(a)^{-1}$ and apply the same argument, we see that if $b \in \phi^{-1}(h)$, then $b \in Na$. Hence $\phi^{-1}(h)$ is a subset of both aN and Na.

Both inclusions are now established, so the Proposition is proved.

■

Example 8: Label the three inscribed equilateral triangles of a regular 9-gon P by with the symbols 1, 2, 3 as shown in Figure 1 (compare this with the constructions in Exercise IV.I.7).

Figure 1. Three triangles inscribed in a regular nonagon P

This defines a homomorphism ϕ from $G = \mathrm{Sym}(P) = D_9$ to $H = S_3$. If we write D_9 in the usual way with generators R, M as shown, then we get the following table showing where every element of D_9 goes under ϕ:

$\boldsymbol{N}$	$R\boldsymbol{N}$	$R^2\boldsymbol{N}$	$M\boldsymbol{N}$	$MR\boldsymbol{N}$	$MR^2\boldsymbol{N}$
id	R	R^2	M	MR	MR^2
R^3	R^4	R^5	MR^3	MR^4	MR^5
R^6	R^7	R^8	MR^6	MR^7	MR^8
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
id	(123)	(132)	(23)	(13)	(12)

The top line shows the left cosets of the kernel N in D_9. Each column on the top is the preimage in D_9 of the element of S_3 at the bottom. Thus, one might think of the homomorphism ϕ in the following way: it "collapses" each coset aN to the image $\phi(a) \in S_3$.

Notation. We denote the set of cosets of N in G by G/N, which we read as "G mod N.

Note that we do not need to specify left- or right- cosets when we refer to G/N above, since by Proposition IV.2.4 every left coset aN is also the right coset Na.

Note that G/N is a set whose elements are themselves sets — The *elements* of the *set* G/N are *cosets* of the form aN. Each coset aN is itself a subset of G. At this point, you should review from Proposition **??** the relationship between cosets and their labels, which are not unique. Also, you should review Appendix E, *Is that function well-defined?*.

Proposition IV.2.5 *Let $\phi : G \to H$ be a homomorphism with kernel N. Then the function*

$$\Phi : G/N \to \phi(G)$$

given by

$$\Phi(aN) = \phi(a)$$

is a well-defined bijection with $\Phi(N) = \Phi(eN) = \phi(e_G) = e_H$.

Proof: There are three things to prove.
Φ is well-defined. Suppose $aN = bN$. We must show $\phi(a) = \phi(b)$. Well,

$$aN = bN \implies a^{-1}b \in N \implies \phi(a^{-1}b) = e_H \implies \phi(a)^{-1}\phi(b) = e_H \implies \phi(a) = \phi(b).$$

The first implication follows from general properties of cosets **??**.

Φ is onto: Given any $h \in \phi(G)$, there exists $a \in G$ with $\phi(a) = h$, by the definition of the image $\phi(G)$. Then $\Phi(aN) = \phi(a) = h$.

Φ is one-to-one. Suppose $\Phi(aN) = \Phi(bN)$. By the definition of Φ, and the fact that ϕ is well-defined, this means $\phi(a) = \phi(b)$. But

$$\phi(a) = \phi(b) \implies \phi(a)^{-1}\phi(b) = e_H \implies \phi(a^{-1}b) = e_H \implies a^{-1}b \in N \implies aN = bN.$$

■

You may have noticed that the two chains of implications above are the reverses of one another. This is a general principle: *showing that a given onto function is one-to-one is the same as showing that its inverse is well-defined.*

Since $\Phi : G/N \to \phi(G)$ is a bijection, you might wonder whether somehow G/N can be turned into a group so that Φ is in fact a homomorphism. This is subtle but possible, and we turn to this in the next sections.

Exercises IV.2

1. Prove that: The kernel of the homomorphism $\pi : \text{Isom}(\mathbb{C}) \to O_2$ is the subgroup $\mathcal{T}$ of Isom $\mathbb{C}$ consisting of all translations.

2. Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be vertical projection onto the x-axis. Draw the kernel N and draw several cosets of N.

3. What is the kernel of $\det : GL_2(\mathbb{R}) \to \mathbb{R}^*$?

4. Suppose G has order 3 and H has order 2. Is there a nontrivial homomorphism from G to H? What if G has order 45 and H has order 14? Generalize.

5. Let G be the infinite cyclic subgroup of Isom$(\mathbb{C})$ generated by a glide-reflection. What is the kernel of $\pi : G \to O_2$?

6. Let m be an integer and let S^1 be the "circle subgroup" of $\mathbb{C} - \{0\}$ under ordinary multiplication, given by $S^1 = \{z \in \mathbb{C} \,|\, |z| = 1\}$. Let $\phi : S^1 \to S^1$ be the homomorphism $\phi(z) = z^m$. Find the kernel N of ϕ. Pick $m = 12$ and draw N and some of its cosets in S^1.

7. What is the kernel of the homomorphism $l_G : G \to S_G$ (see Cayley's Theorem (Thm. II.3.8) and Exercise 6 in Section IV.1.)

IV.3 Normal subgroups

By Proposition IV.2.4 we know that if N is the kernel of a homomorphism $\phi : G \to H$, then $aN = Na$ for all $a \in G$, i.e. every left coset is also a right coset. (See Exercise 1.) It is not true in general that given a subgroup of a group, each left coset is also a right coset (Exercise II.6.5). Therefore, kernels of homomorphisms are special kinds of subgroups. We single out this property with

Definition IV.3.1 (Normal subgroup) *A subgroup N of the group G is a* normal[3] subgroup of G *— and we write $N \triangleleft G$ — if for all $a \in G$, $aN = Na$.*

This definition can also be made in terms of conjugacy:

$$N \triangleleft G \iff ana^{-1} \in N \text{ for all } a \in G,\ n \in N.$$

That these two definitions are equivalent is part of the following proposition:

Proposition IV.3.2 (Characterizations of normal) *N is a normal subgroup of G if and only if any one (hence all) of the following conditions hold:*

1. *All left cosets are right cosets. That is, for all $a \in G$, $aN = Na$.*
2. *$aNa^{-1} = N$ for all $a \in G$. (By definition $aNa^{-1} = \{ana^{-1} \mid n \in N\}$.)*
3. *$aNa^{-1} \subset N$ for all $a \in G$.*
4. *$ana^{-1} \in N$ for all $a \in G,\ n \in N$.*

Proof: Clearly 2 $\implies$ 3 $\iff$ 4.

[3]There are only so many adjectives at one's disposal, and the term "special" sounds unsophisticated. So synonyms, like "normal", are used instead (in ≥ 70 distinct ways, according to one dictionary of mathematics).

Let's prove $3 \implies 2$. Suppose that $a \in G$. Then, by 3, $aNa^{-1} \subset N$ and we need only prove that $aNa^{-1} \supset N$. However $a^{-1} \in G$, so by 3 $a^{-1}N(a^{-1})^{-1} = a^{-1}Na \subset N$. But then

$$\begin{array}{rrcl} & a^{-1}Na & \subset & N \\ \implies & a\big(a^{-1}Na\big)a^{-1} & \subset & aNa^{-1} \\ \implies & N & \subset & aNa^{-1}. \end{array}$$

Now let's prove that $1 \implies 2$. Let $a \in G,\ n \in N$. By 1, there exists $n' \in N$ with $an = n'a$. Hence

$$ana^{-1} = n'aa^{-1} = n' \in N$$

proving that condition 4 holds. But we've already proven that $4 \iff 2.$. Hence $1 \implies 2$.

The implication $2 \implies 1$ is left as Exercise 2.. ∎

Note: We have used the following facts (left as Exercises) concerning elements a, b, c, d and subsets $X,\ Y$ of a group G:

a) If $X \subset Y$ then $cXd \subset cYd$.

b) $a(bXc)d = (ab)X(cd)$.

c) $eXe = X$.

We introduced the notion of conjugacy for geometric purposes in Section III.2. It is also key to the algebraic notion of a normal subgroup. Recall the definition:

Definition III.2.1. *Let G be a group. Two elements $g_1, g_2 \in G$ are* conjugate in G *if there exists an element $f \in G$ such that $g_2 = fg_1f^{-1}$.*

We proved (Lemma III.2.3) that conjugacy is an equivalence relation. We may characterize a normal subgroup $N \triangleleft G$ as a subgroup N of G which contains the entire conjugacy class of each element in N.

Proposition IV.3.3 (Normal iff union of conjugacy classes) *A subgroup N of a group G is normal if and only if N is a union of conjugacy classes in the group G.*

We leave the proof as Exercise 3 ■

Examples of normal subgroups:

1. The kernel N of any homomorphism $\phi : G \to H$ is normal, by Prop. IV.2.3.
2. The subgroups in the Examples of the previous two sections are normal, since they are kernels of homomorphisms.
3. G and $\{e_G\}$ are always normal (though improper) subgroups of G.
4. There are no proper normal subgroups N of $\mathrm{Isom}(\mathbb{C})$ which contain a reflection. To see this, note first that by Proposition IV.3.3, a normal subgroup consists of conjugacy classes. However, any two reflections are conjugate in $\mathrm{Isom}(\mathbb{C})$! So any such normal subgroup of $\mathrm{Isom}(\mathbb{C})$ must contain all reflections. Since N is also a sub*group* of $\mathrm{Isom}(\mathbb{C})$, N must contain all products of all reflections. However, by Proposition II.3.1, reflections generate $\mathrm{Isom}(\mathbb{C})$. So $N = \mathrm{Isom}(\mathbb{C})$.
5. A similar argument shows that there are no proper normal subgroups of S_n which contain a transposition (Exercise 4).
6. Any subgroup of an abelian group is normal.

We conclude this section with an often useful criterion for checking normality:

Proposition IV.3.4 *If H is a subgroup of index two of a group G, then H is normal in G.*

Proof: Let $a \in G$. If $a \in H$ then $aH = H = Ha$. Otherwise, since H has index two,

$$G = H \sqcup aH = H \sqcup Ha.$$

Subtracting off H we see that $aH = Ha$. By Proposition IV.3.2(a), H is normal. ■

It is important to realize that the property of N being a normal subgroup is a property of the *pair* G and N together. For example, let G and N be the dihedral groups $N = D_2 < D_4 = G <$ Isom $\mathbb{C}$. Then N is a subgroup of index two in G. By Proposition IV.3.4 N is normal in G. But N is not normal in Isom($\mathbb{C}$) since it is a proper subgroup containing a reflection (Example 4).

Exercises IV.3

1. Suppose that N is a subgroup of G and that the left coset aN is equal to some right coset. Prove that that right coset is equal to Na.

2. Give the omitted proofs in Proposition IV.3.2 (different characterizations of a normal subgroup).

3. If N is a subgroup of G, prove that $N \lhd G \iff N$ is a union of conjugacy classes in G.

4. Prove the claim in Example 5, that a proper normal subgroup of S_n never contains a transposition.

5. Prove that any subgroup of an abelian group is normal (Example 6).

6. Prove or disprove:

   (a) If K is a normal subgroup of H, and H is a normal subgroup of G, then K is a normal subgroup of G.

   (b) If $N \lhd K$ (a group) and G is a subgroup of K, then $G \cap N \lhd G$.

7. As usual, let $D_n = \langle R, M \rangle$ be the dihedral group, with $R = R_{\frac{2\pi}{n}}$, $M(z) = \overline{z}$. Recall that $R^j M = M R^{-j}$.

   (a) Find the conjugacy class of R^j in D_n $(0 \le j \le n-1)$.

   (b) Find the conjugacy class in D_n of the reflection $R^j M$.
   (Hint: The form of the answer depends on whether n is even or odd.)

   (c) Find all the normal subgroups of D_n.
   Hint: Exercise 3 might be useful.

8. Prove or disprove:

   (a) If $g_1, g_2 \in S_n$ are cycles of the same length l, then g_1 and g_2 are conjugate in the symmetric group S_n.

   (b) If $g_1, g_2 \in A_n$ are cycles of the same length l, then g_1 and g_2 are conjugate in the alternating group A_n.

9. **Definitions:** A group action of G on a set X is called *transitive* if, given any pair $x, y \in G$, there is $g \in G$ with $g(x) = y$. If G is any group and H_1, H_2 are subgroups of G, we say H_1, H_2 are *conjugate in* G if there exists $a \in G$ with $aH_1a^{-1} = H_2$.

   For example, S_n and A_n act transitively on $X = \{1, 2, ..., n\}$ if $n > 2$. $\mathrm{Isom}(\mathbb{C})$ and the subgroup $\mathcal{T}$ of translations act transitively on $X = \mathbb{C}$; the subgroup $\mathcal{R}$ of rotations about the origin does not act transitively on $\mathbb{C}$ (check these claims!).

   Prove that if G acts transitively on X, then the stabilizers

   $$H_1 = \mathrm{Stab}_G(x_1), \quad H_2 = \mathrm{Stab}_G(x_2)$$

   of any two points $x_1, x_2 \in X$ are conjugate subgroups of G.

IV.4 Quotient groups

There are two natural questions which we can ask at this point, which will have answers unified by the concept of *quotient group*.

We have seen that if N is the kernel of a homomorphism ϕ from a group G to another group H, then N is a normal subgroup of G. We ask the following question:

> *If $N \triangleleft G$ is there a homomorphism ϕ from G to another group with $N =$ kernel ϕ?*

We have also seen (Proposition IV.2.5) that when N is the kernel of a homomorphism ϕ from G to another group, then ϕ determines a bijection Φ from the set of cosets G/N to the image group $\phi(G)$.

> *Is there a way to define an operation on this set G/N which makes G/N into a group and which has the property that this function $aN \mapsto \phi(a)$ is actually an isomorphism of G/N with $\phi(G)$?*

One obvious way to try to multiply cosets is to *define*

$$(*) \qquad aN * bN = ab\,N.$$

However, this definition is given in terms of the labels a, b, for the cosets, and these labels are not unique. That is, *a priori it is not clear that this operation of multiplication is well-defined.*

There are actually two other reasonable ways to define the operation $(*)$. In Exercise 5 the reader will be asked to show that these other possibilities all amount to the same thing.

Theorem IV.4.1 *If N is a normal subgroup of the group G, and if multiplication in G/N is defined by the condition $(*)$ then*

1. *The binary operation $(*)$ is well-defined on G/N.*
2. *With the binary operation $(*)$, G/N is a group.*
3. *The function ("quotient map")*

$$q : G \longrightarrow G/N \quad \text{given by} \qquad q(a) = aN$$

is a homomorphism of G onto G/N with kernel N.

The group G/N is called the *quotient (or factor) of G by N*. There are many other places in mathematics in which quotients naturally arise.

Proof: 1. To prove $(*)$ is **well-defined**, we must show:

$$\textit{If } a_1N = a_2N \textit{ and } b_1N = b_2N, \textit{ then } (a_1b_1)N = (a_2b_2)N.$$

Assume the hypothesis. Using general properties of cosets (Proposition **??**),

$$a_1N = a_2N \implies a_1^{-1}a_2 \in N$$

$$b_1N = b_2N \implies b_1^{-1}b_2 \in N.$$

Again using properties of cosets, we can reformulate the conclusion:

$$(a_1b_1)N = (a_2b_2)N \iff (a_1b_1)^{-1}(a_2b_2) \in N.$$

But

$$\begin{array}{lcll} (a_1b_1)^{-1}(a_2b_2) \in N & \iff & b_1^{-1}a_1^{-1}a_2b_2 & \in N \quad (\text{since } (a_1b_1)^{-1} = b_1^{-1}a_1^{-1}) \\ & \iff & b_1^{-1}a_1^{-1}a_2b_1b_1^{-1}b_2 & \in N \qquad (\text{inserting } 1_G \text{ cleverly}) \\ vvv & \iff & b_1^{-1}\underbrace{a_1^{-1}a_2}_{\in N}b_1 & \in N \qquad (\text{since } b_1^{-1}b_2 \in N.) \end{array}$$

Since N is a normal subgroup, by Proposition IV.3.2 we have

$$a_1^{-1}a_2 \in N \implies b_1^{-1}(a_1^{-1}a_2)b_1 \in N.$$

Therefore the last condition in the sequence of "if and only if" statements above is true, and we are done.

2. Under the operation $(*)$, G/N becomes a group because the axioms are satisfied. Here is the verification:

a. **Closure.** By the *definition* of $(*)$, the product of the two elements $aN,\ bN \in G/N$ is the element $abN \in G/N$.

b. **Associativity.** $(aN * bN) * cN = abN * cN = \big((ab)c\big)N = \big(a(bc)\big)N = aN * (bN * cN)$.

c. **Identity.** $N = eN$ serves as identity element: $eN * aN = eaN = aN$.

d. **Inverses.** $a^{-1}N$ is the inverse of aN: $aN * a^{-1}N = a^{-1}N * aN = eN = N$.

3. The function q is a homomorphism since $q(ab) = abN = aN * bN = q(a)q(b)$. For any coset $aN \in G/N$ (the target set of q) we have $q(a) = aN$. Therefore q is onto.

To find the kernel of q, recall from (2) above that the identity element in G/N is just N. Thus

$$a \in \ker(q) \iff q(a) = e_{G/N} \iff aN = N \iff a \in N.$$

■

Theorem IV.4.1 says that the answers to the two questions posed at the beginning of this section are both *yes*. That is, given *any* normal subgroup N of a group G, we can construct a new group G/N. Suppose now that we are in the special case when $N = \ker(\phi)$ for a homomorphism $\phi : G \to H$. We have seen (Proposition IV.2.5) that

ϕ determines a bijection $\Phi : G/N \to \phi(G)$. However, we now know how to turn G/N into a group. We get

Theorem IV.4.2 (The First Isomorphism Theorem)

If $\phi : G \longrightarrow H$ is any homomorphism which is onto and which has kernel N, then the function $\Phi : G/N \longrightarrow H$ given by $\Phi(aN) = \phi(a)$ is an isomorphism.

If we drop the hypothesis that ϕ is onto, then the conclusion should be modified so that Φ is an isomorphism from G/N to the image group $\phi(G)$.

Proof: Φ is a well-defined bijection by Proposition IV.2.5. Further, Φ is a homomorphism because

$$\begin{array}{rcll} \Phi(aN * bN) & = & \Phi(abN) & \text{by the definition of } * \\ & = & \phi(ab) & \text{by the definition of } \Phi \\ & = & \phi(a)\phi(b) & \text{since } \phi \text{ is a homomorphism} \\ & = & \Phi(aN)\Phi(bN) & \text{by the definition of } \Phi. \end{array}$$

Therefore Φ is an isomorphism.

■

What we understand from the First Isomorphism Theorem:

- Two different homomorphisms $\phi_1 : G \longrightarrow H_1$ and $\phi_2 : G \longrightarrow H_2$ with the same kernel N have isomorphic images

$$\phi_1(G) \cong G/N \cong \phi_2(G).$$

Thus, if we count isomorphic groups as the same, the number of different homomorphic images of G is no more than the number of normal subgroups $N \triangleleft G$.

For example, if $G = D_3 = \{id, R, R^2, M, MR, MR^2\}$ where R is rotation about the origin by $\pi/3$ radians and M is reflection in the x-axis, then (exercise!) the only normal subgroups of G are $\{id\}$, $\{id, R, R^2\}$, and D_3 itself. Thus the possible images are isomorphic to

$$D_3/\{\text{id}\} \cong D_3, \qquad D_3/\{id, R, R^2\} \cong \mathbb{Z}_2, \qquad D_3/D_3 \cong \{\text{id}\}.$$

- Every normal subgroup $N \triangleleft G$ does give us a homomorphic image — namely q is a homomorphism with image G/N and kernel N.
- Not only is G/N a copy of $\phi(G)$ when ϕ is a homomorphism with kernel N, but the homomorphism $q : G \longrightarrow G/N$ is a copy of the homomorphism $\phi : G \longrightarrow \phi(G)$. This is because the element $q(a)$ of G/N which a goes to under q is matched by the isomorphism Φ with the element $\phi(a)$ that a goes to under ϕ. In other words

$$\Phi\big(q(a)\big) = \phi(a).$$

This is sometimes summarized in a diagram like that shown in Figure 2:

Figure 2. The three "tick marks" indicate that the diagram commutes: no matter which way you go from G to $\phi(G)$, the result is the same. More precisely, $\Phi \circ q = \phi$.

Examples.

1. S_n/A_n, $\mathrm{O}_2/\mathrm{SO}_2$, and $\mathrm{Isom}(\mathbb{C})/\mathrm{Isom}^+(\mathbb{C})$ are each isomorphic to $\mathbb{Z}_2$.
2. Let $G = \mathbb{R}$ and let $\phi : G \to \phi(G) = S^1$ be given by $\phi(\theta) = e^{i\theta}$. Then the kernel of ϕ is the group $N = \langle\, 2\pi \,\rangle$. The First Isomorphism Theorem then implies $\mathbb{R}/\langle\, 2\pi \,\rangle \cong S^1$.
3. Let $G = \mathrm{Isom}(\mathbb{C})$ and let $\phi = \pi : \mathrm{Isom}(\mathbb{C}) \to \mathrm{O}_2$ be the point map. The kernel of π is the subgroup $N = \mathcal{T}$ of translations. The point map π is onto, so we conclude from the First Isomorphism Theorem that $G/N = \mathrm{Isom}(\mathbb{C})/\mathcal{T} \cong \mathrm{O}_2$.

4. Let $G = D_9$ and let $\phi : G \to S_3$ be the homomorphism given in Example 8 of IV.2, whose kernel is $N = \{\mathrm{id}_{\mathbb{C}}, R^3, R^6\}$. Then $G/N \cong S_3$, and in fact the isomorphism Φ is gotten by examining the effect of an element of G/N on the set of inscribed triangles.

Exercises IV.4

1. Carry out the proof of Theorem IV.4.1 in detail for

   (a) $G = \mathbb{R}^2$ and N the y-axis.

   (b) $G = \mathrm{Sym}(P)$ and N as in Example 8 of Section IV.2.

2. Find, up to isomorphism, all possible quotient groups of D_6 and D_9.

3. Find an explicit isomorphism in each of the following situations between the quotient group G/N and another well-known group.

   (a) $G = (\mathbb{Z}, +)$, $N = n\mathbb{Z} \stackrel{def}{=} \{\, na \mid a \in \mathbb{Z} \,\}$
   (b) $G = (\mathbb{R}^2, +)$, $N =$ the y-axis
   (c) $G = S_n$, $N = A_n$
   (d) $G = (\mathbb{R}, +)$, $N = \mathbb{Z}$
   (e) $G = (\mathbb{R}^2, +)$, $N = \langle (1,0), (0,1) \rangle$
   (f) $G = (\mathbb{C}, +)$, $N = \langle 2\pi i \rangle$

4. Prove or disprove:
   Suppose $N \lhd G$ and $N' \lhd G'$. If $G \cong G'$ and $N \cong N'$, then $G/N \cong G'/N'$.

5. Let G be a group with normal subgroup N. Here, we give alternatives to defining a binary operation on the set of cosets G/N. Recall that an element of G/N is a coset of N in G, and that, in particular, it is a subset of G.

   Let $X, Y \in G/N$ be two cosets.

   (a) Choose *any* $x \in X$ and *any* $y \in Y$, and set $X \cdot Y = xyN \in G/N$. Prove that the binary operation $(X, Y) \to X \cdot Y$ on G/N is well-defined and that it is the same as the operation $(*)$.

(b) Show that the set

$$XY = \{xy \in G \mid x \in X,\ y \in Y\}$$

is itself a coset, i.e. is in G/N. (You'll need to use the fact that N is a normal subgroup.) Hence, the binary operation on G/N given by $(X, Y) \to XY$ makes sense, i.e. that the "product" of two cosets is again a coset. Note that this operation on G/N is automatically well-defined once you know this! Prove that this operation is the same as the operation $(*)$.

6. Let V, W be vector spaces and $T : V \to W$ an onto linear transformation with kernel N. Prove that V/N (as an additive group with addition inherited from vector addition in V) is isomorphic to W. Draw a picture for the case when $V = \mathbb{R}^2, W = \mathbb{R}$, and $T : V \to W$ is given by $(x, y) \to x + y$.

7. Prove that the quotient group $\mathbb{R}/\mathbb{Q}$ has no elements of finite order other than the identity.

IV.5 The Group Extension Problem

The First Isomorphism Theorem opens up fascinating questions about how groups can be constructed. We introduce the *group extension problem*, some tools for analyzing it, and examples of extensions of $\mathbb{Z}$ and $\mathbb{Z} \oplus \mathbb{Z}$ which will be useful in our analysis of the frieze and wallpaper groups in the next chapter.

Definition IV.5.1 *If A, B and G are groups then G is an extension of A by B*[4] *if there is an isomorphism i of A to a normal subgroup $i(A) \triangleleft G$ such that $G/i(A) \cong B$.*

Given groups A and B, the **group extension problem** is the problem of finding all groups (up to isomorphism) such that G is an extension of A by B.

A useful notation for dealing with group extensions is the following:

Definition IV.5.2 *A sequence of groups and homomorphisms of the form*

$$(*) \qquad 1 \longrightarrow A \xrightarrow{i} G \xrightarrow{\phi} B \longrightarrow 1 \ ,$$

is called **exact** *if the kernel of each homomorphism equals the image of the preceding homomorphism. (Here "1" denotes the trivial group; so there are unique homomorphisms $1 \longrightarrow A$ and $A \longrightarrow 1$.) If a sequence of the form $(*)$ is exact it is called* **a short exact sequence**.

Notice that a sequence of the form $(*)$ satisfies:

- $\text{kernel}(i) = \text{image}(1 \longrightarrow A) = \{e_G\} \iff i$ is an injection.
- $\text{image}(\phi) = \text{kernel}(B \longrightarrow 1) = B \iff \phi$ is *onto*.
- $i(A) = \text{kernel}(\phi) \implies i(A) \triangleleft G$.

[4] Unfortunately some authors call this an extension of B by A. Caveat emptor!

- Thus if $(*)$ is exact then the First Isomorphism Theorem tells us that G is an extension of A by B and $B \cong G/i(A)$.

We summarize this with the following theorem:

Theorem IV.5.3 (Extensions $\longleftrightarrow$ short exact sequences) *There is a short exact sequence*

$$1 \longrightarrow A \xrightarrow{i} G \xrightarrow{\phi} B \longrightarrow 1 \ ,$$

if and only if G is an extension of A by B.

■

We will use two tools in analyzing group extensions.

Proposition IV.5.4 (First tool - the action on A by conjugation) . *Assume that we have a short exact sequence $(*)$ in which A is a normal subgroup of G and i is the inclusion map*[5] *Then*

1. *Each element $g \in G$ gives an automorphism α_g of A by conjugation: $\alpha_g(a) = gag^{-1}$.*
2. *If A is abelian then $[\,\phi(g) = \phi(h)\,] \Longrightarrow [\,\alpha_g = \alpha_h\,]$. Thus each element $b \in B$ determines an element $\alpha_b \in$ Aut A by choosing g with $\phi(g) = b$ and setting $\alpha_b = \alpha_g$.*
3. *If A is abelian the function $b \mapsto \alpha_b$ is a homomorphism of B into* Aut A.

Proof: The first assertion follows from Proposition III.2.19. To see the second assertion notice that $[\,\phi(g) = \phi(h)\,]$ implies that g and h belong to the same coset of A. Thus there exists $a_0 \in A$ such that $h = ga_0$. For every $a \in A$ the commutativity of A then yields

$$\alpha_h(a) = hah^{-1} = (ga_0)a(a_0^{-1}g^{-1}) = gag^{-1} = \alpha_g(a),$$

[5]This can be achieved changing "A" to "i(A)" letting the new "i" be the inclusion map.

so that $\alpha_h = \alpha_g$. Given $b \in B$, we may therefore choose any g with $b = \phi(g)$ (remember that ϕ is onto !) and we will get a well-defined automorphism if we define $\alpha_b = \alpha_g$.

Finally, to see that the function $b \mapsto \alpha_b$ is a homomorphism, suppose given $b_1, b_2 \in B$. Choose $g_i\,(i = 1, 2)$ with $\phi(g_i) = b_i$. Notice that $\phi(g_1 g_2) = b_1 b_2$. Then for all $a \in A$ we have

$$\alpha_{b_1 b_2}(a) = \alpha_{g_1 g_2}(a) = (g_1 g_2)a(g_1 g_2)^{-1} = g_1(g_2 a g_2^{-1})g_1^{-1} = (\alpha_{g_1} \circ \alpha_{g_2})(a) = (\alpha_{b_1} \circ \alpha_{b_2})(a).$$

Therefore $b_1 b_2 \mapsto \alpha_{b_1} \circ \alpha_{b_2}$, proving that the map is a homomorphism. ∎

Proposition IV.5.5 (second tool - check for a splitting)
Suppose that a short exact sequence

$$(*) \qquad 1 \longrightarrow A \xrightarrow{i} G \xrightarrow{\phi} B \longrightarrow 1$$

is given, where A is a normal subgroup of G and i is the inclusion map. Then the following statements are equivalent:

1. *There is a homomorphism $j : B \longrightarrow G$ such that $\phi \circ j = id_B$.*
   (j is called a **splitting** *of the short exact sequence or a* **section** *of the map ϕ. We say that thet* **short exact sequence splits** *or that the* **group extension splits**.*)*
2. *There exists a subgroup B_0 of G such that $Ag \cap B_0$ consists of a single element for each coset Ag of A in G.*
3. *There exists a subgroup B_0 of G such that $\phi|B_0 : B_0 \longrightarrow B$ is an isomorphism.*
4. *There exists a subgroup B_0 of G such that every element $g \in G$ is uniquely expressible in the form $g = ab$ where $a \in A,\ b \in B_0$.*

If B_0 is as above then there is a homomomorphism $\alpha : B_0 \longrightarrow Aut\ A$ such that, (denoting $\alpha(b) = \alpha_b$) multiplication in G is given by

$$(a_1 b_1)(a_2 b_2) = (a_1 \alpha_{b_1}(a_2))\,(b_1 b_2) \quad \textit{for all } a_1, a_2 \in A,\ b_1, b_2 \in B_0.$$

Proof: Since the sequence is exact, ϕ is onto and $B = \phi(G)$. We use the fundamental bijection (Proposition IV.2.5) $G/A \longrightarrow \phi(G)$ given by

$$b \longleftrightarrow \phi^{-1}(b) \quad \text{for all } b \in B; \qquad \text{or equivalently,} \qquad Ag \longleftrightarrow \phi(g) \quad \text{for all } g \in G.$$

1. $\Longrightarrow$ 2: Let $B_0 = j(B)$. If $j(b_1)$ and $j(b_2)$ belong to the same coset Ag then

$$\phi(g) = \phi j(b_i) = b_i \quad (i = 1, 2) \implies b_1 = b_2.$$

Thus $Ag \cap B_0$ contains at most one element. But it does contain one element, namely $j(\phi(g))$, since $\phi \circ j = \mathrm{id}_B \implies (\phi j)(\phi(g)) = \phi(g) \implies j\phi(g) \in \phi^{-1}(\phi(g)) = Ag$.

2. $\Longrightarrow$ 3: $\phi \,|\, B_0 = \phi \circ$ inclusion is a homomorphism, and is a bijection since inclusion:$B_0 \longrightarrow G/A$ is a bijection by 2. and $\phi : G/A \longrightarrow B$ is a bijection by Proposition IV.2.5.

3. $\Longrightarrow$ 4: If $g \in G$ then the fact that $\phi \,|\, B_0$ is an isomorphism implies that there exists a unique $b \in B_0$ such that $\phi(b) = \phi(g)$. But then $gb^{-1} = a \in A$ is also uniquely determined. Thus $g = ab$, for unique $a \in A$, $b \in B_0$.

4. $\Longrightarrow$ 3: $B = \{\phi(g) \,|\, g \in G\} = \{\phi(ab) \,|\, a \in A, b \in B_0\} = \{\phi(b) \,|\, b \in B_0\} = \phi(B_0)$. Thus $\phi \,|\, B_0$ is onto. To see that $\phi \,|\, B_0$ is one-one, suppose that $\phi(b') = e_B$, $b' \in B_0$. Then $b' \in A =$ kernel ϕ. If $b' \neq e_G = e_A = e_{B_0}$ then $b' = e_A\, b' = b'\, e_B$ has two different expressions of the form "ab", contrary to 3. Thus $b' = e_G$ and $\phi \,|\, B_0$ is one-one.

3. $\Longrightarrow$ 1: Let $j = (\text{inclusion} : B_0 \longrightarrow G) \circ (\phi \,|\, B_0)^{-1}$. Then $j : B \longrightarrow G$ is a homomorphism with $\phi \circ j = \mathrm{id}_B$.

Therefore statements 1. — 4. are equivalent. Let $\alpha : B_0 \longrightarrow \mathrm{Aut}\, A$ by $\alpha_b(a) = bab^{-1}$. Then, expressing elements of G as $g = ab$ as in 4. above, we have

$$(a_1b_1)(a_2b_2) = a_1(b_1a_2b_1^{-1})b_1a_1 = (a_1\alpha_{b_1}(a_2))\,(b_1b_2) \quad \text{for all } a_1, a_2 \in A,\ b_1, b_2 \in B_0.$$

This completes the proof of Proposition IV.5.5. ∎

Remark. The structure of G described in the preceding proposition, when G results from a split group extension, is formalized as a *semi-direct product*:

Definition IV.5.6 *Suppose that* A *and* B *are groups and that* $\alpha : B \longrightarrow Aut\ A$ *is a homomorphism. Then the* **semi-direct product** $G = A \times_\alpha B$ *is defined as the group with underlying set the cartesian product* $A \times B$ *and rule of multiplication*

$$(a_1,\, b_1)(a_2,\, b_2) = (a_1 \alpha_{b_1}(a_2),\ b_1 b_2).$$

See Exercise 2 for details.

Example IV.5.7 *Suppose that* G *is an extension of* A *by* B*, with short exact sequence* $1 \longrightarrow A \xrightarrow{i} G \xrightarrow{\phi} B \longrightarrow 1$*. Suppose that* A *is torsion-free (*i.*e., has no elements of finite order) and* B *is finite then* G *is a split extension if and only if* G *contains a subgroup* B_0 *isomorphic to* B*.*

Proof: If G is a split extension, the result follows by definition. Conversely, if there is such a subgroup B_0, then $B_0 \cap A = \{e_G\}$, since the non-trivial elements of the finite group B_0 all have finite order, and so cannot belong to A. But then no non-trivial element of B_0 belongs to kernel(ϕ). Hence
$\phi \,|\, B_0 : B_0 \longrightarrow B$ is one-one. Since B and B_0 are finite sets with the same number of elements, we see that $\phi\,|B_0$ is also onto — thus an isomorphism. Therefore G is a split extension by Proposition IV.5.5. ∎

Corollary IV.5.8 *An extension* G *of* $\mathbb{Z}$ *or* $\mathbb{Z} \oplus \mathbb{Z}$ *by a finite group* B *is a split extension if and only if* G *contains a subgroup isomorphic to* B*.*

Exercises IV.5

1. Prove that the dihedral group D_n is an extension of $\mathbb{Z}_n$ by $\mathbb{Z}_2$.
2. (a) The semi-direct product $A \times_\alpha B$ is a group with
$$e_{A \times_\alpha B} = (e_A\, e_B); \qquad (a,\, b)^{-1} = (\alpha_{b_1^{-1}}(a_1^{-1}),\, b_1^{-1}.$$
   (b) In this semi-direct product, $A \times \{1\}$ is a normal subgroup isomorphic to A and $\{1\} \times B$ is a subgroup isomorphic to B. (We denote these subgroups simply as A and B.)
   (c) If $1 \longrightarrow A \xrightarrow{i} G \xrightarrow{\phi} B \longrightarrow 1$ is a split short exact sequence then there is a homomorphism $\alpha : B \longrightarrow$ Aut A such that $G \cong A \times_\alpha B$.

IV.6 Application to the Point Group of G

If G is a subgroup of Isom($\mathbb{C}$), let

$$\mathcal{T}_G = \mathcal{T} \cap G.$$

The function $\pi : Isom\ \mathbb{C} \longrightarrow \mathrm{O}_2$ restricts to a function (which we also call π, though the usual convention would be to call it $\pi \,|\, G$) with the smaller domain G. The kernel of this restricted function is the subgroup $\mathcal{T}_G$. Thus we have from the First Isomorphism Theorem

$$\pi(G) \cong G/\mathcal{T}_G.$$

Definition IV.6.1 (Point group of G) *If $G <$ Isom($\mathbb{C}$), then the group $\pi(G)$ is called the* **point group** *of G.*

It is important to realize that *both* the group G and the point group $\pi(G)$ are subgroups of the *same* larger ambient group Isom($\mathbb{C}$). In particular the point group $\pi(G)$ consists of isometries.

Since G acts on the plane, the stabilizer of the origin

$$H = \mathrm{stab}_G(0) = \{g \in G | g(0) = 0\}$$

is a subgroup of G. If $h \in H$ is an isometry, then it has a formula given by $h(\vec{z}) = A\vec{z} + \vec{b}$ for some $A \in \mathrm{O}_2$ and some $\vec{b} \in \mathbb{R}^2$. But $h \in H \implies h(0) = 0$, so $h(\vec{z}) = A\vec{z}$ and so $h = A = \pi(g) \in \pi(G)$:

$$H = \mathrm{stab}_G(0) < \pi(G).$$

Notice that in general, if you have a homomorphism $\phi : G \to \phi(G)$ and a subgroup H of G, it does not make sense ask whether $H < \phi(G)$, since they live *a priori* in different places. However, in our setting, each of H, G, and $\pi(G)$ are all subgroups of the same larger group, Isom($\mathbb{C}$), so it does make sense to compare them.

In fact, the point group $\pi(G)$ has a very special relationship with $\mathcal{T}_G$.

Proposition IV.6.2 *Let $G < \mathrm{Isom}(\mathbb{C})$ have point group $\pi(G)$ and translation subgroup $\mathcal{T}_G$. Then for all $h \in \pi(G)$ and all $T \in \mathcal{T}_G$,*

$$hTh^{-1} \in \mathcal{T}_G.$$

Note that the conclusion $hTh^{-1} \in \mathcal{T}$ is obvious since $h \in \mathrm{Isom}(\mathbb{C})$ and $\mathcal{T} = \ker(\pi)$ is a normal subgroup of $\mathrm{Isom}(\mathbb{C})$. So what the proposition really says is the stronger conclusion that also hTh^{-1} is an element of G itself.

Proof: Let $h \in \pi(G)$. Then $h = \pi(g)$ for some $g \in G$. From the definition of π this means that $h = T_c g$, where $c = -g(0)$.

Let $T \in \mathcal{T}_G$. Then

$$\begin{array}{rcll} hTh^{-1} & = & (T_c\, g)\, T\, (T_c\, g)^{-1} & \\ & = & T_c\, (g\, T\, g^{-1})\, T_c^{-1} & \\ & = & T_c\, T'\, T_c^{-1} \quad \text{for some } T' \in \mathcal{T}_G & \text{since } \mathcal{T}_G \triangleleft\ G \\ & = & T' & \text{since translations commute.} \end{array}$$

■

Below, recall that the point group consists of isometries of the plane, hence acts on the plane by isometries.

Theorem IV.6.3 (Point group acts on $\mathcal{T}_G.0$) *Let $G < \mathrm{Isom}(\mathbb{C})$ have point group $\pi(G)$ and translation subgroup $\mathcal{T}_G$. Then $\pi(G)$ acts on the orbit of the origin under $\mathcal{T}_G$.*

Remember that $\pi(G)$ is not, in general, a subgroup of G, so it is maybe a bit surprising that $\pi(G)$ sends the orbit of zero under the action of the subgroup $\mathcal{T}_G$ of G to itself.

Proof: Suppose $\vec{b}$ is in the orbit of 0 under the action of $\mathcal{T}_G$. We must show that if $h \in \pi(G)$, then $h(\vec{b})$ is also in the orbit of 0 under $\mathcal{T}_G$, i.e. that there is an element

$\mathcal{T}' \in \mathcal{T}_G$ for which $h(\vec{b}) = T'(0)$. Let $\vec{b} = T(0)$, where $T \in \mathcal{T}_G$. Then

$$\begin{aligned} h(\vec{b}) &= h(\,T(0\,)) \\ &= h\,T(\,h^{-1}(0)) \qquad \text{since } h^{-1}(0) = 0 \\ &= h\,T\,h^{-1}(0) \\ &= T'(0), \quad \text{where } T' \in \mathcal{T}_G \quad \text{by Proposition IV.6.2} \end{aligned}$$

■

Often, knowing that a group acts on a certain kind of set yields a great deal of information about the group. In the next chapter, we will exploit Theorem IV.6.3 when G is a so-called *discrete* group to deduce specific information about the structure of G. Exercise 5 is an example of the kind of argument we will use.

Exercises IV.6

1. What is wrong with the following argument, purportedly proving Proposition IV.6.2? "$\mathcal{T}_G$ is a normal subgroup, so $h\mathcal{T}_G h^{-1} = \mathcal{T}_G$, and so for all $T \in \mathcal{T}_G$ we have $hTh^{-1} \in \mathcal{T}_G$."

2. Compute the point group $\pi(G)$ if

   (a) $G = \langle z \mapsto \overline{z} + 1, \quad z \mapsto z + i \rangle$

   (b) $G = \langle z \mapsto \overline{z}, \quad z \mapsto -\overline{z} + i \rangle$

   (c) $G = \langle z \mapsto e^{2\pi i/3} z, \quad z \mapsto z + 1 \rangle$

3. Find an example of a group $G < \mathrm{Isom}(\mathbb{C})$ for which the stabilizer of the origin is trivial, but the point group is nontrivial.

4. Let G_1, G_2 be two subgroups of $\mathrm{Isom}(\mathbb{C})$ which are conjugate in $\mathrm{Isom}(\mathbb{C})$. Prove
   (i) that their point groups are conjugate in O_2;
   (ii) their point groups need not be identical.

5. Suppose $G < \mathrm{Isom}(\mathbb{C})$ and $\mathcal{T}_G$ is generated by $z \mapsto z + 1$ and $z \mapsto z + i$.

   Prove that the point group $\pi(G)$ is a subgroup of the group generated by R and M, where R is rotation by $\pi/2$ radians about the origin and M is reflection in the real axis.
   Hint: draw the orbit $\mathcal{T}_G.0$ and apply Theorem IV.6.3.

Chapter V

Wallpaper groups

V.1 Equivalence of symmetry groups

We return in this last chapter to the subject with which we started our investigations – symmetries of plane sets and patterns. Recall the definition (Section I.7):

Definition V.1.1 Let P be a subset of the plane. A **symmetry** of P is an isometry $f : \mathbb{C} \to \mathbb{C}$ such that $f(\mathrm{P}) = \mathrm{P}$. **The group of all symmetries of P** is denoted by Sym(P).

A. Frieze groups, wallpaper groups and their patterns

By a *pattern* we shall merely mean a set, although we shall think of patterns as (possibly very complicated) sets whose structure we wish to study. For example, an infinite grid in the plane made of squares of side one is a pattern we might study. It is the set consisting of infinitely many horizontal and vertical lines; say the lines $x = n$ and $y = m$, for all m, $n \in \mathbb{Z}$.

Sometimes a pattern P will have no recognizable repetition of motif – no symmetry

Figure V.1: A pattern with trivial symmetry group – no regular repetition of motif.

Figure V.2: Frieze and wallpaper patterns

– in it, and we would perhaps not refer to it in ordinary English as a *pattern*. The corresponding mathematical statement would be that $\mathrm{Sym(P)} = \{\mathrm{id}\}$. See Figure V.1.

In fact, we are thinking of decorative patterns, as have been found in art and architecture and textiles throughout human history, or of patterns in nature — plane crystal patterns, for example, and beehive patterns. In particular we are concerned with **frieze patterns** (i.e., strip patterns which translate in one dimension) and **wallpaper patterns** (which translate in two directions). See Figure V.2.

In order to be precise about which patterns we are thinking of – to model our intuition about the patterns we have in mind – we name the isometry groups which we will consider. These are the **frieze groups** and **wallpaper groups**. The patterns in question are those patterns which have these isometry groups as symmetry groups.

Definition V.1.2 *A subgroup G of Isom $\mathbb{C}$ is a* **frieze group** *if*

- *The translation subgroup $\mathcal{T}_G$ of G is infinite cylic:*

$$\mathcal{T}_G = \langle T_b \rangle \text{ for some nonzero } b \in \mathbb{C}.$$

- *The point group $\pi(G)$ is finite.*

Definition V.1.3 *A subgroup G of Isom $\mathbb{C}$ is a* **wallpaper group** *if*

- *The translation subgroup $\mathcal{T}_G$ of G is generated by translations by two linearly vectors:*

  $\mathcal{T}_G = \langle T_b, T_c \rangle$ *where b and c are linearly independent vectors in the plane.*

- *The point group $\pi(G)$ is finite.*

Remarks on the definition of frieze and wallpaper groups

1. The requirement in these definitions that the point group of G be finite reflects the fact that we don't want to consider infinite repeated symmetry around any one point. For example, G cannot contain a rotation R of infinite order because then $\pi(G)$ would contain the rotation $\pi(R)$ of infinite order, contradicting the hypothesis that $\pi(G)$ is finite.
2. The total effect of our definitions could be achieved more elegantly but at greater length by a discussion of **discrete groups**. A discrete group G is one for which the orbit Gz of every point $z \in \mathbb{C}$ fails to have any limit points in the plane. It turns out that an isometry group is discrete if and only if it is finite or a frieze group or a wallpaper group. See Appendix **??**.

Examples: As usual, we let R_{θ,z_0} be rotation by θ radians about the point z_0, R_θ be rotation about the origin by θ radians and M_θ be reflection across the the line through the origin which makes angle θ radians with the x-axis. We set $M = M_0 =$ reflection across the x-axis. In Figure V.2, we then have:

The frieze pattern P_1 has symmetry group G_1 with

$$G_1 = \langle\, T_2, R_{\pi,\frac{1}{2}}, MT_1, M_{\pi/2}\,\rangle = \langle\, MT_1, M_{\pi/2}\,\rangle; \qquad \pi(G_1) = D_2 = \{\mathrm{id}, R_\pi, M, M_{\pi/2}\}.$$

The wallpaper pattern P_2 has symmetry group $G_2 = \mathrm{Sym}(P_2)$ with

$$G_2 = \langle\, T_1, T_i, M_{\pi/4}\,\rangle; \qquad \pi(G_2) = \{\mathrm{id}, M_{\pi/4}\}.$$

B. Equivalent isometry groups

Clearly there are infinitely many patterns. But our interest is in the symmetry groups $\mathrm{Sym}(P)$ for various patterns P, and different patterns may contain the same types of symmetries and have "essentially" the same relationships between these symmetries. We have to decide what we mean for two symmetry groups to be the same ("equivalent") or not the same ("inequivalent"). Here is the definition we will use.

Definition V.1.4 *Two isometry groups G_1, G_2 (i.e., subgroups of Isom $\mathbb{C}$) are* **equivalent**[1] *if there exists an isomorphism $\phi : G_1 \longrightarrow G_2$ which takes translations to translations, rotations to rotations, reflections to reflections and glide reflections to glide reflections. We write $G_1 \equiv G_2$ to denote the fact that G_1 and G_2 are equivalent and we call ϕ* **an equivalence**.

Exercise (= Exercise 2): Prove that: (a) If ϕ is an equivalence then ϕ^{-1} is an equivalence. (b) As the name indicates, the definition gives an equivalence relation on the set of subgroups of Isom $\mathbb{C}$.

Examples

1. If P_1 and P_2 are congruent then, although they may be situated differently in the plane (e.g., perhaps one is a rotated version of the other) their symmetries

[1] Perhaps a better name would be "**isometrically isomorphic**" . since this requires an isomorphism which preserves the isometry data and the word "equivalent" is used in so many contexts. But "isometrically isomorphic" is a mouthful. So we won't go that way.

should match up so that Sym P_1 is equivalent to Sym P_2. In fact we have already proved that this is the case, since we have proved (Theorem **??**) that if $f : \mathbb{C} \longrightarrow C$ is an isometry with $f(P_1) = P_2$, then f conjugates Sym P_1 to Sym P_2. Conjugation always takes a group to an isomorphic group, and conjugation by an isometry takes translations to translations, rotations to rotations, etc. Conjugation is our main tool for proving groups equivalent, and we shall strengthen this result in the next section.

2. Let P_n $(n = 1, 2, \ldots)$ be the pattern consisting of the real axis union n line segments of length 1 sticking out at each integer point and making angles $\frac{\pi}{4}$, $\frac{\pi}{5}$, $\ldots \frac{\pi}{n+3}$ with the x-axis. Then $P_1, P_2, \ldots$ are infinitely many patterns all having the same symmetry group $\mathrm{Sym}(P_n) = \langle\, T_1 \,\rangle$.

3. Let P_1 be as in the previous example and let $Q_n = R_n(P_1)$, the result of rotating P_1 by n radians. then the groups $\mathrm{Sym}(Q_n)$ are infinitely many distinct subgroups of Isom $\mathbb{C}$, distinct but nevertheless all equivalent to $\langle\, T_1 \,\rangle$, since the Q_n are all congruent to each other.

C. The classification problem – our grand finale

Our final goal now is to *classify* the frieze groups and the wallpaper groups and thus to delineate exactly what kind of symmetry can occur in the frieze and wallpaper patterns. We wish to answer the following questions.

- Determine when two frieze or wallpaper groups are equivalent to each other.
- List the equivalence classes in each case.

Notice that a frieze group F cannot be equivalent to a wallpaper group G because there is no isomorphism taking the translation group (isomorphic to $\mathbb{Z}$) of F to the translation group (isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$) of G. We shall show in Section V.3 that equivalent groups must have isomorphic point groups as well as isomorphic translation groups; so frieze groups are only equivalent to frieze groups and wallpaper groups to wallpaper groups.

In the remaining sections of this chapter:

The Grand Finale

> *We shall prove that there are* **seven** *equivalence classes of frieze groups and* **seventeen** *equivalence classes of wallpaper groups. We exhibit one group in each class — a "***normal form***" — which we identify by listing a set of geometrically meaningful generators for the group. We show that any other element in the class is not only equivalent to it, but conjugate to it in the affine group (and in most cases, even in the group of similarities.*

See Theorem V.4.1 and Theorem V.5.1 for precise statements of these results.

In Section V.2 the basic facts particular to the frieze and wallpaper groups are given (as opposed to a discussion of equivalence for arbitrary isometry groups, as in Section V.3). These are the facts that a) any non-zero set of translation vectors has a translation vector of smallest positive length and b) the stunning consequence (the **crystallographic restriction**) that the only non-zero rotations which can occur in a wallpaper group are by the angles π, $\frac{\pi}{2}$, $\frac{\pi}{3}$, $\frac{\pi}{4}$ and $\frac{\pi}{6}$. In a frieze group the only non-zero rotations are by π radians.

In Section V.3 we develop the tools for classifying equivalence classes of isometry groups. We show how conjugacy in an appropropriate group can be used to prove that two groups are equivalent (thus conjugacy is a *sufficient* condition for equivalence) and we give necessary conditions for concluding that two groups are not equivalent. (For example we shall show that if two groups are equivalent then their point groups are isomorphic — thus, isomorphism of their point groups is a *necessary* condition for equivalence of two isometry groups.)

Finally, In Sections V.4 and V.5 we prove the classification theorems.

Exercises V.1

1. Prove that the pattern in Figure V.1 has trivial symmetry group: sym(P) = id.

2. We have defined two subgroups G_1, G_2 of Isom $\mathbb{C}$ to be "equivalent" if there exists an isomorphism $\phi : G_1 \longrightarrow G_2$ which takes each isometry to an isometry of the same type (translation, rotation, reflection, glide reflection). Such an isomorphism ϕ is called an *equivalence*. Prove that:

   (a) If ϕ is an equivalence then ϕ^{-1} is an equivalence.

   (b) This is an equivalence relation on the set of subgroups of Isom $\mathbb{C}$.

   (c) Prove that $\mathbb{Z}$ is not isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$ (thus justifying the claim that a frieze group cannot be equivalent to a wallpaper group.

3. Prove that the groups $G_1 = \langle T_1 \rangle$ and $G_2 = \langle T_2 \rangle$ are equivalent to each other but are not conjugate in Isom $\mathbb{C}$.

4. Give an example to show that *Isomorphic isometry groups need not be equivalent.* (*However we shall demonstrate in Section* **??** *the striking fact that isomorphic wallpaper groups are equivalent* (Theorem **??**).

V.2 Smallest non-trivial translations and the crystallographic restriction

In this section we prove that if G is a frieze or wallpaper group then G has two very special properties. We shall prove the following two theorems.

Theorem V.2.1 [Shortest non-trivial translation vectors]
If G is a frieze group or a wallpaper group with translation subgroup $\mathcal{T}_G$ and if $S \subset \mathcal{T}_G$ is a set of non-zero translations then there is a translation $T_b \in S$ such that

$$T_c \in S \implies |b| \leq |c|.$$

Theorem V.2.2 [The crystallographic restriction]
If G is a subgroup of $\mathrm{Isom}(\mathbb{C})$ which which has finite point group and contains a non-zero translation of minimal positive translation length, then any rotation in G has order $1,\ 2,\ 3,\ 4$ or 6.

We shall use the following terminology.

Definition V.2.3 *If $T_b \in G$ then $b = T_b(0)$ is a* **translation vector** *for G and $|b|$ is the* **translation length** *of T_b.*

Note that: The set of all translation vectors for G is the orbit of 0 under $\mathcal{T}_G$, denoted $\mathcal{T}_G(0)$.

Proof of Theorem V.2.1

If G is a frieze group, suppose that

$$\mathrm{id} \notin S \ \subset \ \mathcal{T}_G = \langle\, T_v \,\rangle = \{T_{nv} \,|\, n \in \mathbb{Z}\}.$$

Then $|nv| = |n||v|$ and S has a shortest translation — namely T_{av} where $|a|$ is the minimum of the set of positive integers $\{\, |n| \,|\, T_{nv} \in S\}$. (Geometrically, the set of

translation vectors for G is evenly spaced along a straight line through the origin and av is the closest one to the origin such that $T_{av} \in S$. If $T_{-av} \in S$ then $-av$ would also be a shortest translation vector for S.)

Suppose now that G is a wallpaper group and that S is a set of non-zero translations of G. By hypothesis, there is a pair of linearly independent vectors $v = (a, b), w = (c, d)$ such that

$$\begin{aligned} \mathcal{T}_G(0) &= \langle\, v, w \,\rangle = \{mv + nw \,|\, m, n \in \mathbb{Z}\} = \{(ma + nc,\ mb + nd) \,|\, m, n \in \mathbb{Z}\} \\ &= \left\{ \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} m \\ n \end{pmatrix} \,\middle|\, m, n \in \mathbb{Z} \right\} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \,\middle|\, \begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{T}_G(0). \right\} \end{aligned}$$

We wish to show that there is a *shortest* non-zero vector $\begin{pmatrix} x \\ y \end{pmatrix}$ in S. Since v and w are linearly independent[2] the 2×2 matrix above is invertible. Let us denote

$$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}, \qquad A^{-1} = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}.$$

Every non-zero element of $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{T}_G(0)$ satisfies

$$\begin{pmatrix} m \\ n \end{pmatrix} = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \qquad \text{for some } m, n \in \mathbb{Z}, \quad \text{with } m \neq 0 \ \text{ or } n \neq 0.$$

Claim: There is a real number $\epsilon > 0$ such that every non-zero translation vector $b = \begin{pmatrix} x \\ y \end{pmatrix}$ has $|b| = \sqrt{x^2 + y^2} > \epsilon$.

To see this note that, since A^{-1} is invertible, neither of its rows consists of only zeroes. Therefore $|\alpha| + |\gamma| > 0, \quad |\beta| + |\delta| > 0$. On the other hand, notice that

$$\begin{aligned} |m| &= |\alpha x + \gamma y| \leq |\alpha x| + |\gamma y| \leq \big(|\alpha| + |\gamma|\big)\big(|x| + |y|\big) \\ |n| &= |\beta x + \delta y| \leq |\beta x| + |\delta y| \leq \big(|\beta| + |\delta|\big)\big(|x| + |y|\big) \end{aligned}$$

[2]This is quite important: Suppose that $v = (1, 0),\ w = (\sqrt{2}, 0)$. These are not linearly independent, but there are no non-zero **integers** with $mv + nw = 0$. Nevertheless, the translation group $\langle\, T_v,\, T_w \,\rangle$ has no shortest non-zero translation vector (Exercise **??**).

If $m \neq 0$ then

$$|x| + |y| \geq \frac{1}{|\alpha| + |\gamma|}.$$

Otherwise $n > 0$ and

$$|x| + |y| \geq \frac{1}{|\beta| + |\delta|}.$$

In either case there is a positive number — call it 2ϵ — such that $|x| + |y| \geq 2\epsilon$. Therefore, either $|x| \geq \epsilon$ or $|y| \geq \epsilon$. We conclude that $\sqrt{x^2 + y^2} \geq \epsilon$, proving our claim.

To complete the proof that the set S has a minimal translation vector, consider a square grid ("graph paper") on the plane consisting of squares of diagonal $\delta < \epsilon$. No two distinct vectors $b, c \in T_G(0)$ can belong to the same square in the grid, since $b - c$ is a non-zero translation vector and so has length at least ϵ. Now choose any translation vector c for S. It lies in a square X of this grid and there are only finitely many squares whose most distant point from the origin is less than or equal to the most distant point of X. Look in each of these squares to see if there is a translation vector for S in the square. This gives a finite set of translation vectors for S and a shortest one among these is a shortest translation vector for S. ∎

We shall give two applications of the existence of shortest non-zero translation vectors. The following Proposition tells us how to use this to find geometrically meaningful generators for the translation subgroup $\mathcal{T}_G$. Then we shall use it to prove that the crystallographic restriction holds.

Proposition V.2.4 [Shortest translations generate]

1. *If G is a frieze group and b is a shortest non-zero translation vector for $\mathcal{T}_G$ then $\mathcal{T}_G = \langle T_b \rangle$.*

2. *If G is a wallpaper group, suppose that b is a shortest non-zero translation vector and let c be a shortest non-zero translation vector among all those which are linearly independent of b (i.e., which are not real multiples of b). Then $\mathcal{T}_G = \langle T_b, T_c \rangle$. (***Note:*** *Under the hypothesis here we say that b, c are a* **shortest pair of translation vectors**.*)*

Proof: When G is a frieze group, $\mathcal{T}_G = \langle T_c \rangle = \{T_{nc} \,|\, n \in \mathbb{Z}\}$, for some non-zero $C \in \mathbb{C}$. Then the shortest translation vectors are $b = \pm c$, and $\mathcal{T}_G = \langle T_b \rangle$.

If b, c are a shortest pair for a wallpaper group G, we must show that any element $T_a \in \mathcal{T}_G$ is of the form

$$T_a = T_b^m \, T_c^n = T_{mb+nc.}$$

In other words we must show that any other translation vector a is of the form $a = mb + nc$ for some integers m, n. Suppose, on the contrary, that there is a translation vector a which is not of this form.

Let P be the parallelogram with vertices 0, b, c, $b + c$. Tesselate the plane by the parallelograms $T_{mb+nc}(P)$. Since a lies on one of these parallelograms and is assumed not to be a vertex (a point of the form $mb + nc$) then some translation $T_w = T_{-(mb+nc)}T_a$ gives a translation vector w which lies in P and is not one of the vertices of P.

Let $M = \max\{|b|, |c|\}$. If v is the closest vertex of P to w then $w - v$ is a non-zero translation vector. So, since b, c is a shortest pair, $|w - v| \geq M$. However (Exercise 2), every non-vertex point of P — in particular the point w — has distance less than M from its nearest vertex. Thus the assumption that $\mathcal{T}_G \neq \langle T_b, T_c \rangle$ has led to a contradiction. ∎

The crystallographic restriction

We have just shown that frieze and wallpaper groups satisfy the hypothesis of Theorem V.2.2. In proving that this theorem (*the crystallographic restriction*) indeed holds, we first give a useful lemma on the construction of translations.

Recall that the rotation $R_{\theta,\, z_0}$ by θ radians about the point z_0 has the formula

$$R_{\theta,\, z_0}(z) \;=\; e^{i\theta}\, z + (1 - e^{i}\theta)\, z_0.$$

Lemma V.2.5

1. *If* $z_0 \neq z_1$ *then* $R_{\theta,\,z_1} \circ R_{-\theta,\,z_0}$ *is a translation. Indeed*

$$R_{\theta,\,z_1} \circ R_{-\theta,\,z_0}(z) = z + (1 - e^{i\theta})(z_1 - z_0).$$

2. *For any isometry* g*, the isometry* $gR_{\theta,\,z_0}g^{-1}$ *is a rotation. Indeed*

$$gR_{\theta,\,z_0}g^{-1} = R_{\theta,\,g(z_0)}.$$

3. *If* T_b *is the translation by* $b \in \mathbb{C}$ *then*

$$T_b\, R_{\theta,\,z_0}\, T_{-b}\, R_{-\theta,\,z_0}\,(z) = z + (1 - e^{i\theta})\, b.$$

4. *If* T_b *is the translation by* $b \in \mathbb{C}$ *then*

$$T_b\, R_{\theta,\,z_0}\, T_b\, R_{-\theta,\,z_0}\,(z) = z + (1 + e^{i\theta})\, b.$$

5. *If* $\alpha,\, \beta \in \mathbb{R}$ *then*

$$R_{\alpha,\,z_1} R_{\beta,\,z_2} R^{-1}_{\alpha,\,z_1} R^{-1}_{\beta,\,z_2}\,(z) \;=\; z + (1 - e^{i\alpha})(1 - e^{i\beta})\,(z_1 - z_2).$$

Proof: 1. $R_{\theta,\,z_1} \circ R_{-\theta,\,z_0}$ is a translation because it is a direct isometry which does not change the angle which any line makes with the horizontal. The exact formula is found by calculating (or visualizing) where z_0 goes.

2. $gR_{\theta,\,z_0}g^{-1}$ is a direct isometry with fixed point set $\{g(z_0)\}$ since, in general, $\mathrm{Fix}\,(g\,f\,g^{-1}) = g\,(\mathrm{Fix}(f))$. The only such direct isometry is $R_{\theta,\,g(z_0)}$.

3. If in 2. we take $g = T_b$, we get $T_b R_{\theta,\,z_0} T_{-b} = R_{\theta,\,z_0+b}$. Applying 1., we see that $T_b\, R_{\theta,\,z_0}\, T_{-b}\, R_{-\theta,\,z_0}\,(z) = z + (1 - e^{i\theta})\, b$.

4. Direct calculation shows that

$$T_b R_{\theta,\,z_0} T_b = R_{\theta,\,z_1} \quad \text{where} \quad z_1 = z_0 + b\frac{1 + e^{i}\theta}{1 - e^{i}\theta}$$

The result then follows from 1.

5. Using 2. we see that

$$R_{\alpha,\,z_1} R_{\beta,\,z_2} R_{-\alpha,\,z_1} = R_{\beta,\,z_3} \quad \text{where} \quad z_3 = R_{\alpha,\,z_1}(z_2) = e^{i\alpha}\, z_2 + (1 - e^{i\alpha})z_1.$$

We then apply 1. to see that $R_{\beta,\,z_3}\,R_{-\beta,\,z_2} = z + (1-e^{i\alpha})(1-e^{i\beta})\,(z_1 - z_2)$. ∎

Proof of the crystallographic restriction, Theorem V.2.2

Let $T_b \in G$ be a translation with translation vector b of minimal positive length. Since the point group of G is finite any rotation $R \in G$ about some point z_0 has finite order – say n. Then $\langle R \rangle$ contains $R_{\theta,\,z_0}$ where $\theta = 2\pi/n$. But then by the preceding lemma and the closure axiom for groups, the translation given by

$$T_b\,R_{\theta,\,z_0}\,T_{-b}\,R_{-\theta,\,z_0}\,(z) = z + (1-e^{i\theta})\,b$$

is an element of G. Since there is no translation vector for G shorter than b, we see that $|1-e^{i\theta}| \geq 1$. This means that $e^{i\theta}$ is a point on the circle of radius one about the origin and is on or outside the circle of radius one about the number 1. Clearly then (see Figure V.3) $\theta = \frac{2\pi}{n} \geq \frac{\pi}{3}$. Thus $n \in \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$.

But if rotation g by $2\pi/5$ radians belonged to G, so would rotation g^2 by $\alpha = 4\pi/5$ radians. However then the translation, with the formula given by the previous lemma,

$$T_b R_{\alpha,\,z_0} T_b R_{-\alpha,\,z_0}(z) = z + (1+e^{i}\alpha)b,$$

would belong to G. However, the minimality of the length of b would require that

$$|1+e^{i}\alpha| \;=\; |e^{i\alpha} - (-1)| \;\geq\; 1.$$

This is false if $2\pi/3 \leq \alpha \leq \pi$. (Draw a picture analogous to Figure V.3.) Thus it is false for $\alpha = 4\pi/5$. This proves that G contains no rotation of order 5. ∎

Note: It is easy to construct patterns P, using hexagons or using squares, such that Sym(P) is a wallpaper group containing containing rotations of orders $2, 3, 4$ or 6. (Exercise 3.)

From the Crystallographic Restriction we get the important Corollary:

Corollary V.2.6 *Suppose that B is a frieze group or a wallpaper group. Then the point group $\pi(G)$, and also the stabilizer of any point of $z \in \mathbb{C}$, $Stab_G(z)$, is one of the following:*

- $\{id\}$,

Figure V.3: The region $|1 - e^{i\theta}| \geq 1$ (shaded)

- *a rotation group of order* $2,\ 3,\ 4,$ *or* $6,$
- *a conjugate of one of the dihedral groups* $D_1,\ D_2,\ D_3,\ D_4,$ *or* $D_6.$

Proof: Every rotation R of order n in Isom $\mathbb{C}$ goes to a rotation $\pi(R)$ of the same order in the point group. Thus, by the Crystallographic Restriction the rotations in the point group $\pi(G)$ can only have order $1, 2, 3, 4,$ or 6. But, by hypothesis, $\pi(G)$ is a finite group of isometries fixing the origin — hence (Theorem III.4.2) is trivial or a finite rotation group or conjugate to a finite dihedral group. Putting these pieces together, we get the possible groups stated in the theorem.

If $z \in \mathbb{C}$, $\mathrm{Stab}_G(z)$ contains no translations other than the identity (since these couldn't fix z), so the kernel of the homomorphism $\pi \,|\, \mathrm{Stab}_G(z)$ is trivial. Therefore π is one-one when restricted to this stabilizer and maps it isomorphically onto a subgroup of $\pi(G)$. Since subgroups of rotation groups and conjugates of dihedral groups are rotation groups or conjugates of dihedral groups, the result follows. ■

Exercises V.2

1. Suppose that $v = (1, 0),\ w = (\sqrt{2}, 0)$. Prove that
   - These vectors are not linearly independent.
   - There are no non-zero **integers** m, n with $mv + nw = 0.$

- The translation group $\langle T_v, T_w \rangle$ has no shortest non-zero translation vector.

2. Suppose that b, c are linearly independent vectors and that P is the parallelogram whose vertices are the points $0,\ b,\ c,\ b+c$. If w is a point of P which is not one of the vertices, prove that the distance from w to its nearest vertex is less than $M = \max\{|b|,\ |c|\}$.

   Hint: Divide P into four subparallelograms, each with two sides of length $\frac{|b|}{2}$ and two of length $\frac{|c|}{2}$. Use the facts that w lies in one of these parallegrams and that, since a and b are independent,

$$\left| \frac{b}{2} \pm \frac{c}{2} \right| < \frac{|b|}{2} + \frac{|c|}{2}.$$

3. Give examples of wallpaper groups containing rotations of orders $2,\ 3,\ 4$ or 6.
   Hint: Use a checkerboard or hexagonal pattern to define these groups.

V.3 The Keys to Classifying Frieze and Wallpaper Groups

Our plan is to show how to conjugate any given frieze group or wallpaper group to an equivalent group in "normal form", and to choose these normal forms so that no two are equivalent to each other.[3] Then we can determine whether two groups G_1, G_2 are equivalent by conjugating them to their normal forms N_1, N_2 and checking whether these are equal. Each normal form will be specified, using the following template, as a subgroup of Isom $\mathbb{C}$ generated by a certain set of elements. (The translation subgroup of the group N is denoted $\mathcal{T}_N$.)

$N = \langle T_1, g_1, \ldots g_q \rangle$ when G is a **frieze group**,
where $\mathcal{T}_N = \langle T_1 \rangle$ and $\langle \pi(g_1) \ldots \pi(g_q) \rangle = \pi(N)$.

$N = \langle T_b, T_c, g_1, \ldots g_q \rangle$ when G is a **wallpaper group**,
where $\mathcal{T}_N = \langle T_b, T_c \rangle$, with b, c a shortest pair of generating translation vectors,
and $\langle \pi(g_1) \ldots \pi(g_q) \rangle = \pi(N)$.

For example, consider the frieze groups $G_1 = \mathrm{Sym}(P_1)$, $G_2 = \mathrm{Sym}(P_2)$, where P_1, P_2 are as in Figure V.3. It will turn out that $N_1 = N_2 = \langle T_1, R_\pi \rangle$. Thus G_1 and G_2 are equivalent.

This section lays the foundation for the two aspects of this program:

1) In the subsection *Moving towards a normal form*, we show how to gather information about the group. We demonstrate how to conjugate a given frieze or wallpaper group G to an equivalent group with standardized point group $\mathcal{R}_n$ or D_n; we discuss how to find a generating set for the group, how to determine whether the short exact sequence $1 \longrightarrow \mathcal{T}_G \xrightarrow{i} G \xrightarrow{\pi} \pi(G) \longrightarrow 1$ splits, and how $\pi(G)$ determines what the translation vectors for $\mathcal{T}_G$ are. Much

[3] There are many situations in mathematics where one chooses one special element from each equivalence class of a given equivalence relation and calls these the "normal forms" under that equivalence relation.

Figure V.4: Symmetry groups of these patterns have normal form $N = \langle T_1, R_\pi \rangle$

of the subtlety and beauty comes from the interaction of the opposite isometries (those which π takes to reflections) with the translations.

2) In the subsection, *Tools for proving groups inequivalent*, we give necessary conditions for proving that groups are equivalent. We can prove two groups inequivalent by proving that some necessary condition for equivalence is not met. From the definition we know that equivalent groups must be isomorphic and have isomorphic translation subgroups. We show further that their point groups must be isomorphic (and it will follow, conjugate in Isom O_2!); that if one has a splitting of its short exact sequence then so does the other, and that the translation subgroups are conjugate in s way influenced by the point groups. The final determination of the normal forms will be given in Sections V.4 and V.5.

Moving towards a normal form

Proposition V.3.1 (Conjugating to achieve a standard point group.) *If G is a frieze group or a wallpaper group whose point group (necessarily finite) contains n elements then G is conjugate in Isom $\mathbb{C}$ to a group G_2 satisfying:*

1. *If G contains only rotations and/or translations:*

$$\pi(G_2) = \mathcal{R}_n \quad \textit{and} \quad \mathcal{R}_n < G.$$

2. *If G contains a reflection or glide reflection:*

$$\pi(G_2) = D_m \ (n = 2m) \quad \textit{and} \quad \mathcal{R}_m < G.$$

Proof of Proposition V.3.1:

Note that the two cases given in the Proposition may be restated as the dichotomy that $\pi(G)$ does or does not contain a reflection. By hypothesis $\pi(G)$ is a finite subgroup of O_2. We apply Corollary III.4.3 (with $\pi(G)$ playing the role of the finite group G in that corollary). This yields

$$\begin{array}{lrcl}
\pi(G) \text{ contains no reflections} & & \Longrightarrow & \exists\ g \in \mathrm{O}_2 \text{ with } g\pi(G)g^{-1} = \mathcal{R}_n. \\
\pi(G) \text{ contains a reflection} & & \Longrightarrow & \exists\ g \in \mathrm{O}_2 \text{ with } g\pi(G)g^{-1} = D_m. \\
\textbf{Note that:} & g \in \mathrm{O}_2 & \Longrightarrow & \pi(g) = g. \\
& g \in \mathrm{Isom}\ \mathbb{C} & \Longrightarrow & gGg^{-1} \sim G \quad (\sim \text{ means "conjugate in Isom } \mathbb{C}").
\end{array}$$

With g as in the preceding step, let $G_1 = gGg^{-1}$. We use the fact that π is a homomorphism to see that

$$\begin{aligned}
\pi(G_1) &= \pi(gGg^{-1}) \\
&= \pi\left(\{gxg^{-1} \,|\, x \in G\}\right) \\
&= \{\pi(gxg^{-1} \,|\, x \in G\} \\
&= \{\pi(g)\pi(x)\pi(g)^{-1} \,|\, x \in G\} \\
&= \{g\pi(x)g^{-1} \,|\, x \in G\} \\
&= g\pi(G)g^{-1} \\
&= \begin{cases} \mathcal{R}_n & \text{if } G \text{ contains only rotations and translations.} \\ D_m & \text{if } G \text{ contains a reflection.} \end{cases}
\end{aligned}$$

Let $R \in \pi(G_1)$, $R = R_{\frac{2\pi}{n}}$ or $R_{\frac{2\pi}{m}}$, according as $\pi(G_1) = \mathcal{R}_n$ or $\pi(G_1) = D_m$. Denote $k = \mathrm{order}(R)$ $(= n \text{ or } m)$. Then $R = \pi(h)$ for some isometry $h \in G_1$. By the definition of π (think of where π takes each kind of isometry), we see that h must be a rotation of order k about some point z_0. Then $T_{-z_0} h T_{z_0} = R_{\frac{2\pi}{k}}$. Let $G_2 = T_{-z_0} G_1 T_{z_0}$. Then $G_2 \sim G_1 \sim G$ and

$$\pi(G_2) = \pi(T_{-z_0} G_1 T_{z_0}) = \pi(G_1) = \begin{cases} \mathcal{R}_n & \text{if } G \text{ contains no reflections,} \\ D_m & \text{if } G \text{ contains a reflection.} \end{cases}$$

and $$\langle\, R_{\frac{2\pi}{k}} \,\rangle = \mathcal{R}_k < G_2,$$

completing the proof of the Proposition. ∎

We wish to be able to name generators for the groups we're studying. We use the following lemma.

Lemma V.3.2 *Suppose that* $1 \longrightarrow \mathcal{T} \xrightarrow{i} G \xrightarrow{\pi} H \longrightarrow 1$ *is a short exact sequence of groups, where* i *is the inclusion map. Suppose that we are given generating sets* $S_{\mathcal{T}}$ *and* S_H *for* $\mathcal{T}$ *and* H*, and suppose that* $X \subset G$ *is a subset with* $\pi(X) = S_H$*. Then* $G = \langle S_{\mathcal{T}} \cup X \rangle$.

Proof: Since $G \supset S_{\mathcal{T}} \cup X$, we know that $G \supset \langle S_{\mathcal{T}} \cup X \rangle$. To prove the opposite inclusion we must show that each $g \in G$ is a product of elements of $S_{\mathcal{T}} \cup X$, and their inverses.

To see this, suppose that $\pi(g) = h_1^{n_1} h_2^{n_2} \dots h_k^{n_k}$ where $h_i \in S_H$. By hypothesis, there exists $x_i \in X$ with $\pi(x_i) = h_i$. Then

$$\pi\left(x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}\right) = h_1^{n_1} h_2^{n_2} \dots h_k^{n_k} = \pi(g).$$

Since g and $x = x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}$ have the same image under π, they belong to the same coset of the kernel of π, namely $\mathcal{T}$. Therefore there exists $t \in \mathcal{T}$ with $g = tx$. But $t = t_1^{m_1} \dots t_\ell^{m_\ell}$ for some $t_1, \dots, t_\ell \in S_{\mathcal{T}}$. Hence

$$g = t_1^{m_1} \dots t_\ell^{m_\ell} x_1^{n_1} x_2^{n_2} \dots x_k^{n_k} \quad \text{as claimed.} \qquad \blacksquare$$

Proposition V.3.3 (Generators for frieze and wallpaper groups)

1. *Suppose that* G *is a frieze group with* $\mathcal{T}_G = \langle T_a \rangle$ *and* $\pi(G) = \langle h_1 \dots h_n \rangle$*. Suppose that* $x_i \in G$ *satisfies* $\pi(x_i) = h_i$, $1 \le i \le n$*. Then*

$$G = \langle T_a, x_1, \dots, x_n \rangle.$$

2. *Suppose that* G *is a wallpaper group with* $\mathcal{T}_G = \langle T_a, T_b \rangle$ *and* $\pi(G) = \langle h_1 \dots h_n \rangle$*. Suppose that* $x_i \in G$ *satisfies* $\pi(x_i) = h_i$, $1 \le i \le n$*. Then*

$$G = \langle T_a, T_b, x_1, \dots, x_n \rangle.$$

Proof: This follows immediately from V.3.2 using the short exact sequence $1 \longrightarrow \mathcal{T}_G \xrightarrow{i} G \xrightarrow{\pi} \pi(G) \longrightarrow 1$. ∎

We may wish to change the scale of the figure we are looking at (for example if we want the length of a given translation to be exactly one unit). Recall that, for $0 \neq \lambda \in \mathbb{R}$, the function $m_\lambda : \mathbb{C} \longrightarrow \mathbb{C}$ is defined by $m_\lambda(z) = \lambda z$ and is called a *dilation*. The function m_λ dilates everything by a factor of λ. (If $\lambda < 0$ it will also rotate the entire picture by 180 degrees; note that $R_\pi = m_{-1}$.)) This won't change the equivalence class of the symmetry group G. Indeed the new figure will have symmetry group $m_\lambda G m_\lambda^{-1}$. This is equivalent to G, as we see by summarizing our discussion in Section III.2 with the following lemma. (See III.2.8, III.2.9, III.2.17.)

Lemma V.3.4 *Suppose that G is a subgroup of Isom $\mathbb{C}$.*

- *If $0 \neq \lambda \in \mathbb{R}$, then $m_\lambda T_b m_\lambda^{-1} = T_{\lambda b}$ for all $b \in \mathbb{C}$.*
- *If $0 \neq \lambda \in \mathbb{R}$, then $m_\lambda G m_\lambda^{-1}$ is equivalent to G.*
- *If $b, c \in \mathbb{C}$ are non-zero then there is a rotation R and a dilation m_λ such that*

$$(Rm_\lambda)T_b(Rm_\lambda)^{-1} = T_c.$$ ∎

We now illustrate the tools given thus far by determining the classification of frieze and wallpaper groups in the simplest cases, when there are only rotations and translations.

Theorem V.3.5 (Classification of groups of direct isometries)

1. *If G is a frieze group consisting completely of direct isometries then G is equivalent to one and only one of the following two subgroups of Isom $\mathbb{C}$ – the one which has the same point group as G. Indeed G is conjugate to this group in Sim $\mathbb{C}$ or Aff_2, as indicated.*

	normal form	**point group** $\pi(G)$	**conjugate in**
•	$\langle T_1 \rangle$	$\{id\}$	Aff_2
•	$\langle T_1, R_\pi \rangle$	$\{id, R_\pi\} = \langle R_\pi \rangle$	Aff_2

2. *If G is a a wallpaper group consisting completely of direct isometries then G is equivalent to one and only one of the following subgroups of Isom $\mathbb{C}$ – the one which has the same point group as G. Indeed G is conjugate to this group in Sim $\mathbb{C}$ or Aff$_2$, as indicated*

normal form	**point group** $\pi(G)$	**conjugate in**
• $\langle T_1, T_i \rangle$	$\{id\}$	*Aff*$_2$
• $\langle T_1, T_i, R_\pi \rangle$	$\langle R_\pi \rangle$	*Aff*$_2$
• $\langle T_1, T_i, R_{\pi/2} \rangle$	$\langle R_{\pi/2} \rangle$	*Sim* $\mathbb{C}$
• $\langle T_1, T_{\frac{1}{2}+\frac{\sqrt{3}}{2}i}, R_{\pi/3} \rangle$	$\langle R_{\pi/3} \rangle$	*Sim* $\mathbb{C}$
• $\langle T_1, T_{\frac{\sqrt{2}}{2}+\frac{\sqrt{2}}{2}i}, R_{\pi/4} \rangle$	$\langle R_{\pi/4} \rangle$	*Sim* $\mathbb{C}$
• $\langle T_1, T_{\frac{\sqrt{3}}{2}+\frac{1}{2}i}, R_{\pi/6} \rangle$	$\langle R_{\pi/6} \rangle$	*Sim* $\mathbb{C}$

Proof: We note first that none of the eight groups listed above are equivalent to each other: A frieze group cannot be equivalent to a wallpaper group because their translation subgroups are not isomorphic (so there exists no isomorphism taking translations to translations). Between the two frieze groups listed or within the six wallpaper groups listed, no two are isomorphic because the rotations of highest orders in two distinct groups, G_1, G_2 have the same orders as the rotations of highest orders in $\pi(G_1), \pi(G_2)$. But these highest orders are different if $G_1 \neq G_2$ are both frieze or both wallpaper groups in the above list. Since an isomorphism showing these groups equivalent would take a rotation of given order to a rotation of the same order, we see that G_1 is not equivalent to G_2.

Frieze groups: If G is a frieze group, then G contains only rotations by 0 radians (the identity map) and π radians. For if R were a rotation by θ radians, $\theta \neq 0, \pi$ let T_a be a non-trivial translation. Then

$$RT_aR^{-1} \;=\; T_{(\pi(R))(a)}$$

would be another translation in G with translation vector $(\pi(R))(a)$, linearly independent of A. This would contradict the fact that G is a frieze group.

If G contains no rotations (*i.e.,* only the identity rotation) then G consists entirely of translations. Thus $G = \langle T_a \rangle$ for some non-zero $a \in \mathbb{C}$. Then $\pi(G) = \{\text{id}\}$ and by Part 3. of Lemma V.3.4, we may conjugate G in Aff$_2$2 to the group $\langle T_1 \rangle$.

If G is a frieze group which contains a rotation by π radians and no reflections or glide reflections, then G is conjugate in Isom $\mathbb{C}$ (Proposition V.3.1) to a frieze group G_2 with $\pi(G_2) = \{\mathrm{id}, R_\pi\} \subset G_2$. Then $G_2 = \langle T_a, R_\pi \rangle$ by the proposition on generators V.3.3. If we conjugate T_a by an appropriate similarity $g = Rm_\lambda$, as in lemma V.3.4, then we get

$$\begin{aligned} gG_2g^{-1} &= g\langle T_a, R_\pi \rangle g^{-1} = \langle T_1, R_\pi \rangle \\ \text{since} \quad (Rm_\lambda)R_\pi(Rm_\lambda)^{-1} &= R_\pi. \end{aligned}$$

Wallpaper groups:

TO BE CONTINUED

Exercises V.3

1. Prove that in the third case of Definition **??** there is no splitting $j : D_m \longrightarrow G$.
2. Show that $m_\lambda^{-1} = m_{\lambda^{-1}}$
3. Prove Lemma V.3.4
4. Prove Lemma **??**.

V.4 The classification of frieze groups

This section will be devoted to proving the following theorem:

Theorem V.4.1 (The classification of frieze groups) *Every frieze group is equivalent to one and only one of the following seven subgroups of Isom* $\mathbb{C}$*:*

1. $\langle T_1 \rangle \cong \mathbb{Z}$
2. $\langle T_1 M \rangle \cong \mathbb{Z}$
3. $\langle T_1, M \rangle \cong \mathbb{Z} \oplus \mathbb{Z}_2$
4. $\langle T_1, R_\pi \rangle \cong \mathbb{Z} \times_\alpha \mathbb{Z}_2 = \langle x, y \,|\, y^2 = e, yxy^{-1}x = e \rangle$
5. $\langle T_1, M_{\pi/2} \rangle \cong \mathbb{Z} \times_\alpha \mathbb{Z}_2 = \langle x, y \,|\, y^2 = e, \quad yxy^{-1}x = e \rangle$
6. $\langle T_1,\ M,\ M_{\pi/2},\ R_\pi \rangle$
7. $\langle T_2 M,\ \mathbb{R}_\pi,\ T_1 M_{\pi/2} T_{-1} \rangle$.

Proof:

V.5 The classification of wallpaper groups

Theorem V.5.1 (The classification of wallpaper groups) *Every wallpaper group is equivalent to one and only one of the following groups:*

Appendix A

The Language of Implication

In this chapter, we suppose that P and Q are statements.

The following all mean the same thing:

- If P then Q. (short for: "If P is true then Q is true.")
- $P \implies Q$. ("P implies Q.")
- P is *sufficient* for Q.
- Q is *necessary* for P.
- Q is true if P is true.
- P is true only if Q is true.
- If not Q then not P. (i.e., [not Q] $\implies$ [not P]).

Note: The last statement is called the *contrapositive* of the first statement. (We emphasize that they all mean the same thing.) It is often natural to prove that

$P \implies Q$ by proving its contrapositive, The argument is called *proof by contradiction* and goes like this:

"Oh yeah? Suppose [not Q]. *Then .... Then we would have* [not P]. "

and everyone agrees that $P \implies Q$ is then proved.

Converses. The *converse* of the statement "$P \implies Q$" is the statement "$Q \implies P$" .

Common logical fallacies often involve confusing a statement with its converse — confusing what is sufficient with what is necessary. This fallacy is illustrated in the example,

"$\theta = 45°$ since $\sin\theta = \sqrt{2}/2$."

The statement quoted is wrong because the fact that $\sin\theta = \sqrt{2}/2$ is *necessary* but not *sufficient* for the fact that $\theta = 45°$ (for example, θ might be 135°, whose sine is $\sqrt{2}/2$).

If and only if.

The following statements all mean the same thing:

- $P \implies Q$ and $Q \implies P$.
- P is true if and only if Q is true.
- P is true iff Q is true.
- $P \iff Q$.
- P is *necessary and sufficient* for Q.
- Q is *necessary and sufficient* for P.
- P and Q are *equivalent*.

Induction. The *principle of mathematical induction* is a procedure by which one establishes that every one of a certain sequence $P_n,\ n = 1, 2, 3, 4, ...$ of statements is true. Here is the precise statement:

If $P_1, P_2, \ldots, P_n, \ldots$ is an infinite sequence of statements and **if**

i) P_1 is true, **and**

ii) $P_n \Longrightarrow P_{n+1}$, for all integers $n \geq 1$,

then P_n is true for all positive integers n.

Example and counterexample.

Suppose as before that P and Q are statements, and you want to decide whether the implication $P \implies Q$ is true (as in many of the Exercises in this book). For example, suppose P is the statement, "x is a real number" and Q is the statement "x has a real square root r".

A good place to begin is to look for *examples*, i.e. cases where the hypothesis P is satisfied. Then, check to see if the conclusion Q is true *for this particular example.* If it is, then you have discovered *evidence in favor* of the implication $P \implies Q$ being true. *However, this is not enough to prove that the implication $P \implies Q$ holds.* For example, while the numbers $0, 1, 2, 3, 4, 5$ all have real square roots, the number -1 does not have a real square root. Said another way, this last fact is a *counterexample* to the assertion, $P \implies Q$: we have *found a particular instance when P is true but Q is not true.*

When constructing your list of examples, start simple. Then, gradually add other examples to your list. Try to make the examples of as varied a flavor as possible, so that your list of examples doesn't fall into some special class. This would have the effect of adding a hypothesis to P. For example, in the list $x = 1, 2, 3, 4, 5$ above all the numbers are positive, and positive numbers do indeed have real square roots.

This process actually applies to definitions as well. When encountering a new definition for the first time, try to come up with *examples* which fit the definition, as well as *counterexamples* which do *not.* Start simple, and gradually increase the complexity of the examples until you comprehend what the definition is really about.

Appendix B

How to Read a Proof

The crucial element in reading a proof is that it be an *active* rather than a passive process. The goal is not simply to hear what the author says. Rather, the goal is to own the proof at the end of the process. The following steps give a guide as to how to proceed.

1. Before starting to read the proof

Step A: read the claim. Ask yourself:

- do I understand the meaning of each term used?
- do I understand what is being asserted?

If so: Continue on to **Step B**.

If not:

- Review the definitions.
- Think of an *example* where the hypothesis is satisfied, and check if the conclusion is satisfied.

Step B:

Close the book and try to think of *why* the assertion should be true. How would you start to try to prove it? How would you use each of the hypotheses? You must struggle for some time at this point, so that when reading the author's proof you will understand what battles she/he fought — understand the issues involved. (Sometimes this step will suffice, without having to read what the author wrote! You will prove the assertion yourself, and the sun will shine upon you. It's a great feeling.)

2. Read the proof. At each step, ask yourself:

- Do I understand the logic at each point of the proof?

If not:

- Think through the argument in an example,
- Try to figure out what earlier facts and definitions the author is using,
- Try again. Try harder.

Reading mathematics is a slow business. The race is not to the swift, but to the thorough. There is no reason to be discouraged if a proof is not understood at first reading (or second or...).

3. After reading the proof, ask yourself:

- Do I know where each of the hypotheses was used?

If so good! This raises the odds that the proof was correct and that you understand it.

If not, either

- the proof is incorrect,

- the theorem could be stated with fewer hypotheses,
- or you missed something in reading the proof: go back to **Step 2.**

Appendix C

Sets and equivalence relations

If mathematics has a "periodic table of elements", as Paul Halmos has asserted, then the notion of "set" plays a role something like carbon: it permeates every living thing.

Common synonyms for the term "set" include "collection" and "family". Sets are comprised of "elements", often called "members".

The following are all equivalent ways of defining the same thing:

- Let $\mathcal{F}$ be the set of functions $f : \mathbb{C} \to \mathbb{C}$ that are given by the formula $f(z) = az + b$, where a, b are complex numbers with $|a| = 1$.
- Let
$$\mathcal{F} = \{f : \mathbb{C} \to \mathbb{C} \mid f(z) = az + b,\ a, b \in \mathbb{C},\ |a| = 1\}.$$
- Let $\mathcal{F}$ be the family of functions from the complex plane to itself such that $f(z) = az + b$, where a, b are complex numbers and $|a| = 1$.

Note the differences in language between the first and third descriptions.

It is important to realize that one can make a set whose elements are just about anything: numbers, functions, subsets of other sets, vector spaces, cars, children, etc.

Cartesian products.

Here is a common way of constructing new sets out of old ones.

Definition C.1 (Cartesian product) *If X and Y are sets, the* **Cartesian product** *$X \times Y$ is the set of ordered pairs*

$$X \times Y = \{(x, y) \mid x \in X,\ y \in Y\}.$$

So e.g. $\mathbb{R}^2$ is the Cartesian product of $\mathbb{R}$ and itself.

Relations.

Definition C.2 (Relation) *A* **relation** *on a set X is a nonempty subset $\mathcal{R}$ of $X \times X$. If $(x, y) \in \mathcal{R}$ then we write $x\mathcal{R}y$. The relation is called:*

1. **reflexive** *if for all $x \in X$ we have: $x\mathcal{R}x$.*
2. **symmetric** *if for all $x, y \in X$ we have: $x\mathcal{R}y \implies y\mathcal{R}x$.*
3. **transitive** *if, for all $x, y, z \in X$ we have: $x\mathcal{R}y$ and $y\mathcal{R}z \implies x\mathcal{R}z$.*

A relation satisfying all three properties is called an **equivalence relation** *and we write $x \sim y$ instead of $x\mathcal{R}y$.*

Examples. Let X be the set of people in the world. Let $\mathcal{R}$ be the relation on X given by

1. "is the sibling of": If you consider yourself your own sibling, then this relation is reflexive, symmetric, and transitive, i.e. is an equivalence relation.
2. "is an ancestor of": This is neither reflexive nor symmetric, but is transitive.
3. "is at least as old as": This is reflexive and transitive but not symmetric.

Partitions and equivalence relations.

Suppose $\sim$ is an equivalence relation on a set X. Given $x \in X$, we can look at all of the elements which are equivalent to x and lump them together into a subset of x, called the *equivalence class* of x. The equivalence class of x is often denoted by $[x]$.

At this point, we define the notion of a *partition* of a set, which, as we shall see, is another way of looking at an equivalence relation.

Definition C.3 (Partition) *A* **partition** *of a set X is a family $\mathcal{F}$ of subsets of X satisfying the following properties:*

1. $X = \bigcup_{A \in \mathcal{F}} A$
2. *If* $A, B \in \mathcal{F}$ *and* $A \neq B$ *then* $A \cap B = \emptyset$.

In words: *The sets in the family $\mathcal{F}$ are pairwise disjoint and their union is all of X.* For example, the set X of undergraduate students at a university is partitioned into a family $\mathcal{F}$ of four disjoint subsets $A_1, ..., A_4$ consisting of freshmen, sophomores, juniors, and seniors. Sometimes (1) and (2) are combined using the "disjoint union" notation $\sqcup$ into the single condition:

$$X = \bigsqcup_{A \in \mathcal{F}} A.$$

Every equivalence relation determines a unique partition, and vice versa, according to the following proposition.

Proposition C.4 (Equivalence classes partition) *1. Given a set X and an equivalence relation $\sim$ on X, the equivalence classes form a partition of the set X.*

2. *A partition $\mathcal{F}$ of a set X defines an equivalence relation $\sim$ by*

$$x \sim y \text{ if and only if } \{x, y\} \subset A \text{ for some } A \in \mathcal{F},$$

i.e. two points are called equivalent precisely when they lie in the same set in the family.

Proof: (1) Each element belongs to its own equivalence class by the reflexivity property, so the union of all the equivalence classes is the whole set X and condition (1) of definition C.3 holds. If $A = [a], B = [b]$ are two equivalence classes, and if $c \in A \cap B$, then $a \sim c,\ c \sim b$, so by transitivity and symmetry we have $a \sim b$. Thus $A = B$ and condition (2) holds.

(2) On the other hand, suppose given the partition $\mathcal{F}$. If $x \in X$ then by Condition (1), $x \in A \in \mathcal{F}$ for some $A \in \mathcal{F}$. Then $\{x, x\} \subset A$ and we have $x \sim x$. The relation is symmetric because

$$x \sim y \implies \{x, y\} \subset A \text{ for some } A \in F \implies \{y, x\} \subset A \implies y \sim x.$$

Finally, $x \sim y$ and $y \sim z$ implies that $\{x, y\} \subset A$, and $\{y, x\} \subset B$ for some $A, B \in \mathcal{F}$. But then $A \cap B \neq \emptyset$, so by Condition (2) of the definition, $A = B$. Hence $\{x, z\} \subset A \in \mathcal{F}$ and $x \sim z$. Thus the relation is transitive. ∎

Examples.

1. Let G be a group acting on a set X, and define $x \sim y$ if x, y belong to the same orbit. Then $\sim$ is an equivalence relation on X. Proposition II.5.8 says that the orbits partition X, and therefore this proposition is a special case of the above Proposition C.4.
2. Let G be a group and H a subgroup. Define $x \sim y$ if $xH = yH$. Then $\sim$ is an equivalence relation on G, and the fact that the cosets of H in G form a partition of G is another special case of Proposition C.4.
3. Let G be a group and let $\sim$ be the relation: $g_1 \sim g_2$ if g_1, g_2 are conjugate in G. Then $\sim$ is an equivalence relation on G (Lemma III.2.3). The "conjugacy classes" are precisely the equivalence classes under this equivalence relation, and form a partition of G.

Exercises C

1. Let X be the set of subsets of the plane, and define $\mathrm{P}_1 \sim \mathrm{P}_2$ if P_1 is congruent to P_2, i.e there exists an isometry $f : \mathbb{C} \to \mathbb{C}$ with $f(\mathrm{P}_1) = \mathrm{P}_2$. Check that $\sim$ is an equivalence relation on subsets of the plane. What properties of the set of isometries do you use in this verification?

Appendix D

Functions

In this section, we discuss generalities about functions. Let X and Y be any sets. Recall that a *function* $f : X \to Y$ is a rule which assigns, to each element x of X, an element y of Y, denoted by $f(x)$. It will be important for us to remember that *the set X, the set Y, and the rule f are all part of the data of the definition of the function.*

If $f : X \to Y$ is a function, the set X is called the *domain* of f, and we shall call Y the *target set* of f.

Images and preimages. If A is any subset of X, we let

$$f(A) = \{f(x) \mid x \in E\}$$

and call $f(A)$ the *image of the set A under the function f* . We'll sometimes shorten this and call $f(A)$ simply the *image of A* if we both know which function we're talking about. [1]

[1] Our use of the term "target set" is nonstandard. Some texts refer to Y as the *range* of f. However, other texts use the term *range* to refer to the image of X under f. Our term *target set* is meant to spare you this confusion later.

If $B \subset Y$, then *inverse image of B under f* is defined as

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}.$$

Note that the function f need not have an inverse (see below). The term *preimage* is used synonymously with inverse image. When B is a single element of Y, the preimage of y is sometimes called the *fiber over y* when the function is understood.

Examples are discussed below.

The symbol $\mapsto$. In many settings, we want to describe the rule for a function without giving a name for it. In this case we use the symbol $\mapsto$ to indicate the image of an element under the function being described. So e.g.

$$x \mapsto x^2$$

denotes the rule which sends x to x^2.

Extensions and restrictions. If $f : X \to Y$ is a function, there are a couple of standard ways of constructing new functions from this data. If we keep the rule and the target set the same, and replace X with a *subset* $A \subset X$, we get a new function called the *restriction of f to A*, typically denoted by $f|A$. The inverse of this procedure is called *extension*. Suppose $X \subset W$ and we have a function $f : X \to Y$. If we can find a function $g : W \to Y$ such that $g|X = f$, then we call g an *extension of f to W*.

Definition D.1 (1-1,onto,bijection) *A function $f : X \to Y$ is called*

1. one-to-one *if $f(x_1) \neq f(x_2)$ whenever $x_1 \neq x_2$;*
2. onto *if for every $y \in Y$, there is an element $x \in X$ for which $f(x) = y$;*
3. *a* bijection *if f is both one-to-one and onto.*

Said another way, a function $f : X \to Y$ is one-to-one if the preimage of an element in Y consists of *at most one, possibly no* elements of X. The function f is onto if the preimage of an element in Y consists of *at least one, possibly more* elements of X.

Examples.

1. Take $X = Y = \mathbb{R}$ and $f(x) = x^2$. Then f is neither one-to-one (since e.g. $f(-1) = f(1) = 1$) nor onto (since e.g. there are no real numbers x for which $x^2 = -1$).

   If $y < 0$ the preimage of y is empty; if $y = 0$ then the preimage of y is just 0, i.e. is one point; if $y > 0$ then the preimage of y consists of the two square roots of y.

2. Take $X = \mathbb{R}$, $,Y = [0, +\infty)$, and $f(x) = x^2$ Then f is not one-to-one (since e.g. $f(-1) = f(1) = 1$) but is now onto (since every nonnegative number has a square root).

3. Take $X = [0, +\infty)$ and $Y = \mathbb{R}$. Then f is one-to-one (since $x_1^2 = x_2^2$ only when $x_1 = \pm x_2$, and we've thrown away the negative numbers) but is not onto (since there are no real numbers x for which $x^2 = -1$.)

4. Take $X = Y = [0, +\infty)$ and $f(x) = x^2$. Then f is one-to-one and onto. (To see this, put together the discussion of the preceding examples.) Hence f is a bijection from X to Y.

5. Let Z be any set. Then the *identity function on* Z defined by $id_Z(z) = z$ for $z \in Z$ is a bijection. Though it's not a very interesting one, it is important.

The above examples are meant to emphasize the point raised in the first paragraph: that *a function has three pieces of information: the domain, the target set, and the rule*. In the above examples, we've kept the rule the same, but changed the domain and target set. Thus, properties like being one-to-one and onto are properties involving not just the rule, but of all three parts of the function.

We continue with some further generalities about functions; more will be established in the Exercises for this section.

Proposition D.2 *Let $f : X \to Y$ be a function. Then the following three conditions are equivalent:*

1. *f is a bijection;*
2. *there is a function $g : Y \to X$ such that*
$$f \circ g = id_Y \text{ and } g \circ f = id_X.$$
3. *there is a bijection $g : Y \to X$ such that*
$$f \circ g = id_Y \text{ and } g \circ f = id_X.$$

When these conditions hold, we write $f = g^{-1}$ and $g = f^{-1}$.

Proof:
$1 \implies 2$:Suppose f is a bijection. Set

$$g(y) = \text{the only point } x \in X \text{ with } f(x) = y.$$

(Existence of at least one x with $f(x) = y$ is guaranteed by the fact that f is onto. That there is exactly one such x is guaranteed by the fact that f is one-to-one.) We now check that $g \circ f = \mathrm{id}_X$. Well, choose $x \in X$ and let $y = f(x)$. We have

$$g(y) = g(f(x)) = \text{the only point } x' \in X \text{ with } f(x') = f(x) = y$$

by the definition of g applied to $y = f(x)$. So $x' = x$, since f is one-to-one. So $g \circ f(x) = f(x)$ for all $x \in X$ and we have shown $g \circ f = \mathrm{id}_X$. Showing that $f \circ g = \mathrm{id}_Y$ is similar.

$2 \implies 3$: We must show g is a bijection. Pick $y_1, y_2 \in Y$ and suppose $g(y_1) = g(y_2)$. Then

$$\begin{array}{rcccl} y_1 & = & \mathrm{id}_Y(y_1) & = & f(g(y_1)) \quad \text{since } f \circ g = \mathrm{id}_Y \\ & & & = & f(g(y_2)) \quad \text{since } g(y_1) = g(y_2) \\ & & & = & \mathrm{id}_Y(y_2) \quad \text{since } f \circ g = \mathrm{id}_Y \\ & & & = & y_2 \end{array}$$

and so $y_1 = y_2$. Hence g is one-to-one.

To see that g is onto, let $x \in X$ be any point in the target set of g, which is the domain of f. We must produce some $y \in Y$ for which $g(y) = x$. Take $y = f(x)$. Then $g(y) = g(f(x)) = \mathrm{id}_X(x) = x$ and we're done.

$3 \implies 1$**:**(Sketch) Imitate the proof above, with the roles of f and g interchanged, to show that f is one-to-one and onto. ■

If $f : X \to Y$ is any bijection and $g : Y \to X$ satisfies $f \circ g = \mathrm{id}_Y, g \circ f = \mathrm{id}_X$, we say that *$g$ is an inverse of f* and that *f is an inverse of g*. *A priori* f may have more than one inverse. But in fact this does not happen:

Proposition D.3 *If $f : X \to Y$ is a bijection, then there is a unique function $g : Y \to X$ such that $f \circ g = id_Y$ and $g \circ f = id_X$, and this g is a bijection.*

Proof: Existence of g, and the fact that it is a bijection, follows from the previous proposition. So it remains only to show uniqueness. Suppose $f \circ g_1 = f \circ g_2 = \mathrm{id}_Y$ and $g_1 \circ f = g_2 \circ f = \mathrm{id}_X$. Let $y \in Y$ be arbitrary. Then

$$\begin{array}{rcll} g_1(y) & = & g_1(f \circ g_2(y)) & \text{since } f \circ g_2(y) = y \\ & = & (g_1 \circ f)(g_2(y)) & \text{since composition is associative} \\ & = & \mathrm{id}_X g_2(y) & \text{since by assumption } g_1 \circ f = \mathrm{id}_X \\ & = & g_2(y). & \end{array}$$

Hence $g_1 = g_2$. ■

So from now on, when we speak of inverses, we can speak of *the* inverse.

We conclude with a general property of compositions of functions.

Proposition D.4 *If $f_1 : X_1 \to X_2, f_2 : X_2 \to X_3$, and $f_3 : X_3 \to X_4$, then $f_3 \circ (f_2 \circ f_1) = (f_3 \circ f_2) \circ f_1$.*

Proof: Let $x \in X_1$. Then $((f_3 \circ f_2) \circ f_1)(x) = (f_3 \circ f_2)(f_1(x)) = f_3(f_2(f_1(x)))$. But $(f_3 \circ (f_2 \circ f_1))(x) = f_3(f_2(f_1(x)))$ also, so the two compositions are the same. ■

As a special case, consider the above when $X_1 = X_2 = X_3 = X_4 = X$. Then *the binary operation of composition of functions from X to itself is associative.* In particular, if the elements of a group are functions from a set to itself and the operation is composition, the associative axiom for groups is automatically satisfied.

The symbol $\mapsto$, read "maps to", is used when we want to talk about a function without specifying a name or symbol for that function. For instance, if z is a complex number, and the domain and range are both taken to be $\mathbb{C}$, then $z \mapsto z^2$ means the function which sends z to its square z^2. We could have written this as $f(z) = z^2$, but this requires us to give a *name* f.

Exercises D

1. **If** $f_1 : X_1 \to X_2$ and $f_2 : X_2 \to X_3$ are both one-to-one, **then** the composition $f_2 \circ f_1 : X_1 \to X_3$ is one-to-one.

2. **If** $f_1 : X_1 \to X_2$ and $f_2 : X_2 \to X_3$ are both onto, **then** the composition $f_2 \circ f_1 : X_1 \to X_3$ is onto.

3. **If** $f_1 : X_1 \to X_2$ and $f_2 : X_2 \to X_3$ are both bijections, **then** the composition $f_2 \circ f_1 : X_1 \to X_3$ is a bijection.

4. **If** $f : X \to Y$ and there exists a function $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$, **then** f is one-to-one.

5. **If** $f : X \to Y$ and there exists a function $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$, **then** f is onto.

6. **If** X and Y are finite, $\#X = \#Y$, and $f : X \to Y$ is one-to-one, **then** f is a bijection. Compare this with the following: If $T : V \to W$ is a linear transformation between finite-dimensional vector spaces of the same dimension, and if T is one-to-one, then T is invertible.

7. **If** X and Y are finite, $\#X = \#Y$, and $f : X \to Y$ is onto, **then** f is a bijection. Compare this with the following: if $T : V \to W$ is a linear transformation between finite-dimensional vector spaces of the same dimension, and if T is onto, then T is invertible.

8. Give an example of sets X, Y and $f : X \to Y$ which is one-to-one, but not onto.

9. Show that there is a bijection between $\mathbb{Z}$, the set of integers, and $\mathbb{Q}$, the set of rational numbers.

10. Let X denote the positive real numbers, Y the set of complex numbers, and let $f(x) = \sqrt{x}$ for the usual positive square root of x. Find an extension of f to the set W of complex numbers whose real part is positive.

11. If f is one-to-one, then any restriction of f is one-to-one.

12. If f is onto, then any extension of f is onto.

13. If f is a bijection, then any extension of f is a bijection.

Appendix E

"Is That Function Well-Defined?"*

Frequently the key assertion in a proof is that *the function is well-defined*. To understand the issue involved, recalled that, by definition,

> a function $f : X \longrightarrow Y$ is a rule which associates to each element $x \in X$ *one* element $y \in Y$, usually denoted by $y = f(x)$.

Different elements x_1, x_2 might have $y = f(x_1) = f(x_2)$, — in which case the function is not one-one — but the essence of being a function is that for each $x \in X$ there is only one value $f(x)$ given by the rule.

The difficulty which frequently occurs is that the natural method of proceeding in a discussion often fails to give a clear and unambiguous value for "$f(x)$". When this happens the function is *not well-defined* and we must sharpen our rule before proceeding. When we wish to stress the fact that the rule given allows exactly one value for $f(x)$, so that f really is a function, we say that "f *is a well-defined function*".

Examples.

0* It is a mark of sophistication to be able to understand and ask this question. At a social gathering of mathematicians, nothing will make you one of the cognoscenti faster than to ask "*Is that function well-defined?*". This is even more effective than the old classic, "*But what happens in the non-abelian case?*" or that conversation stopper "*... yes, but only if the action is discrete*".

Example E.1 *"Let* $f : \mathbb{C} \longrightarrow \mathbb{C}$ *by* $f(z) = \sqrt{z}$*."*

This is not well-defined unless we are more precise about what "$\sqrt{z}$" means. For each complex number $z \neq 0$ there are two numbers whose square is z. For example, if $z = i$ we have

$$z = \left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right)^2 = \left(-\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i\right)^2.$$

Which one of these should we call "$\sqrt{z}$"?

Example E.2 *"Let* $f : \mathbb{C} - \{0\} \longrightarrow \mathbb{R}$ *by* $f(z) = \cos(\theta)$*, where* z *has polar coordinates* (r, θ) *for some* $r > 0$*."*

Using the language of section I.2, this can be written as: $f(z) = \cos(\arg z)$. Since $\theta = \arg(z)$ takes on infinitely many different values (any two of which differ by an integer multiple of 2π), this is an example which is immediately suspect and for which one must definitely prove that the function f is well-defined. It turns out that the cosine function has the property that

$$\cos\theta = \cos(\theta + 2n\pi) \quad \text{for all } n \in \mathbb{Z}$$

so that $f(z)$ indeed takes on only one value, no matter which of the values of θ is used. It follows that f is indeed well-defined.

The next two examples come from Chapters II and IV. The reader who hasn't gotten this far yet might want to delay reading them. Alternatively the reader might try to understand the issues that will be coming up.

Example E.3 (See Section II.7) *Let*

$G = \{-1, 1\}$, *under the operation of ordinary multiplication,*

$S_n =$ *the group of permutations of* $\{1, 2, \ldots, n\}$ *under composition,*

$\sigma : S_n \longrightarrow G$ *by*

$\sigma(g) = +1$ *if f is the product of an even number of transpositions (i.e., f is an even permutation).*

$\sigma(g) = -1$ *if f is the product of an odd number of transpositions (i.e., f is an odd permutation).*

The potential problem here is that I might think of a way to write a given permutation g as a product of an even number of transpositions, while my friend might think of a way of writing g as a product of an odd number of transpotions. If this were possible, then σ would not be well-defined, since $\sigma(g)$ would equal both $+1$ and -1. A major result in Section II.7 is Theorem II.8.9, which states that this cannot happen. Thus σ is well-defined.

Example E.4 (See Section IV.4) *If H is a subgroup of the group G, we define the quotient group G/H by*

$G/H = \{xH \mid x \in G\}$ *(the set of left cosets of H in G),*
$(xH)(yH) = (xy)H$ *for all $x, y \in G$.*

In this situation one might ask: *Is the multiplication well-defined?*

To understand the meaning of this question, recall that a multiplication rule is a function $M : G/H \times G/H \longrightarrow G/H$. Thus in this situation we are defining a function M as follows: *If C_1 and C_2 are left cosets of H in G, choose elements $x \in C_1$, $y \in C_2$, so that $C_1 = xH$, $C_2 = yH$, and define*

$$M\Big(C_1, C_2\Big) = M\Big(xH, yH\Big) = (xH)(yH) = (xy)H.$$

The question then becomes: *Is the function M well-defined?* To prove it's well-defined one must prove that if different elements x', y' are chosen from C_1 and C_2, then we will get the same answer for $M(C_1, C_2)$:

$$(xH)(yH) = (x'y')H.$$

This will *not* always happen. The multiplication M will not always be well-defined. From Section IV.4 we get the following, whose proof is a recommended exercise:

Theorem E.5 *Suppose that G is a group and that H is a subgroup of G. Let G/H be the set of all left cosets of H in G. Then the multiplication on G/H given by*

$$(xH)(yH) = (xy)H$$

is well-defined if and only if H is a normal subgroup of G.

More generally, suppose X is a set, $\sim$ is an equivalence relation on X, and suppose f is a function defined on X. Let $X/\sim$ denote the set of equivalence classes of X. Under what circumstances does f determine, in a reasonable fashion, a function on $X/\sim$?

The obvious thing to do is the following: let A be an equivalence class. Then we can write $A = [x]$ for some $x \in X$. Define

$$F([x]) = f(x).$$

The difficulty is that this need not be well-defined, since there may be many different elements $x, y, z, ...$ in X for which $A = [x] = [y] = [z]$. Thus, we don't know which element to use in the formula: $\phi(x), \phi(y), \phi(z), ...$ The outcome will be successful, however, if we have

$$x \sim y \implies f(x) = f(y).$$

This is sometimes summarized in a diagram like Figure 1, where $q : X \to X/\sim$ denotes the "quotient" function $x \mapsto [x]$.

Figure 1. The function f on X determines a function F on $X/\sim$ exactly when $x \sim y \implies f(x) = f(y)$.

Examples.

1. The construction of the homomorphism $\Phi : G/N \to \phi(G)$ of Section IV.4 is an example of this kind of construction: $f = \phi$, $X = G$, and $\sim$ is the equivalence relation determined by the partition of G into cosets of N.

2. We can interpret the construction of the multiplication operation on $G/N \times G/N$ of Example E.4 in this context as well.

   To do this, put $X = G \times G$. Multiplication gives a function:

   $$m : G \times G \to G.$$

   The partition of G into cosets of a normal subgroup N gives a partition of $G \times G$ as well, hence an equivalence relation:

   $$(a_1, b_1) \sim (a_2, b_2) \text{ if } a_1 \sim a_2 \text{ and } b_1 \sim b_2,$$

   i.e. (a_1, b_1) and (a_2, b_2) are equivalent if $a_1 N = a_2 N$ and $b_1 N = b_2 N$. The quotient map $q : G \to G/\sim = G/N$ gives us a quotient map $q \times q : G \times G \to G/N \times G/N$ by applying q in each factor. In the language above, let $f : G \times G \to G/N$ be given by $f = q \circ m$. Since the multiplication operation is well-defined, we get a function $F = M : G/N \times G/N \to G/N$; see Figure 2.

Figure 2. Multiplication $m : G \times G \to G$ "descends" to a multiplication $M : G/N \times G/N \to G/N$ when N is a normal subgroup of G.

Appendix F

Linear algebra

This will be a brief summary of definitions, ideas, and theorems from linear algebra which we have used. No proofs will be given. Details can be found in almost any text on the subject.

Vector spaces.

Definition F.1 (Vector space) *Let k denote either the real numbers or complex numbers. A* vector space over k *is a nonempty set V together with two operations:*

$$V \times V \to V, \;\; k \times V \to V$$

called vector addition *and* scalar multiplication*, respectively, such that the following axioms hold for all $\vec{u}, \vec{v}, \vec{w} \in V$ and all scalars $c, d \in k$,*

1. *if $+$ denotes vector addition,*
   (a) $\vec{u} + \vec{v} = \vec{v} + \vec{u}$
   (b) $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$
   (c) *There is a vector $\vec{0} \in V$ such that $\vec{v} + \vec{0} = \vec{v}$ for all $\vec{v} \in V$*
   (d) *For each $\vec{v} \in V$, there is a vector $-\vec{v} \in V$ such that $\vec{v} + (-\vec{v}) = \vec{0}$.*

2. *If $c\vec{u}$ denotes scalar multiplication, then*
   (a) $(cd)\vec{v} = c(d\vec{v})$
   (b) $1\vec{v} = \vec{v}$
3. *Vector addition and scalar multiplication are related by the following rules:*
   (a) $(c+d)\vec{v} = c\vec{v} + d\vec{v}$
   (b) $c(\vec{u}+\vec{v}) = c\vec{u} + c\vec{v}$

Notice that condition (1) says that $(V, +)$ is an abelian group.

Examples.

1. Let
$$V = \mathbb{R}^n = \{(x_1, ..., x_n) \mid x_i \in \mathbb{R}\}$$
where vector addition and scalar multiplication are performed coordinatewise. Then V is a vector space.
2. Let $V = \mathbb{C}$ and $k = \mathbb{R}$, and define addition by addition of complex numbers and multiplication by multiplication of real number by complex number. Then V is a vector space which is isomorphic (see below) to $\mathbb{R}^2$.
3. (a digression) Let V be the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. For $f, g \in V$ define $f + g$ to be the function whose value at t is $f(t) + g(t)$ and define cf to be the function whose value at t is $cf(t)$. Set the zero function to be, the function whose value at t is zero for all t. Then V is a vector space.

Linear independence. A subset $\{\vec{v}_1, ..., \vec{v}_p\} \subset V$ is called *linearly independent* if
$$c_1\vec{v}_1 + ... + c_p\vec{v}_p = \vec{0} \implies c_i = 0, \text{ all } i.$$
Otherwise, we say that the set is *linearly dependent.*

For example, set $\vec{u}, \vec{v}, \vec{w} = (1,0),\ (0,1),\ (1,1)$, respectively. Then $\{\vec{u}, \vec{v}\}$ is linearly independent, since $c_1\vec{u} + c_2\vec{v} = (c_1, c_2) = \vec{0} \implies c_1 = c_2 = 0$. The set $\{\vec{u}, \vec{v}, \vec{w}\}$, however, is linearly dependent since
$$1 \cdot \vec{u} + 1 \cdot \vec{v} + (-1) \cdot \vec{w} = \vec{0}.$$

Spanning sets. A subset $\{\vec{v}_1, ..., \vec{v}_p\} \subset V$ is said to *span* V if given any $\vec{v} \in V$, there are $c_1, ...c_p$ such that

$$c_1\vec{v}_1 + ... + c_p\vec{v}_p = \vec{v}.$$

Bases and dimension. A subset $\mathcal{B}$ of V which is both linearly independent and spans V is called a *basis* of V. We leave it as an exercise to check that *if* $\mathcal{B} = \{\vec{b}_1, ..., \vec{b}_n\}$ *is a ordered* [1] *basis of* V*, then every* $\vec{v} \in V$ *can be written in exactly one way as:*

$$\vec{v} = c_1\vec{v}_1 + ... + c_n\vec{v}_n.$$

It turns out that if V has a finite basis, then any two bases $\mathcal{B}, \mathcal{B}'$ have the same number of elements, and this common number is called the *dimension* of V.

The *standard basis* for $\mathbb{R}^2$ is the basis:

$$\vec{e}_1 = (1, 0) \quad \text{and} \quad \vec{e}_2 = (0, 1).$$

Notice that, to express any vector $(x, y) = c_1\vec{e}_1 + c_2\vec{e}_2$ in terms of this basis, we just have $c_1 = x$ and $c_2 = y$.

Linear transformations. Let V, W be vector spaces over k. A function $T : V \to W$ is called *linear* if for all $\vec{u}, \vec{v} \in V$ and all $c, d \in k$, we have $T(c\vec{u} + d\vec{v}) = cT(\vec{u}) + dT(\vec{v})$. Taking $c = d = 1$, we see that a linear transformation is a homomorphism between abelian groups $(V, +)$ and $(W, +)$. A linear transformation which is a bijection is called an *isomorphism*. If V is n-dimensional, then V is isomorphic to $\mathbb{R}^n$: a choice of ordered basis $\mathcal{B}$ defines a function:

$$\vec{v} \mapsto (c_1, ..., c_n) \in \mathbb{R}^n$$

[1] This means that we can talk about the *first* element of $\mathcal{B}$, the *second*, etc.

where the c_i's are the uniquely determined coefficients appearing in the discussion of basis, and this function an isomorphism from V to $\mathbb{R}^n$. It is easily checked that the set of isomorphisms from a vector space V to itself form a group called the *general linear group of* V and denoted $GL(V)$ and .

Matrices. Let

$$A = [\vec{a}_1 ... \vec{a}_n]$$

be an m-row by n-column array, or *matrix*, of numbers whose columns are vectors in $\mathbb{R}^m$. Let $\vec{x} = (x_1, ..., x_n) \in \mathbb{R}^n$. The *product* $A\vec{x}$ is given by

$$x_1\vec{a}_1 + ... + x_2\vec{a}_n \in \mathbb{R}^m.$$

Thus the rule $T(\vec{x}) = A\vec{x}$ defines a function from $\mathbb{R}^n$ to $\mathbb{R}^m$; it is easily checked that this function is linear. Thus, matrix multiplication yields a linear transformation.

Conversely, suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear. Let $\mathcal{B}_n = \{\vec{e}_1, ..., \vec{e}_n\}$ be the standard basis for $\mathbb{R}^n$ and $\mathcal{B}_m$ the standard basis for $\mathbb{R}^m$. The *standard matrix* of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is the matrix

$$A = [T(\vec{e}_1)...T(\vec{e}_n)]$$

and it is easily checked using the linearity properties that the functions T and $\vec{x} \mapsto A\vec{x}$ are the same. Hence, every linear transformation arises as matrix multiplication.

Matrix multiplication. Suppose A is a p-column by m-row matrix, and suppose

$$B = [\vec{b}_1 \ ... \ \vec{b}_n]$$

is a n-column by p-row matrix. Thus, as linear transformations, $B : \mathbb{R}^p \to \mathbb{R}^m$ and $A : \mathbb{R}^n \to \mathbb{R}^p$. The *product* of A and B is defined by

$$AB = [A\vec{b}_1 \ ... \ A\vec{b}_n]$$

and is the standard matrix of the linear transformation which is the composition $\vec{x} \mapsto A\vec{x} \mapsto B(A(\vec{x}))$. Therefore, *matrix multiplication is associative since composition of functions is associative.*

The n-by-n matrix with ones on the diagonal and zeros everywhere else is the standard matrix of the identity function (which is a linear transformation) and is denoted I_n. A matrix A is *invertible* if there is an n-by-n matrix C for which $AC = I_n$. If such C exists, it is unique, satisfies $CA = I_n$ too, and is denoted A^{-1}. The set of invertible n-by-n matrices form a group under matrix multiplication, denoted $GL_n(\mathbb{R})$. The groups $GL(\mathbb{R}^n)$ and $GL_n(\mathbb{R})$ are isomorphic via the function which sends an invertible linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ to its standard matrix.

Similarity. Two n-by-n matrices A_1, A_2 are called *similar* if there is an invertible matrix B such that $A_2 = BA_1B^{-1}$. The relation of similarity is an equivalence relation on the set of n-by-n matrices. This is almost, but not quite, like the notion of conjugacy: we do not require that A_1, A_2 themselves are invertible. If we restrict to the setting where the A_i are invertible, then the equivalence relation of similarity is the same as the equivalence relation of conjugacy in $GL_n(k)$.

Determinants. The determinant of a two-by-two matrix is given by

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

and vanishes exactly when A is not invertible. It satisfies a multiplicative property:

$$\det(AB) = \det(A)\det(B).$$

More general determinants are somewhat tricky to define and compute. We refer the reader to [**?**] for a precise definition and characterization of determinants. For our purposes, we will simply state that there exists a function det defined on n-by-n matrices taking values in k such that:

1. $\det(A) = 0$ if and only if A is not invertible,
2. $\det(AB) = \det(A)\det(B)$

In fact, it can be computed recursively (at least for moderately sized matrics) using something called *cofactor expansion* as follows. Let $A = (a_{ij})$ be the array of entries of A. Let M_{ij} denote the submatrix obtained by deleting the ith row and jth column. Then

$$\det(A) = \sum a_{ij}(-1)^{i+j} \det(M_{ij})$$

where the sum is over any fixed row or column of A.

Inner products. Let V be a real vector space. An *inner*, or *dot* product on a vector space V is a function

$$V \times V \to \mathbb{R}$$

and written $\langle \vec{u}, \vec{v} \rangle$, such that for all $\vec{u}, \vec{v}, \vec{w} \in V$ and $c \in \mathbb{R}$ we have

1. $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$
2. $\langle \vec{u} + \vec{v}, \vec{w} \rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle$
3. $c\langle \vec{u}, \vec{v} \rangle = \langle c\vec{u}, \vec{v} \rangle = \langle \vec{u}, c\vec{v} \rangle$
4. $\langle \vec{v}, \vec{v} \rangle \geq 0$, with equality if and only if $\vec{v} = \vec{0}$.

The *standard inner product* on $\mathbb{R}^n$ is given by

$$\langle (x_1, ..., x_n), (y_1, ..., y_n) \rangle = x_1 y_1 + ... + x_n y_n.$$

On $\mathbb{R}^n$, this yields a definition of the length of a vector:

$$||\vec{v}|| = \sqrt{\langle \vec{v}, \vec{v} \rangle}.$$

It is easy to check that if $n = 1, 2, 3$ the definition of length of a vector $||\cdot||$ just given agrees with the usual notion. Having notions of lengths of vectors, we can easily define the distance between points (vectors) in $\mathbb{R}^n$ by setting the distance between $\vec{u}$ and $\vec{v}$ to be $||\vec{u} - \vec{v}||$.

An inner product also leads to a definition of angle between two vectors:

$$\theta = \cos^{-1} \frac{\langle \vec{u}, \vec{v} \rangle}{||\vec{u}|| \; ||\vec{v}||}.$$

Orthogonal transformations. Let V be a real vector space equipped with an inner product. If $T : V \to V$ is linear and satisfies

$$\langle T(\vec{u}), T(\vec{v})\rangle = \langle \vec{u}, \vec{v}\rangle$$

for all $\vec{u}, \vec{v} \in V$, i.e. if T *preserves the inner product*, then we say that T is *orthgonal* with respect to this inner product. Such transformations preserve lengths of vectors and angles between vectors, i.e. are *isometries* of V with respect to the distance determined by the inner product. An orthogonal transformation is necessarily invertible, and the set of orthogonal transformations forms a group under composition.

If $V = \mathbb{R}^n$ and the inner product is the standard one, and if T has standard matrix A, then T is orthgonal if and only if $AA^t = I_n$, where A^t denotes the transpose of A: the (i, j)th entry of A^t is the (j, i)th entry of A. The set of matrices A satisfying $AA^t = I_n$ forms a group under matrix multiplication (the proof is either by manipulation of matrices, or by appeal to the correspondence between matrix multiplication and composition of functions again). This group is called simply *the orthogonal group* and is denoted O_n.

Index